Measurements of the production of a W boson in association with jets and the search for Higgs pair production in the bbZ($\ell\ell$)Z(jj) decay channel with the CMS detector at the LHC

by Apichart Hortiangtham

B.Sc. in Physics, Mahidol University, Thailand
M.Sc. in Radiation Detection and Instrumentation, University of Surrey, UK

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

July 25, 2019

Dissertation directed by

Emanuela Barberis
Professor of Physics

Acknowledgements

First of all, I would like to express my gratitude to my advisor, Professor Emanuela Barberis for her guidance and support throughout my PhD journey. The completion of this thesis would not be possible without her encouragement and understanding.

I would also like to thank my Thesis Committee members, Professor Darien Wood, Professor Toyoko Orimoto, and Professor Pran Nath. It has been my privilege to work with and learn from them.

The studies in this thesis would not be successful without my colleagues Dr. Darin Baumgartel, Dr. Tomislav Seva, and Dr. Philippe Gras. They initiated the W+jets analysis framework used in this thesis and willingly answered my questions when I started the analysis. In addition, I would like to thank Dr. Bhawan Uppal and Dr. Kadir Ocalan for being great co-workers working on this analysis together. In addition, I would like to give a big thanks to Dr. David Morse, who has been working together with me in the double Higgs analysis. I have learned a lot from him during the past years.

I am also grateful to my collaborators at Northeastern University and CMS experiment. In particular, I would like to thank Professor George Alverson, Dr. Tim Cox, Dr. Daniele Trocino, Dr. Andrea Massironi, Dr. Yacine Haddad, Dr. David Nash, Dr. Matthew Chasco, Dr. Ren-Jie Wang, Dr. Rafael de Lima, Metin Yalvac, Andrew Wisecarver, Chad Freer, Tanvi Wamorkar, Bingran Wang, Gabriel Madigan, Abraham Tishelman-Charny, Vivan Nguyen, and Jingyan Li for their collaboration, help and support.

I would like to thank my family for all the love and understanding they have given me. Finally, I would like to thank my wife, Sirinrat Sithajan, who always supports and encourages me to pursue my goals.

Abstract of Dissertation

This PhD thesis presents two experimental particle physics analyses, the measurements of the production of a W boson in association with jets (W+jets) in proton-proton collisions using the muon decay channel and the search for resonant Higgs pair (diHiggs/double Higgs) production in proton-proton collisions using the bbZ($\ell\ell$)Z(jj) final state. Both analyses use data recorded by the Compact Muon Solenoid (CMS) experiment at the CERN Large Hadron Collider (LHC).

The W+jets analysis uses data from the proton-proton collisions at a center-of-mass energy ($\sqrt{s}$) of 8 TeV and 13 TeV collected during the LHC RunI and RunII (2015) periods, corresponding to an integrated luminosity of 19.6 fb^{-1} and 2.2 fb^{-1} respectively. The measured cross sections are reported as functions of jet multiplicity, jet transverse momenta (p_{T}), jet pseudorapidity (η), and the scalar sum of jet transverse momenta (H_{T}) for different jet multiplicities. Distributions of the angular correlations between the jets and the muon are examined, as well as the average number of jets as a function of H_{T} and as a function of angular variables. The measurements are compared to the theoretical predictions from next-to(-next-to)-leading-order calculations and event simulations that utilize matrix element calculations interfaced with parton showers.

The Higgs pair production analysis uses data at $\sqrt{s} = 13$ TeV collected in the LHC RunII (2016) period, corresponding to an integrated luminosity of 35.9 fb^{-1}. The bbZZ final state considered here is the one where one Z decays into two oppositely charged leptons, and the other Z decays hadronically into two or more jets. Upper limits at 95% confidence level are placed on the production cross section of narrow-width spin-0 and spin-2 particles decaying to Higgs boson pairs. This is the first search of its kind in this final state.

In addition to the analysis work, the thesis also describes the work of maintaining the conditions data and support the operation of the CMS Cathode Strip Chamber (CSC),

which is the main sub-detector for measuring muons in the forward/backward regions and thus a critical detector for these analyses.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction and Overview

Particle physics studies the structure of matter at the most fundamental level. After more than a century of theoretical and experimental research, scientists have come to believe that matter is made up of, at the subatomic level, a small number of fundamental building blocks which are referred to as elementary particles. In additions, these elementary particles are governed by specific rules, in forming atoms which make up us, the Earth, and the universe. Scientists have been refining the theory used to describe these elementary particles and to explain their interactions. The current and most widely accepted theory is known as the Standard Model (SM) of particle physics. On the experimental side, one of the modern methods to probe the interactions and test the structure of elementary particles is to perform high energy scattering experiments in which particles are accelerated and fired at others and the outcomes are examined. The advances in accelerator and detector technologies have allowed for a deeper understanding of particle interactions and unprecedented precision tests of the Standard Model.

Although the Standard Model has been very successful in predicting the results of numerous experiments, including the discovery of several particles, for example the discovery of the Higgs boson in 2012, the theory is far from complete as there are phenomena that cannot be explained by the Standard Model. Examples of such phenomena are the strong CP problem, neutrino oscillations, matter-antimatter asymmetry, and the explanation of dark matter and dark energy. Theoretical developments to account for the deficiencies of the Standard Model include various extensions of the SM, such as supersymmetry (SUSY), as well as entirely new theoretical frameworks, such as string theory and extra dimensions. These theories, collectively referred to as physics beyond the Standard Model (BSM), often predict the existence of new particles, hence the driving force for searching for new particles at particle physics colliders.

At present, the most powerful accelerator in the world is the Large Hadron Collider (LHC) at CERN which began its operation in 2008. The accelerator is approximately 27 km in circumference. It accelerates two beams of protons and collides them together. The LHC is designed to reach a maximum centre-of-mass energy of 14 TeV, the highest energy ever achieved. On the LHC ring, there are four detectors situated in huge underground caverns. These are the ATLAS, CMS, ALICE and LHCb detectors. The Compact Muon Solenoid (CMS) is a general-purpose detector designed to investigate the broadest range of physics possible.

This thesis work explores two complementary approaches, precision testing of the SM and searches for physics BSM, using the proton-proton collision data produced at the LHC and collected by the CMS detector. The precision tests of the SM are conducted through the W+jets measurements analysis. As for BSM, the search for resonant Higgs pair production, which stems from extra dimensions models is presented.

This thesis is organized in the following manner: Chapter 1 gives the overview of the thesis and the related theories. Chapter 2 describes the experimental apparatus used in acquiring

the data including a brief description of the LHC accelerator and of the CMS detector. The reconstruction of physics object is described in Chapter 3. The W+jets analysis is discussed in Chapter 4, where the 8 TeV and 13 TeV analyses are presented separately in Chapter 4.6 and 4.7, respectively. Chapter 5 presents the search for Higgs pair production. Finally, Chapter 6 concludes and summarizes the thesis.

1.2 The Standard Model of Particle Physics and Beyond

The Standard Model (SM) of particle physics is a relativistic quantum field theory (a consistent theory based on both quantum mechanics and special relativity) [1, 2, 3, 4]. The theory successfully describes the elementary particles of matter and their interactions through fundamental forces: the strong, weak and electromagnetic forces. The fourth known force, gravity, is not included in the framework. The gravitational effect is so tiny compared to the effects of the other forces on the short distances of particle interactions in high energy physics measurements, that gravity can be completely ignored.

The SM divides particles into fermions, gauge bosons and the Higgs boson according to their spin (intrinsic angular momentum) with values $1/2$, 1 ,and 0 respectively. According to the SM, matter is made of fermions which are further divided into quarks (q) and leptons (ℓ). Fig. 1.1 illustrates the fundamental particles in the SM and gives an overview of their properties. In addition to what is shown in Fig. 1.1, anti-particles exist for every fermion and quark. In the case of fermions, the anti-particles have the same mass, but opposite-sign electric charge. In the case of quarks, the anti-particles also have inverse color charge.

The quarks have electric charge of $+2/3$ or $-1/3$ in units of the electron charge (top and bottom purple rows in Fig. 1.1 respectively). In addition, they can be categorized into three generations, with increasing in mass, such that the first generation are up (u) and down (d), the second generation are charm (c) and strange (s), and the third generation are top (t) and

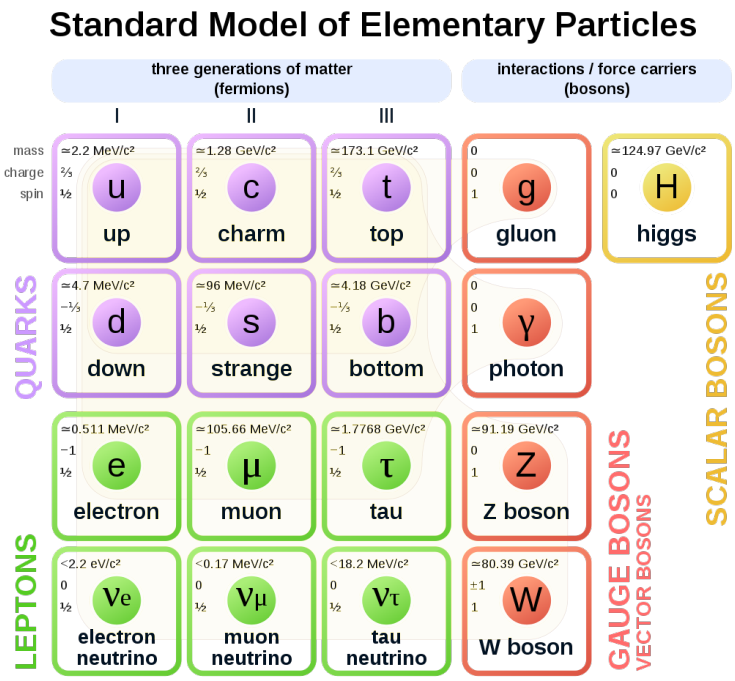


FIGURE 1.1: Overview of the fundamental particles in the Standard Model. Matter is made of quarks (in purple) and leptons (in green) which are collectively called fermions. Gauge bosons (in orange) act as the force carriers in the theory. The Higgs boson (in yellow) is responsible for the mass of each of these fundamental particles [5].

bottom (b). In addition to the electric charge, quarks carry weak and strong charges, where the latter is also known as the color charge. The color charge has three color states: red, green, and blue. Only color-neutral objects can exist freely, therefore quarks form composite states, called hadrons, which can be further categorized into baryons, bound states of three quarks, and mesons, a quark and its anti-quark bound state.

Leptons are arranged, similarly to quarks, into three generations in which each generation has two particles as depicted in Fig. 1.1. Each generation essentially shares similar properties to the first generation, except for the higher mass. Within each generation, there is a negatively-charged lepton and an electrically neutral neutrino. The charged leptons are the electron (e^-), muon (μ^-) and tau lepton (τ^-) with an electric charge of -1. The corresponding neutrinos are electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). Recent neutrino-oscillation experiments have revealed that their mass is non-zero although

they were initially thought to be massless.

In the SM, the interactions between fundamental particles are described by the exchange of gauge bosons that mediate the forces. The photon is the mediator of the electromagnetic interaction and couples to electric charge. The electromagnetic interaction is described by the theory of quantum electrodynamics (QED) [6, 7, 8].

The three massive gauge bosons, $W^{\pm}$ and Z^0 which couple to the weak charge of the particles are the mediators of the weak interaction. The photon is massless while $m_W = 80.4\,\text{GeV}$ and $m_Z = 91.2\,\text{GeV}$. The range of the electromagnetic force is infinite due to the zero rest mass of the photon, while the weak interaction has a very short range. The electroweak theory, by Glashow, Salam and Weinberg [9, 10, 11], unifies weak and electromagnetic interactions into a single framework.

The quantum field theory that describes the strong interaction is quantum chromodynamics (QCD) [12, 13]. The mediators of the strong interaction are eight gluons which couple to particles carrying color charge. Gluons are massless and can self-interact.

Charged leptons interact electromagnetically and weakly. Neutrinos are electrically neutral and can only interact weakly, hence, are difficult to be detected. Quarks interact electromagnetically, weakly, and strongly. Quarks may absorb or emit a W boson, causing a flavor transformation and allowing for radioactive decay processes such as beta decay and inverse beta decay.

While leptons are observed as single entities, single quarks have never been observed. This is due to a phenomena known as "quark confinement" or "color confinement". Confinement describes the phenomenon that upon growing separation among quarks in a color-neutral bound state, the energy stored in the field between them increases linearly with the distance. This continues until it becomes favorable to create a quark-antiquark pair from the vacuum,

which then form color-neutral bound states with the quarks previously undergoing separation. In collider experiments, quark confinement leads to the production of narrow jets of color-neutral mesons and baryons around the direction of flight of quarks produced in the collision.

Finally, the Brout-Englert-Higgs mechanism [14, 15] describes how particles acquire mass through the interaction with the Higgs field. The mechanism explains the mass of the $W^{\pm}$ and Z^0 bosons, while the photon can remain massless. The fermion masses originate from interactions with the Higgs field via a Yukawa couplings. The Higgs boson is the quantum manifestation of the Higgs field. It is a scalar boson with zero spin, and without electric charge nor color charge.

Although the success of SM, various observed phenomena remain unexplained. The SM does not offer candidates for dark matter, a mysterious matter whose existence have been supported by astrophysical observations. Another example is the lack of description of gravity. The SM has no explanation for the relative difference in the strength of the electroweak and gravitational forces. This is known as the "hierarchy problem". Several extensions to the SM attempt to address this problem. Such extensions often lead to the introduction of new particles which await to be discovered at the LHC.

1.3 Cross Section and Proton-Proton Collisions

The SM allows us to calculate the cross section σ of a process involving fundamental particles of the theory. Cross section essentially determines the likelihood of occurrence of a particular interaction, and is measured in units of area (barn = 10^{-24} cm^2). It can be generally computed as the ratio of the transition rate Γ_{fi} and the incoming particle flux Φ_{in}.

$$\sigma = \frac{\Gamma_{fi}}{\Phi_{in}} \tag{1.1}$$

The Γ_{fi} can be generally interpreted as a combination of two parts, the matrix element $\mathcal{M}_{fi}$ and the density of (final) states. For example, for a 2→2 process with a and b being the incoming and 1 and 2 the outgoing particles, the Γ_{fi} can be written as:

$$\Gamma_{fi} = \frac{(2\pi)^4}{4E_a E_b} \int |\mathcal{M}_{fi}|^2 \delta^{(4)}(p_a + p_b - p_1 - p_2) \cdot \frac{d^3\mathbf{p}_1}{2E_1(2\pi)^3} \frac{d^3\mathbf{p}_2}{2E_2(2\pi)^3} \tag{1.2}$$

where the p_a, p_b, p_1 ,p_2 are the four-momentum vectors of the incoming and outgoing particles, $\mathbf{p}_1$ and $\mathbf{p}_2$ are the three-vector momenta of the outgoing particles and the process-specific information is contained in the matrix element $\mathcal{M}_{fi}$. Feynman diagrams associated to a process can be deduced and the $\mathcal{M}_{fi}$ can be calculated using Feynman rules (when a perturbative approach can be used). An example of Feynman diagrams illustrating the 2→2 process is shown in Fig. 1.2.

FIGURE 1.2: Example of Feynman diagrams of a 2→2 QCD process (left) and a 2→2 EWK/QED process (right) which are possible in pp collisions. The figure is taken from K. S. Grogg.

However, the cross section mentioned is for the interaction of fundamental particles, e.g. u and ū in the left picture of Fig.1.2. It is also called partonic cross section, which means cross section for direct parton-parton interaction. As mentioned in the previous section, only bound states of quarks, e.g. protons, can be observed. Therefore, in colliding physics, beams

of protons are collided which can provide the collisions of partons that constitute them. The collisions have to be very energetic such that the constituent partons directly interacts, hence called "hard scattering" process. Physicist, therefore, have to study and calculate the cross section of pp collisions in order to study parton interactions. However, it is impossible to know or control the momentum of the partons inside the hadrons. Furthermore, the constituent quarks may exchange and radiate gluons producing virtual "sea quarks".

The QCD factorization theorem is used to compute the hadronic cross section from the parton-level cross section. The theorem states that the cross section for a hard scattering process between two hadrons can be computed by weighting the corresponding partonic cross sections $\hat{\sigma}_{ij}$ with the parton distribution function (PDF) $f_i(x, \mu_F^2)$ defined as the probability density for finding a parton i with a certain longitudinal momentum fraction x at a resolution scale μ_F^2 inside the incoming hadron. The cross section can be written as:

$$\sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 \left\{ f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \times d\hat{\sigma}_{ij}\left(p_1, p_2, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_R^2}, \mu_F^2\right) \right\}, \quad (1.3)$$

where P_1 and P_2 are the four-momentum of the colliding hadrons, and the μ_F is the factorization scale. The sum runs over all partons where $p_1 = x_1 P_1$ and $p_2 = x_2 P_2$. Q^2 is the characteristic scale of the hard scattering and μ_R is the renormalization scale of the QCD running coupling α_s.

For large energies (Q^2), α_s becomes small and $d\hat{\sigma}_{ij}$ can be perturbatively calculated as:

$$d\hat{\sigma}_{ij} = \sum_{m=0}^{n} d\hat{\sigma}_{ij}^{(m)} \alpha_s^m, \quad (1.4)$$

where n is the order of the perturbative calculation; n = 0, 1, and 2 corresponds to the leading-order (LO), next-to-leading-order (NLO), and next-to-next-to-leading-order (NNLO) calculations in perturbative QCD (pQCD).

1.4 Motivation

1.4.1 W+Jets

The W boson is a fundamental particle which was first discovered in 1983 by the UA1 and UA2 collaborations at CERN.[16, 17]. $W^{\pm}$ and Z bosons are responsible for the electroweak interaction governing the behavior of subatomic particles over short distances.

W bosons can be produced in hadron colliders via the electroweak interaction of a quark with an antiquark of different flavor. An example of the processes is illustrated in the left graph of Fig. 1.3 where an up quark and an anti-down quark annihilate and produce a W boson which subsequently decays. In practice, such processes can be accompanied by quark and gluon interactions governed by strong interaction, yielding additional partons and resulting in the associated production of W boson and jets (W+jets). For example, the colliding quark emits quantum chromodynamics (QCD) radiation in the form of gluon(s) before annihilation, as depicted in the right graph of Fig. 1.3. These W+jets events are ubiquitous in the modern hadron colliders such as Tevatron and LHC.

Measurements of vector bosons produced in association with jets provide stringent tests of perturbative QCD. In addition, the production of W+jets is a main background to rarer Standard Model (SM) processes, such as top quark production and Higgs boson production in association with a W boson, and it is also a prominent background to several searches for physics beyond the SM. Therefore, a thorough study of these processes is vital at the LHC.

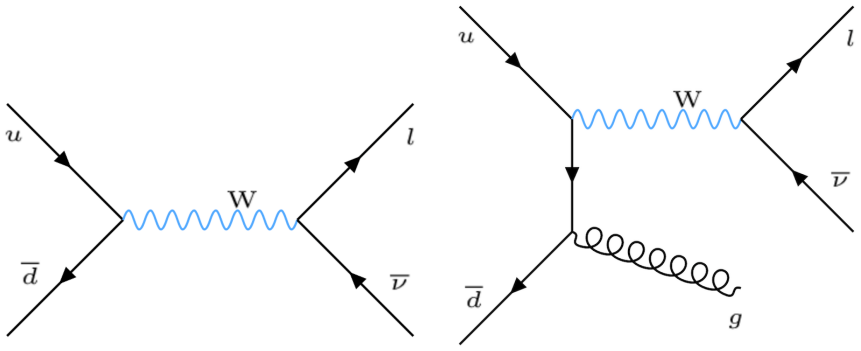


FIGURE 1.3: Leading-order Feynman diagram for the W production process (left), and $O(\alpha_s)$ Feynman diagram with real gluon emission (right).

Previous measurements of W+jets production were performed by the CDF [18] and D0 [19, 20] collaborations in proton-antiproton collisions at $\sqrt{s}$ =1.96 TeV at the Fermilab Tevatron collider. The ATLAS [21] and CMS [22] collaborations measured W+jets production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC, using the data corresponding to integrated luminosities of 4.6 and 5.0 fb^{-1}, respectively. The more recent measurement by the ATLAS collaboration uses the data from proton-proton collisions at $\sqrt{s} = 8$ TeV, and with an integrated luminosity of 20.3 fb^{-1} [23].

1.4.2 Doulbe Higgs Production

The discovery of the Higgs boson (H), the missing piece of the Standard Model of particle physics, by the ATLAS and CMS collaborations in 2012 [24, 25] with the mass $m_H \approx 125$ GeV has confirmed the Brout-Englert-Higgs (BEH) mechanism of electroweak symmetry breaking which explains the origin of mass [14, 15]. After the discovery, the properties of the Higgs boson have also been measured and found to be consistent with the SM. In addition to predicting the existence of the Higgs boson, the BEH mechanism also requires the Higgs field to couple to itself. Therefore, one of the crucial tasks is to understand the Higgs boson self-interactions which could provide insight into the Higgs scalar potential structure. An example of processes involving the self-coupling λ_{HHH} is illustrated in the left graph

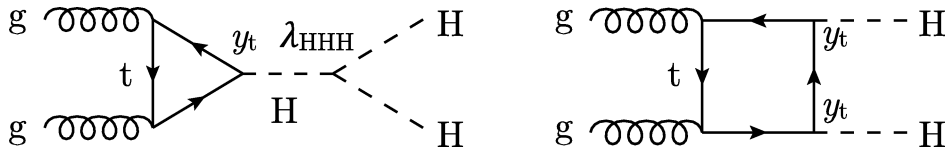


FIGURE 1.4: Feynman diagrams contributing to Higgs boson pair production via gluon fusion at leading order [26].

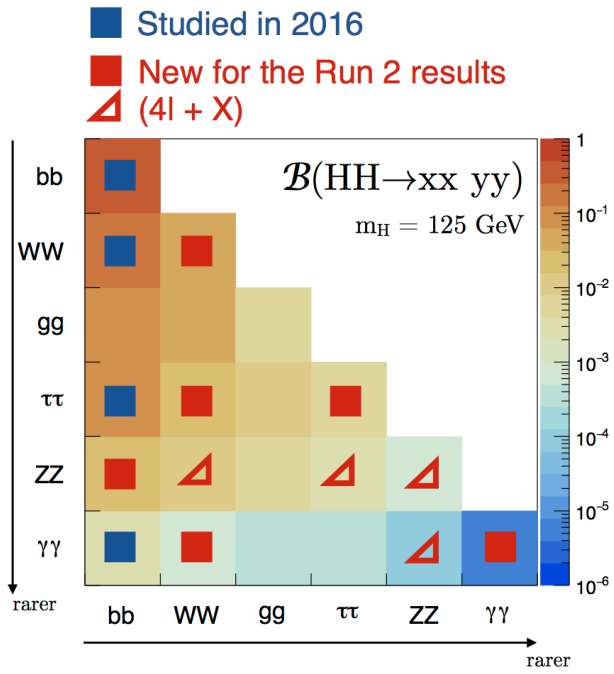
of Fig. 1.4 where a pair of Higgs bosons is generated. Higgs boson pair (HH) production measurement, from which the self-coupling can be extracted, provides an independent and crucial test of the SM and the BEH mechanism.

In the SM, the production of HH from pp collision at the LHC occurs mainly via gluon-gluon fusion and involves two processes for which the diagrams are shown in Fig. 1.4. The production cross section of HH is $\sigma_{HH} = 33.49$ fb at 13 TeV [27] which is too low to be observed with the current dataset from the experiments at the LHC. In spite of this fact, the HH search is still vital because any deviation, for example an increase in the production rate, from the SM prediction would constitute evidence of new physics beyond the SM (BSM).

The discovery of the Higgs boson also opens a new window for the search for BSM using the Higgs particle as a probe. In particular, HH production can be used to search for a new heavy particle with a spin of zero or two and with mass at least $2m_H$ (so that the decay to HH is kinematically possible). Several models predict the existence of heavy resonances that subsequently decay to HH with a large enough cross section to be observed with current LHC data. Such theories are, for example, models with two Higgs doublets (2HDMs) [28], the minimal supersymmetric model (MSSM) [29, 30], and the models with warped extra dimensions (WED) [31].

Various decay modes can be used in a search for HH production. Fig. 1.5 shows the branching ratio of several decay channels [32] as well as that of the $b\bar{b}ZZ$ channel utilized in this thesis.

Previous searches for resonant production of HH have been performed by the ATLAS and CMS collaborations with pp collisions at $\sqrt{s} = 8$ and 13 TeV, in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau\tau$, $b\bar{b}\gamma\gamma$,



HH decay channel	Branching Ratio
$b\bar{b}b\bar{b}$	33%
$b\bar{b}WW^*$	25%
$b\bar{b}\tau\tau$	7.3%
$bbZZ$	3.1%
$WW^*\tau\tau$	2.7%
$ZZWW^*$	1.1%
$b\bar{b}\gamma\gamma$	0.26%
$\gamma\gamma WW^*$	0.10%

FIGURE 1.5: The branching fractions of each decay channel of HH (left), showing also the markers indicating the analysis channels that have been performed (blue) as well as those are being carried out (red) [L. Cadamuro]. The table (on the right) shows the numerical values of the branching fractions for some decay channels.

$b\bar{b}$WW, and $\gamma\gamma$WW final states. Both ATLAS and CMS collaborations also published the combined results of these channels based on both $\sqrt{s} = 8$ TeV [33, 34] and 13 TeV [35, 36] data. The summary of recent CMS Run-II results on the 95% CL upper limits on the resonant HH production cross section is displayed in Fig. 1.6.

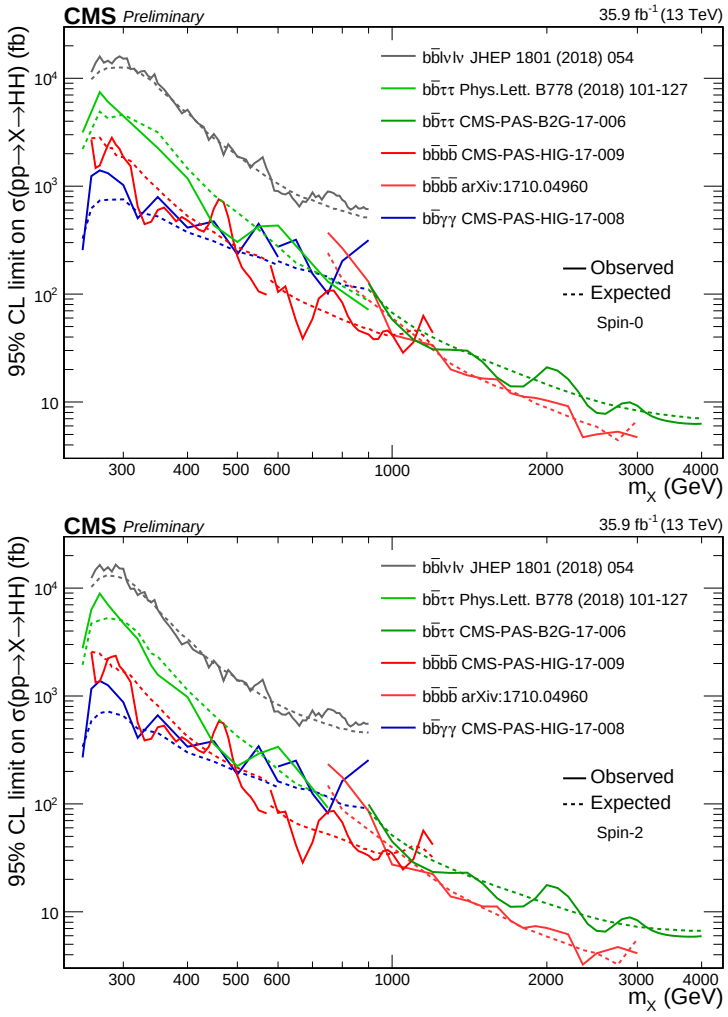


FIGURE 1.6: Observed and expected 95% CL upper limits on the product of cross section and the branching fraction $\sigma(gg \to X) \times B(X \to HH)$ obtained by different analyses assuming spin-0 (top) and spin-2 (bottom) hypotheses [37].

Chapter 2

Experimental Apparatus

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator operated by the European Organization for Nuclear Research (CERN) situated in Geneva, Switzerland. The LHC is housed in a 27 km-long and 3.8 m-wide circular tunnel, located about 100 meters underground on average. The tunnel spans across the border between Switzerland and France. The LHC has two adjacent and parallel beam pipes in which two particle beams are accelerated in opposite directions. The beam pipes have four crossing points around the tunnel where the particle beams are brought to collide. Located at the crossing points are four detectors, ALICE, ATLAS, CMS and LHCb, each designed for certain kinds of research as depicted in Fig. 2.1.

The LHC is primarily a proton-proton (pp) collider with a design center-of-mass energy ($\sqrt{s}$) of 14 TeV (beam energy of 7 TeV) and a nominal luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. It can also be used to accelerate heavy ions, for example lead ions (Pb) to an energy of 1.38 TeV, allowing for the possibility of performing various experiments including pp, pPb, and PbPb collisions as well as fixed-target experiments. The LHC was designed to be a discovery machine which

aimed to find evidence for the Higgs boson and to study energies never reached before. It began colliding protons with $\sqrt{s} = 7$ TeV in 2010 and reached $\sqrt{s} = 8$ TeV and 13 TeV in 2012 and 2015 respectively. The summary of the LHC operation parameters is shown in Table 2.1. Additionally, the tentative plan for future operations, including the High-Luminosity (HL-) LHC project is shown in Fig. 2.2.

TABLE 2.1: The summary of the LHC operating parameters including running (Run) and long shutdown (LS) periods.

LHC Period	Year	$\sqrt{s}$	Integrated Luminosity	Bunch Spacing
Run1 (early)	2010 - 2011	7 TeV	6.14 fb^{-1}	50 ns
Run1 (late)	2012	8 TeV	23.3 fb^{-1}	50 ns
LS1	2013 - 2014	-	-	-
Run2 (early)	2015	13 TeV	4.2 fb^{-1}	50 ns, 25 ns
Run2 (late)	2016 - 2018	13 TeV	158.8 fb^{-1}	25 ns
LS2	2019 - 2020	-	-	-

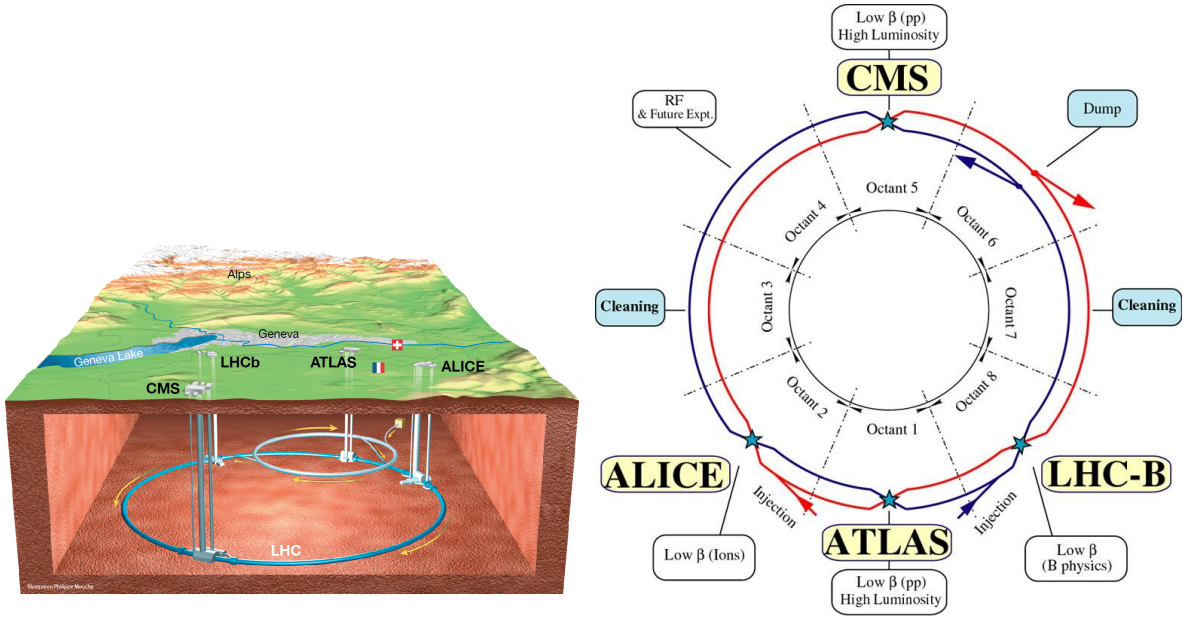


FIGURE 2.1: The left figure shows an overview of the locations of the LHC tunnel and the four main experiments (ALICE, ATLAS, CMS and LHCb) [38]. The right figure is an overview schematic showing the four main experiments and the two ring structure of the LHC [39].

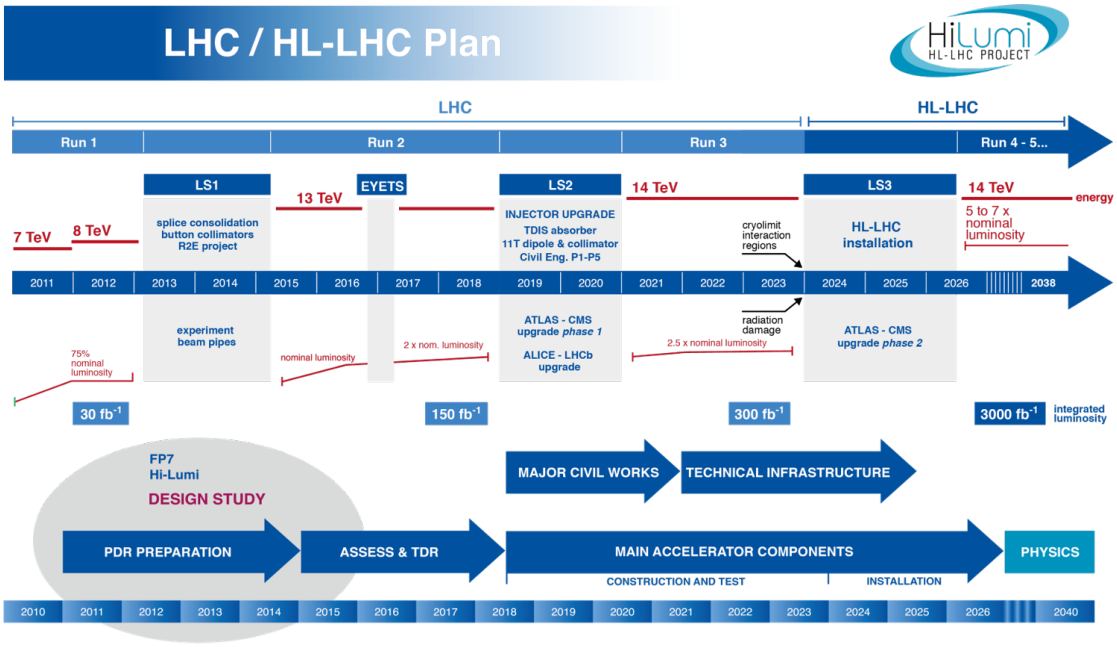


FIGURE 2.2: LHC/HL-LHC planned schedule [40].

2.1.1 The LHC Components

The LHC is composed of eight arcs and eight linear sections (also called "insertions") alternately placed, hence it is not a perfect circle. The arcs contain dipole magnets required to bend the protons trajectories. In total 1,232 dipole magnets (154 in each arc) made of Niobium-Titanium (NbTi) coils generate a magnetic field of around 8.33 T from currents of about 11.7 kA, resulting in an effective bending radius of 2,804 m. The cross-sectional diagram of an LHC dipole magnet is shown in Fig. 2.3.

The linear sections contain additional sets of special 392 quadrupole magnets, which are responsible for focusing the beams down to the smallest possible size. Each straight section has a different layout depending on the specific use of the section: beam collisions within an experiment, beam injection, beam dumping, or beam cleaning. One of the linear sections contains accelerating radio-frequency systems (RF cavities) that are used to accelerate the particle beams to their nominal energy as well as to maintain their speed by compensating

for their energy losses. The diagram on the right of Fig. 2.1 shows where these linear sections are positioned around the LHC ring, together with their specific uses.

Both dipole and quadrupole magnets, 15 m and 3 m long respectively, are superconducting magnets maintained at their operating temperature of 1.9 K (-271.25 °C). This temperature is achieved by using superfluid helium for the cryogenic system.

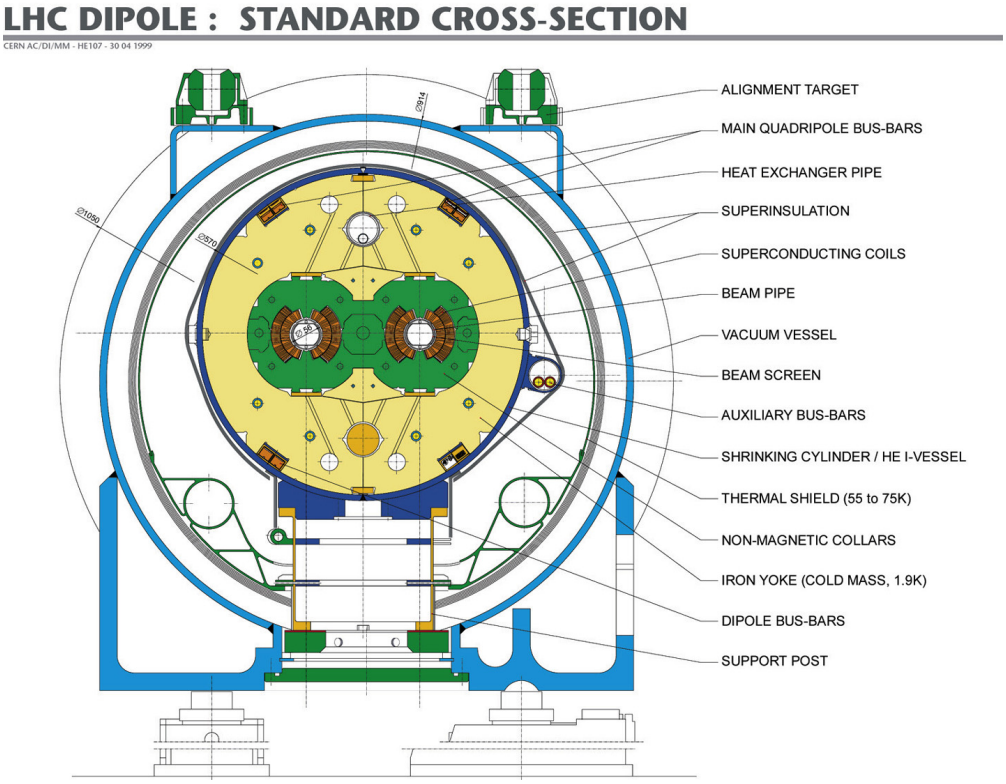


FIGURE 2.3: Diagram showing the cross-section of an LHC dipole magnet [41].

2.1.2 The LHC Beams and Luminosities

Each proton beam nominally consists of numerous packets of protons called bunches. Each bunch contains on the order of 10^{11} protons. At the design operation, bunches are spaced 25 ns apart and 2,808 bunches are filled in each beam. A bunch of protons is about 30 cm in length with a transverse size of the order of a mm in the LHC pipes, but can be squeezed as small as 16 μm at the collision points.

The collision rate of protons at a particle accelerator is quantified by the instantaneous luminosity (L_{inst}) which is a measure of the *beam brightness*. The instantaneous luminosity is defined in terms of the beams parameters as:

$$L_{inst} = \frac{N_1 N_2 n_p f}{4\pi\sigma_x\sigma_y} F = \frac{N_1 N_2 n_p f \gamma}{4\pi\epsilon_n \beta^*} F \tag{2.1}$$

where N_1 and N_2 are the numbers of proton bunches in each beam, n_p is the number of protons contained in each bunch, f is the beams crossing frequency, and σ_x and σ_y are the beam transverse size in the horizontal (x) and vertical (y) planes. When two beams of protons are brought to collide, the beams are not perfectly aligned but cross with a finite beam crossing angle ϕ. The luminosity reduction factor F is introduced to account for the loss of using a beam crossing angle with respect to head-on collisions.

The product $\sigma_x\sigma_y$ can be expressed in terms of the Lorentz factor γ, the beam normalized transverse emittance ϵ_n and the beam parameter β^*. The beam emittance ϵ_n measures the spread of the beam particles in the position-momentum phase space. The beam parameter β^* is the distance between the point where the beams cross and the point where the transverse size of the beams is twice as large as the transverse size at collision point.

In principle, an increase in the instantaneous luminosity, hence a higher rate of interactions, can be achieved by adjusting these beam parameters, e.g. increasing f, n_p, N_1, and N_2, and focusing (squeezing) the beams by reducing ϵ_n and β^*. The values of these parameters at the design operating conditions are listed in Table 2.2.

The instantaneous luminosity can also be expressed as

$$L_{inst}(t) = \frac{1}{\sigma_p}\frac{dN(t)}{dt} \tag{2.2}$$

where $N(t)$ is the number of events produced at a certain time t, and σ_p is the proton-proton interaction cross section. In this sense, it is clear that L_{inst} has a dimension of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The integrated luminosity L can then be calculated as $\mathrm{L} = \int L(t)dt$. The total number of events produced in the collisions over a period of time is then related to the integrated luminosity by $N = \sigma L$. In this sense, the number of observed events for a physics process N_{obs} is related to the cross section of the physics process σ and the integrated luminosity L with the proper accounting of detector acceptance A (fraction of the events that falls within the detector geometrical coverage) and the efficiency of detection ϵ as follows:

$$\sigma = \frac{N_{obs}}{A \times \epsilon \times L}. \tag{2.3}$$

TABLE 2.2: Design parameters of the Large Hadron Collider.

Parameters	Design Value
Circumference Length	26659 m
Depth	50 - 175 m
Bunch spacing	25 ns
Bunch Crossing Rate f	40.08 MHz
Number of bunches per proton beam	2808
Number of protons per bunch	1.1×10^{11}
Crossing angle ϕ	285 μrad
Normalized transverse emittance ϵ_n	3.75 μm
β^*	0.55 m
Nominal proton energy	7 TeV
Center-of-mass energy $\sqrt{s}$	14 TeV
Design instantaneous luminosity	10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Average number of collisions per crossing	20

2.1.2.1 Pile Up

At the LHC, high density proton bunches are collided. The probability to have more than one pp collision increases with increasing luminosity. Indeed, many pp interactions occur during a single bunch crossing (approximately 21 interactions on average in the 2012 run

period). Theses additional collisions in the same or adjacent bunch crossings are called "pileup", PU.

Simulations used in an analysis have to account for pileup in order to properly mimic the environment of real collisions. In the simulations, this is achieved by adding soft inelastic collisions to a hard process based on the average number of interactions occurring in the LHC. However, the average number of interactions is dependent on the accelerator conditions during the data taking, and hence varies from run to run. MC samples are generally produced with a well-defined number of interactions distribution derived with the best projection of the LHC condition. Therefore, an additional reweighting of the simulated events, on a run by run basis, has to be performed such that pileup distribution in the simulation matches the distribution in the observed data.

2.1.3 The CERN Accelerator Complex

Protons are injected through a chain of accelerators in the CERN accelerator complex, as shown in Fig. 2.4, so that the energy of the proton beams is progressively increased before delivery to the LHC. First, protons are obtained by ionizing atoms of hydrogen gas using an electric field, and they are passed to a linear accelerator section, Linac2. After being accelerated in Linac2 to reach the energy of 50 MeV, the protons are injected into the Proton Synchrotron Booster (PSB). The Booster boosts them to 1.4 GeV and feeds them to the Proton Synchrotron (PS), where they reach the energy of 25 GeV. The protons are then sent to the Super Proton Synchrotron (SPS), where they are further accelerated to 450 GeV. Then they are finally injected into the LHC where they are split into two beams with a clockwise and an anti-clockwise direction. This injection process takes about 9 hours, after which the beams are accelerated, for 20 minutes, until the final beam energy is achieved. The counter-rotating beams are then brought together to collide at four interaction points where the detectors are placed. The beams circulate for many hours inside the LHC beam pipes

under normal operating conditions, until the remaining number of protons is low. Then, the beams are dumped and a new cycle of these processes restarts.

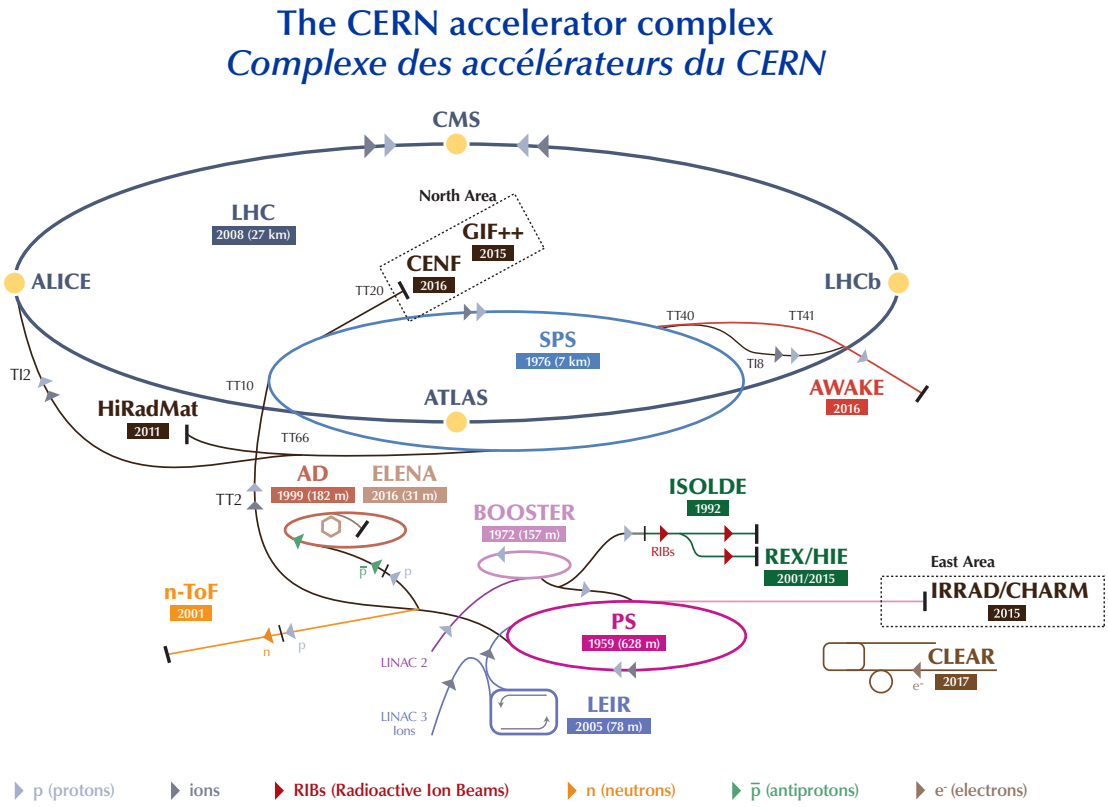


FIGURE 2.4: The CERN accelerator complex [42].

2.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general-purpose particle detector that is designed to measure and identify a wide range of particles and phenomena produced in the collisions at the LHC. It is constructed as a cylindrical onion, where different layers of detectors measure different particles (see Fig. 2.5). As discussed in Sec. 2.1, bunches of protons are

collided. In each of the collisions, most of the protons are just passing through or glancing each other. Only some of the protons collide head-on and produce very energetic collisions. Identifying the type and properties of the particles emerging from the collision is essential for reconstructing the underlying process of interaction.

The CMS detector consists of an inner tracking system and electromagnetic (ECAL) and hadron (HCAL) calorimeters surrounded by a 3.8 T superconducting solenoid. The inner tracking system consists of a silicon pixel and a strip tracker which reconstructs the position of charged particles in successive layers. This allows for the precise reconstruction of the trajectories of charged particles, and vertices of origin. The crystal ECAL and the brass and scintillator HCAL are used to measure the energies of photons, electrons, and hadrons. The three muon detector systems surrounding the solenoid are used to identify and to further track muons. Muon detectors are composed of drift tubes in the barrel region, cathode strip chambers in the endcaps, and resistive plate chambers in both the barrel region and the endcaps. Fig. 2.6 illustrates how the fundamental particles, e.g. photons, electrons, charged hadrons, neutral hadrons, and muons, emerging from the collision of partons within the protons are detected.

The CMS experiment uses a right-handed coordinate system, with the origin at the nominal collision point, the x-axis pointing to the center of the LHC ring, the y-axis pointing up (perpendicular to the LHC plane), and the z-axis pointing along the anticlockwise beam direction. The polar angle (θ) is measured from the positive z-axis and the azimuthal angle (ϕ) is measured from the positive x-axis in the x-y plane. The radius (r) is the perpendicular distance from the z-axis and the pseudorapidity (η) is defined as $\eta = -\ln[\tan(\theta/2)]$. The coordinate system is illustrated in Fig. 2.7.

As mentioned in Chapter 1, since the colliding protons are composite objects having partons (gluons, quarks, antiquarks) as substructures, the hard scatter is a scattering between partons. Therefore, the z component of the momentum of the colliding partons is not known

and the conservation of momentum in the z-direction cannot be used. However, the conservation of momentum in the transverse plane can always be applied, since the collisions are always along the z axis. Therefore, it is natural that the observables in collider experiments are measured in the transverse plane, for example p_{T}, and $E_{\mathrm{T}}^{\mathrm{miss}}$.

A brief description of each of the CMS subdetector systems is given in Sec. 2.2.1-2.2.5. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [43].

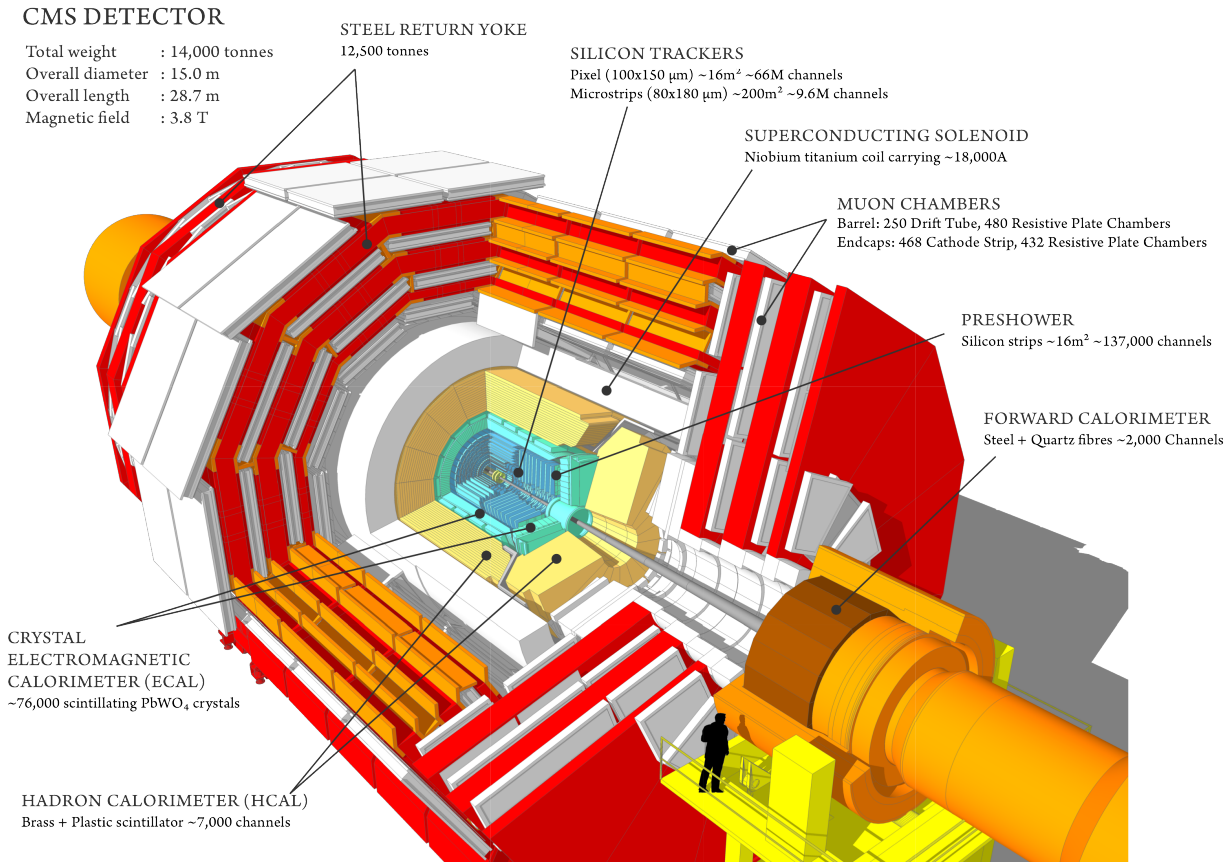


FIGURE 2.5: A cutaway image of the CMS detector showing different layers of the detector [44].

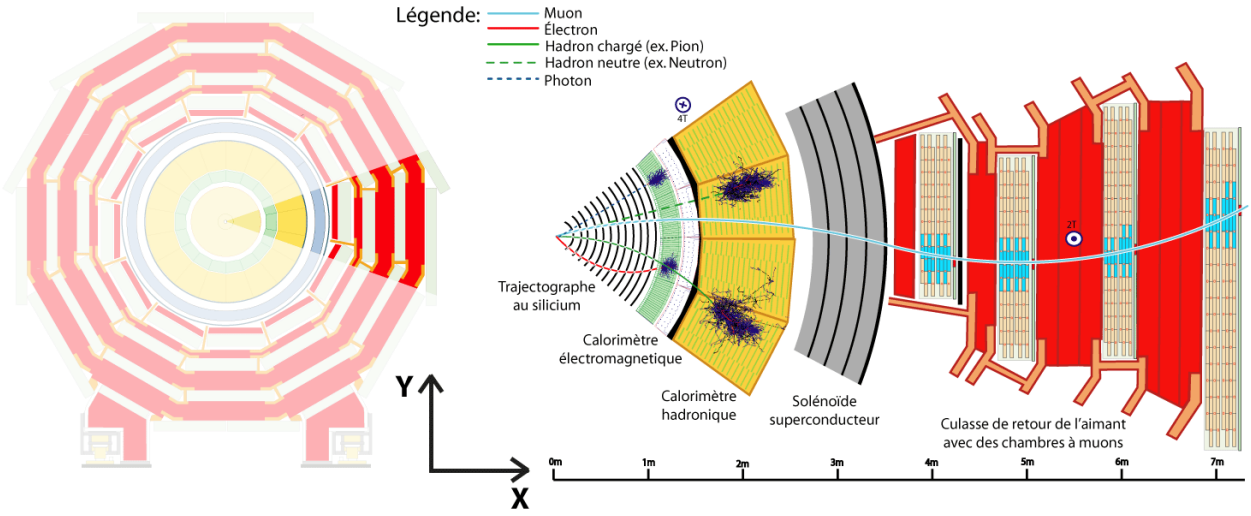


FIGURE 2.6: A transverse cross-section of the CMS detector illustrating how different particles interact with various sections of the detector [44].

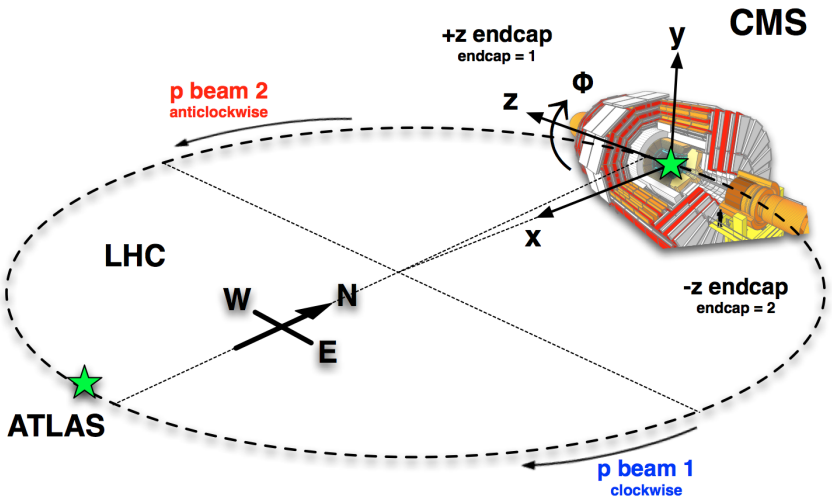


FIGURE 2.7: The global coordinate system of CMS [T. Cox].

2.2.1 Superconducting Solenoid Magnet

The momentum and charge of particles generated in LHC collisions are measured based on the bending of particles' trajectories when traversing a nearly uniform magnetic field. The relation between the transverse momentum p_T of a charged particle and its radius of curvature R can be written as:

$$p_T \propto BR \tag{2.4}$$

, where B is the magnetic field strength.

The higher the p_{T} of a particle is, the stronger the magnetic field is needed to bend its trajectory, therefore, the CMS detector needs a very high magnetic filed to measure particles with large momentum. Such a strong magnetic field is generated by the superconducting solenoid magnet made of coils of niobium-titanium wires, operated at a temperature of 4.5 K. The dimensions of the solenoid are 12.5 m in length and 6 m in diameter, surrounding the inner tracker and the calorimeter systems as depicted in Fig 2.5. It is designed to deliver a magnetic field strength of 4 T. However, for longevity, the operating value is set to 3.8 T.

Even with such an intense magnetic field, the trajectories of particles with high p_{T} ($\sim$100 GeV) are almost straight with respect to the scale of the CMS detector. This field strength is indeed necessary to determine the electric charge of muons with $p_{\mathrm{T}} > 1$ TeV. In contrast, particles with $p_{\mathrm{T}} < 0.75$ GeV are so bent that they do not reach the electromagnetic calorimeter whose inner radius is $\sim$1.3 m.

The magnetic flux is returned through an iron yoke composed of five wheels, and two endcaps which are each composed of three disks. The iron yoke also serves as the structure hosting the muon system for which the return field in the yoke provides the bending field.

2.2.2 Inner Tracking System

The function of the inner tracking system or "tracker" is to precisely and efficiently measure the trajectories of charged particles emerging from the LHC collisions, as well as precisely reconstruct the (secondary) vertices of decaying particles. It is the detector system situated closest to the beam pipe, surrounding the interaction point. Overall, it has the shape of

a cylinder with a length of 5.8 m and a diameter of 2.5 m. The CMS tracker uses silicon detector technology and consists of a pixel detector and a silicon strip tracker. Both of them are further divided into barrel and endcaps parts. The magnetic field in the tracker volume provided by the CMS solenoid is homogeneous with the value of 3.8 T. A schematic drawing of the CMS tracker is shown in Fig. 2.8.

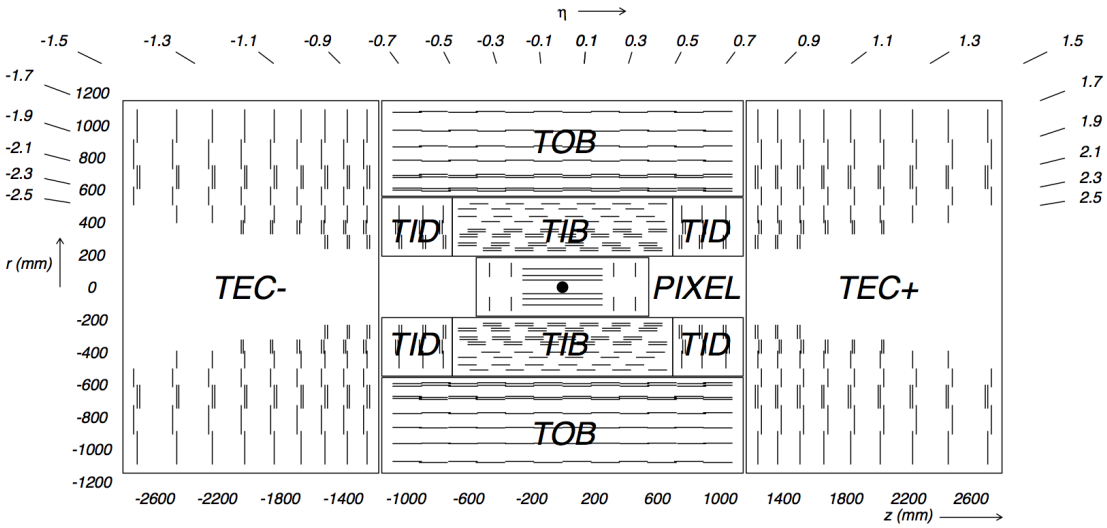


FIGURE 2.8: Schematic drawing showing the longitudinal cross section of the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits [43].

The pixel detector system consists of three barrel layers (BPix) with two forward pixel detectors (FPix). The BPix are three 53-cm-long cylindrical layers of hybrid pixel detector modules surrounding the interaction point at radii of 4.4, 7.3 and 10.2 cm. The FPix, also referred to as "pixel endcap disks", consists of four disks of pixel modules extending from about 6 to 15 cm in radius from the beam axis and located on each side at $z = \pm 34.5$ and $z = \pm 46.5$ cm. Collectively, the pixel detector covers a pseudorapidity of $|\eta| < 2.5$.

The silicon strip tracker consists of different subsystems including the Tracker Inner Barrel and Disks (TIB/TID), the Tracker Outer Barrel (TOB), and the Tracker EndCaps (TECs), which span the region from 20 cm to 116 cm in radius, as measured from the beam axis and the z range of $|z| < 282$ cm. All subsystems use silicon micro-strip sensors with their strips

parallel to the beam axis in the barrel, and in the radial direction on the disks. TIB and TOB have 4 and 6 barrel layers respectively. Each TID has 3 disks, each one consisting of 3 rings while each TEC is composed of 9 disks, carrying up to 7 rings.

The modules in the first two layers of the TIB and TOB, the first two rings of the TID, and rings 1, 2, and 5 of the TECs incorporate a second micro-strip detector module which is mounted back-to-back with a stereo angle of 100 mrad. The stereo layers provide a measurement of the second co-ordinate (z in the barrel and r on the disks).

2.2.3 Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) is designed to measure the energy of particles primarily interacting via electromagnetic interactions. In particular, its function is to identify electrons and photons and to measure their energies. The ECAL is a homogeneous calorimeter, meaning that its entire volume is sensitive and it is used to measure the energy deposits. It has a cylindrical geometry consisting of an ECAL Barrel (EB) containing 61,200 scintillating lead tungstate (PbWO4) crystals and two ECAL Endcaps (EE) containing 7,324 PbWO4 crystals each. The PbWO4 crystals possess appropriate properties for operating at LHC, such as high density (8.28g/cm^3), short radiation length (0.89 cm) and being optically clear, providing a fine granularity, fast and compact calorimeter. Avalanche photodiodes (APDs) are used as photodetectors in the barrel while vacuum phototriodes (VPTs) are used in the endcaps.

The location and the layout of the CMS ECAL is shown in Fig. 2.5 and Fig. 2.9 respectively. The EB covers the pseudorapidity range of $|\eta| < 1.479$. The granularity of the EE is 360-fold in ϕ and (2×85)-fold in η. The EB front face is at a radius of 1.29 m from the beam axis and its total length in the z-direction is 6 m. The EEs cover the range $1.479 < |\eta| < 3.0$ and is located at $z = 315.4$ cm.

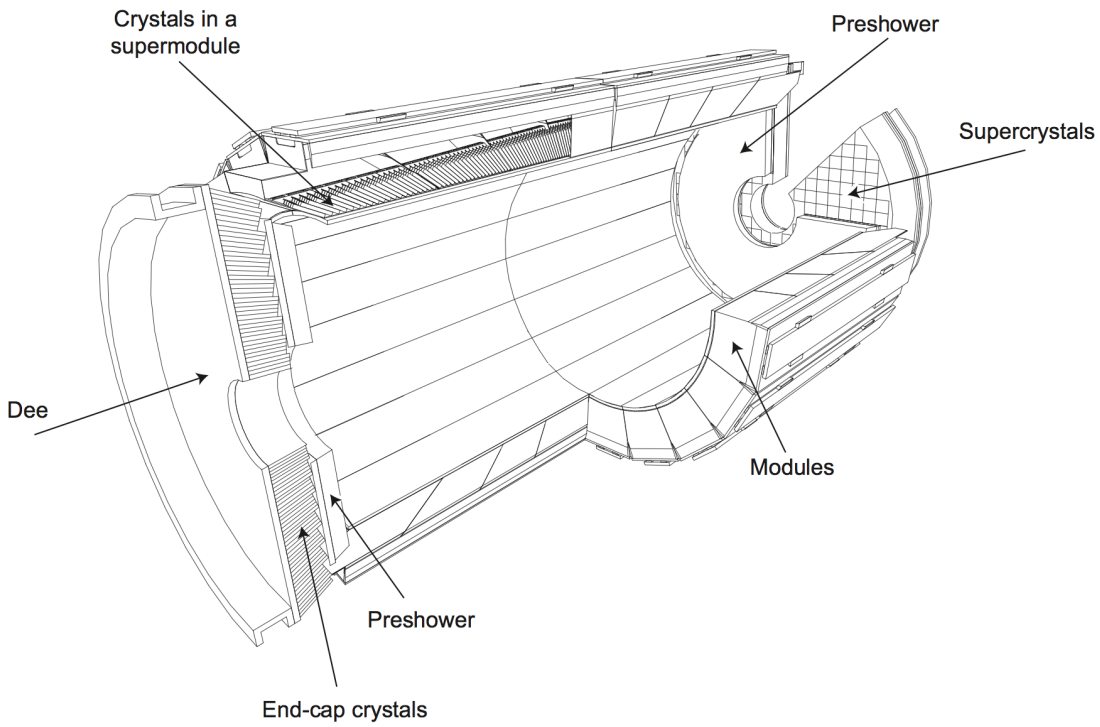


FIGURE 2.9: Schematic drawing of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front [43].

The EB crystals have a front face and rear face area of 22×22 mm^2 and 26×26 mm^2 respectively. The crystal length is 230 mm. The crystals are contained in a thin-walled alveolar structure called submodule. The submodules are put together into modules of different types, depending on the position in η. Each module contains 400 or 500 crystals. A supermodule is obtained by assembling four modules, each separated by aluminum conical webs with 4-mm thickness. Each supermodule contains 1,700 crystals.

The crystals in the EEs are identical and are grouped in units of 5×5 crystals called supercrystals (SCs). Each endcap is divided into 2 halves, or "Dees", each holding 3,662 crystals. Each crystal has a front face and rear face area of 28.62×28.62 mm^2 and 30×30 mm^2 respectively, and a length of 220 mm.

In addition, endcaps preshower detectors (ES) are placed in front of the EEs. The ES is a two-layer sampling calorimeter. It uses lead radiator layers to initiate the electromagnetic

showers from photons and electrons interaction. The deposited energy and shower profiles are measured by silicon strip sensors placed behind the radiators. The total thickness of the ES is 20 cm, and it covers a pseudorapidity range of $1.653 < \eta < 2.6$. Its main function is the identification of neutral pions through $\pi^0 \to \gamma\gamma$ decays.

2.2.4 Hadronic Calorimeter

The CMS hadronic calorimeter (HCAL) measures the energy and direction of hadrons and, hence, provides the measurement of jets and missing transverse energy. The HCAL consists of 4 subdetectors: the hadronic barrel (HB), the hadronic outer (HO), the hadronic endcap (HE), and the hadronic forward (HF). The location of these subdetectors within the CMS detector can be seen in Fig. 2.5 and Fig. 2.11. Collectively, HCAL covers the region of $|\eta| < 5$.

The HB and HE are sampling calorimeters that use brass as the absorbing material and plastic scintillator tiles as the active material. HB is 9 m in length and spans the region between the ECAL and the solenoid magnet (177 m$< r <$ 295 m) covering the pseudorapidity range $|\eta| < 1.3$ with granularity $\Delta\eta \times \Delta\phi = 0.087{\times}0.087$. The HE covers $1.3 < |\eta| < 3.0$. The granularity of the HE varies from $\Delta\eta \times \Delta\phi = 0.087{\times}0.087$ at $|\eta| = 1.3$ to $\Delta\eta \times \Delta\phi = 0.35{\times}0.174$ at $|\eta| = 3.0$.

The HO is placed outside and surrounds the solenoid magnet to cover the high energy hadron showers that pass through HB. It is mounted along the inner part of the magnetic flux return yoke as shown in Fig. 2.10. It covers the pseudorapidity range $|\eta| < 1.26$. In the central region ($|\eta| < 0.33$), it consists of two plastic scintillator layers separated by a layer of 18 cm-thick iron absorber while in the rest of the η range a single scintillator layer is placed behind the iron absorber. The HO's granularity matches that of the HB.

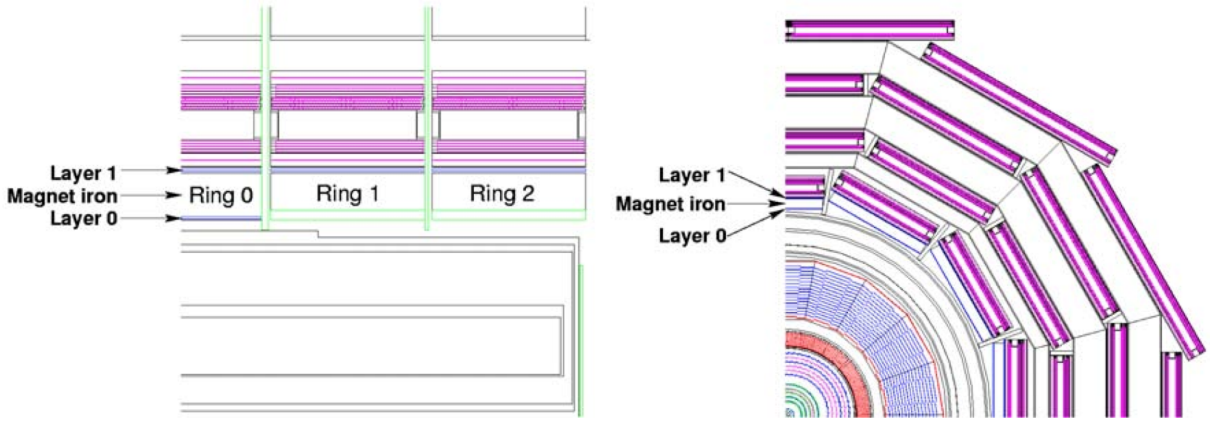


FIGURE 2.10: Diagrams showing the longitudinal (left) and transverse (right) views of HO layers [43].

The HF extends to the pseudorapidity range $3.0 < |\eta| < 5.0$ where there is a high particle flux. Therefore, the active material is made of quartz fibre which is known to be radiation resistant. It is placed in front of the CMS muon system, at a distance of 11.2 m from the interaction point. Its granularity varies from $\Delta\eta \times \Delta\phi = 0.111 \times 0.174$ at $|\eta| \approx 3$ to $\Delta\eta \times \Delta\phi = 0.302 \times 0.348$ at $|\eta| \approx 5$.

2.2.5 Muon System

The main purpose of the muon system is to provide robust muon identification and precise muon momentum measurement covering $|\eta| < 2.4$. The system consists of three types of gas ionization chambers including 250 drift tubes (DTs), 540 cathode strip chambers (CSCs), 1056 resistive plate chambers (RPCs) arranged in layers interleaved with steel flux-return yokes. The return yokes serve as a confinement of the magnetic field, a hadron absorber, and a support for the muon chambers. A schematic of the CMS muon system, illustrating the major components is shown in Fig. 2.11

The DTs are rectangular drift cells that measure the positions of muons by determining the drift time of the induced charges through an electric field. They are arranged as 4 concentric

layers or stations around the beam axis and in 5 wheels along the beam axis. This is called the muon barrel (MB) system, and it spans a range of $|\eta| < 1.2$. The inner three DT layers have 60 drift chambers each, and the outer layer has 70. The drift cells technology is suitable in this region because of the uniform magnetic field, low background flux induced by neutrons as well as low muon rate in the barrel region.

The muon endcap (ME) systems are equipped with cathode strip chambers (CSCs). One ME is placed on the +z direction and the other one on the -z direction with respect to the interaction point (IP). The MEs cover $0.9 < |\eta| < 2.4$, the region where the magnetic field is non-uniform, the background levels are high, and the muon rates are also high. On each ME, CSCs are arranged in 4 stations along the the beam axis and two or three rings along the r axis. The first stations (closest to the IP) have three rings while the other stations have two rings.

The RPCs are added to both the MB and MEs, covering $|\eta| < 1.9$. They are gaseous parallel-plate detectors. They have fast response, providing good time resolution and are used mainly as trigger detectors. There are 6 layers of RPCs embedded in the barrel iron yoke and 4 layers in each ME.

2.2.5.1 Cathode Strip Chambers

The CSCs are multi-wire gaseous proportional counters with a finely segmented cathode strip readout. Each CSC chamber is trapezoidal in shape and contains six anode wire planes alternated with seven trapezoidal panels as depicted in Fig 2.12. The anode wires are arranged azimuthally and provide a precise measurement of the muon position in radial (r) space or η space. The cathode strips with a constant $\Delta\phi$ width are arranged radially and are approximately perpendicular to the anode wires, and hence give a precise measurement of the muon position in ϕ space. When a muon passes through the gas volume, atoms of the gas are ionized, leaving electrons which are drifted by the electric field to the anode wires,

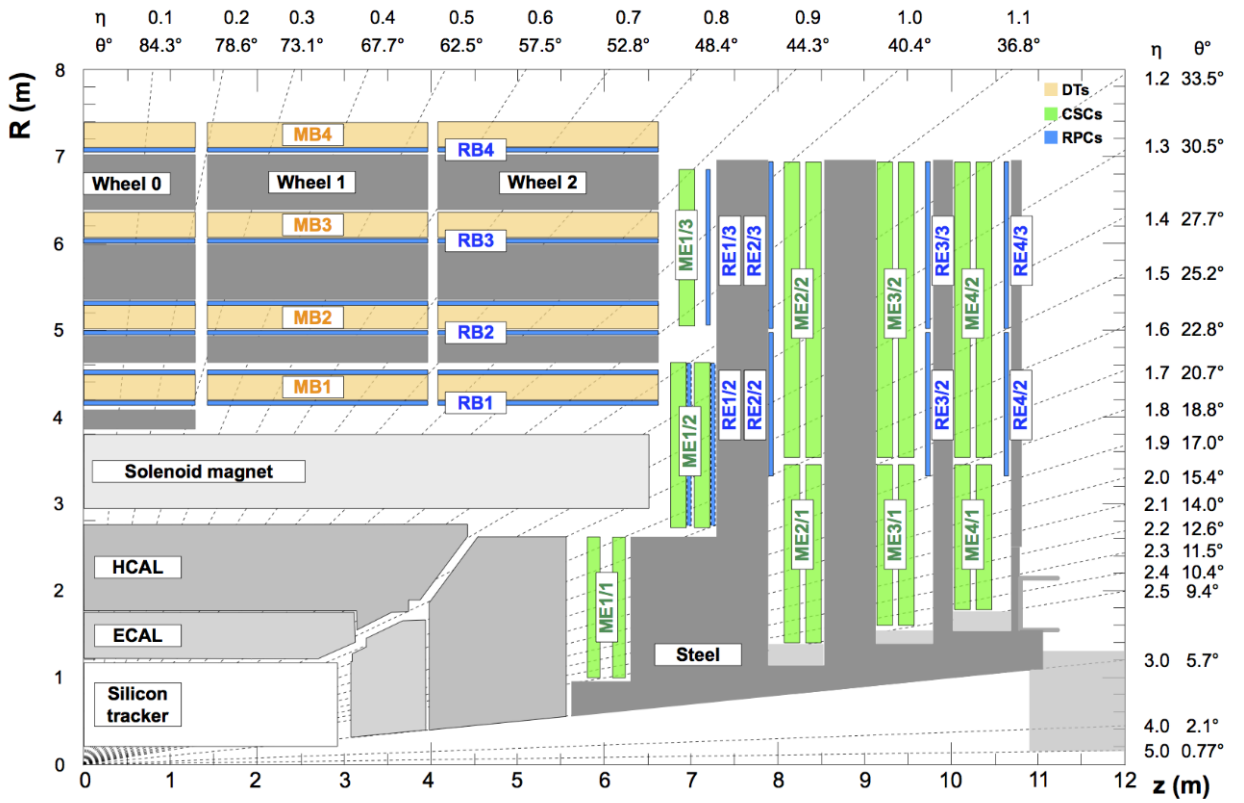


FIGURE 2.11: An r-z cross section of a quadrant of the CMS detector illustrating the locations of the various muon stations and the steel flux-return disks constituting the CMS muon system. The DTs are labeled MB ("Muon Barrel") and the CSCs are labeled ME ("Muon Endcap"). The RPCs are mounted in both the barrel and endcaps of CMS, where they are labeled RB and RE, respectively [45].

causing an avalanche. The ions, on the other hand, drift to the cathode strip, and induce a charge pulse. These give the ability to measure the position in the r-ϕ plane at which the muon crosses the chamber.

The seven trapezoidal panels with a thickness of 16.2 mm each, shown in Fig. 2.13, when stacked together, form six 9.5 mm gas gaps or layers. The panels are coated with 1.6-mm-thick FR4 skins on each side. The FR4 skins are clad with 36-μm-thick copper on their outer surfaces. A pattern of 80 strips is milled on each side of the trapezoidal panels (panels 1, 3, 5, 7 in Fig. 2.13) with a 0.5 mm gap between strips. The panels (2, 4, 6), called *anode panels*, have anode wires wound around them with a wire spacing of approximately 3.2 mm. The wires are made of gold-plated tungsten with a diameter of 50 μm. The nominal gas

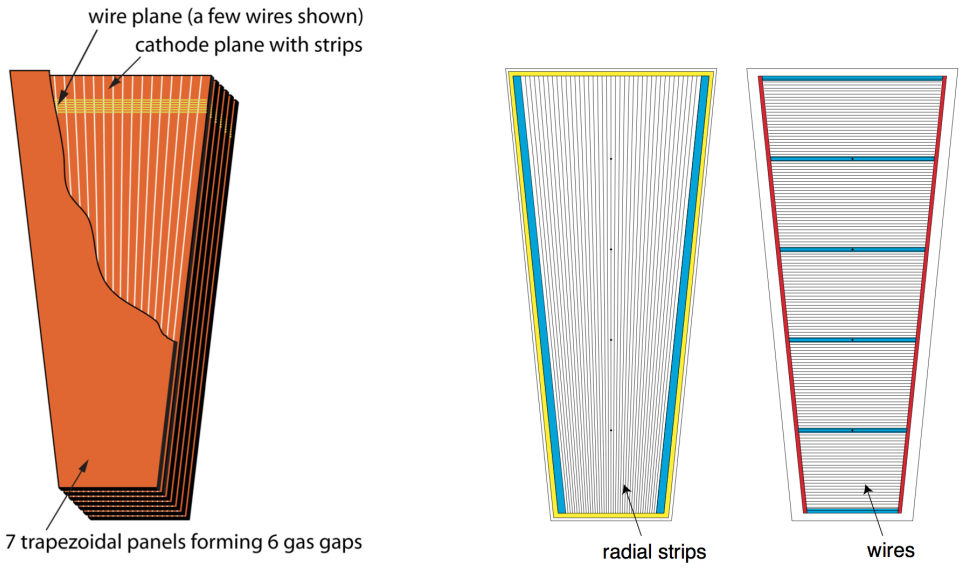


FIGURE 2.12: Layout of a CSC chamber showing orientation of the strips (radial) and wires (perpendicular to the central strip), and 6 layers of the chamber stacked in the direction perpendicular to the strip-wire plane (left), and finer detail of the strip-wire plane in a single layer of a CSC (right) [46].

mixture used is 40%Ar + 50%CO_2 + 10%CF_4. The nominal operating HV for the chambers is 3.6 kV, which corresponds to a gas gain of 7×10^4.

The concentric rings of chambers assembled between disks of the steel flux-return yoke at approximately the same value of z, form a CSC station (see Fig. 2.11). There are four CSC stations (S1-S4) in each endcap. The chambers in these stations are labeled as ME1-ME4 and they contain a total of 540 CSCs as shown in Table 2.3.

TABLE 2.3: Summary of the number of CSC chambers per endcap muon system. There are 2 rings (R1,R2) in each station (S2-S4) except for station 1 (S1) where there are 3 rings. These constitute a total of 540 CSCs in the entire endcap muon system.

CSCs per endcap	S1	S2	S3	S4	All
R1	36	18	18	18	90
R2	36	36	36	36	144
R3	36	-	-	-	36
In station	108	54	54	54	270

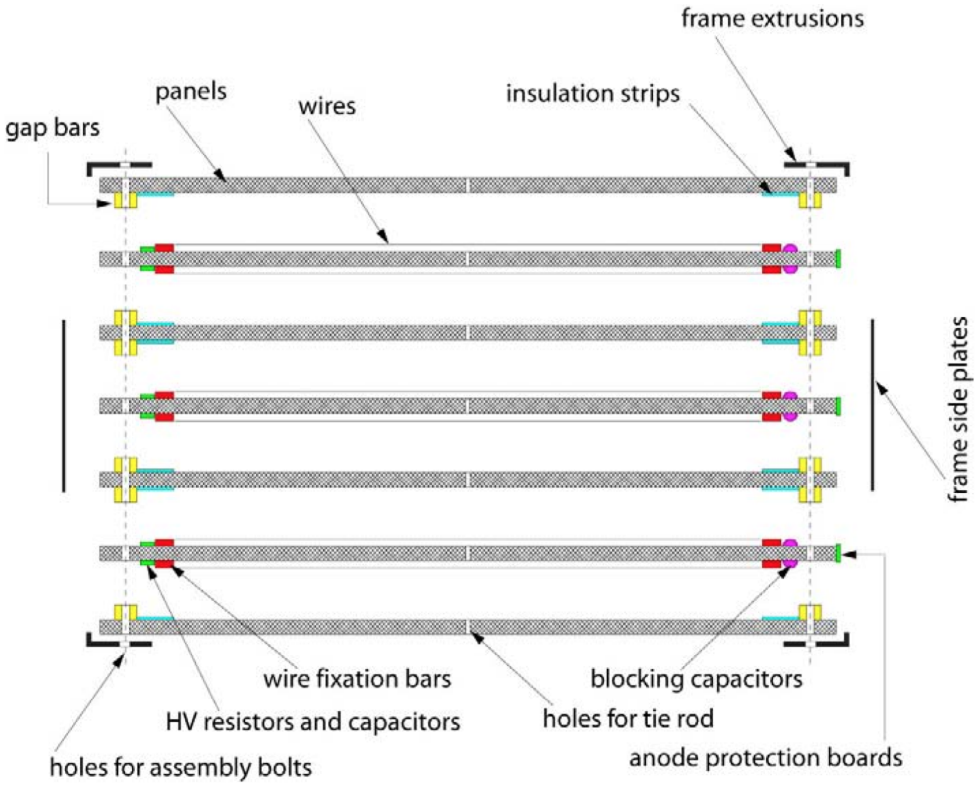


FIGURE 2.13: Mechanical design of the CMS cathode strip chambers [43].

TABLE 2.4: Summary of the number of strip channels per layer of a CSC and per endcap muon system.

(A) Strip channels per layer.

Strip Channels	S1	S2	S3	S4
R1	64	80	80	80
R2	80	80	80	80
R3	64	-	-	-
R4 = ME1/1A	48	-	-	-

(B) Strip channels per endcap.

Strip Channels	S1	S2	S3	S4	All
R1	13824	8640	8640	8640	
R2	17280	17280	17280	17280	
R3	13824	-	-	-	
R4 = ME1/1A	10368	-	-	-	
In station	55296	25920	25920	25920	133056

TABLE 2.5: Summary of the number of wire channels per layer of a CSC and per endcap muon system.

(A) Wire channels per layer.

Wire Channels	S1	S2	S3	S4
R1	48	112	96	96
R2	64	64	64	64
R3	32	-	-	-

(B) Wire channels per endcap.

Wire Channels	S1	S2	S3	S4	All
R1	10368	12096	10368	10368	
R2	13824	13824	13824	13824	
R3	6912	-	-	-	
R4 = ME1/1A	10368	-	-	-	
In station	31104	25920	24192	24192	105408

2.2.6 Calibration Constants for CSC

The CSC system consists of 540 chambers with over 260k cathode (strip) channels and 210k anode (wiregroup) channels readout. The readout of the front-end cathode (strip) electronics of the CSC system is calibrated and monitored continually over time through dedicated calibration runs taken between periods of data taking. In addition to the calibration, the functioning of CSC electronic components (including chambers, strips, and wires) is also constantly monitored. The electronics calibration involves a series of tests and measurements [47] the results of which, called "calibration constants", are values characterizing the performance of the CSC detectors and electronics.

The time-dependent (or run-dependent) calibration constants are stored in a central database as CMS's "conditions data" and can be accessed as required by the CMS software. Conditions data can be any run-dependent values needed by the CMS event reconstruction software. For example, the bad strips, bad wires, and bad chambers objects which identify known nonfunctioning components needing special treatment in either reconstruction or simulation

are included in conditions data. During CSC local reconstruction and simulation, the appropriate set of values from the conditions data is applied to ensure the best performance of the CSCs and their readout electronics.

The electronics calibrations include parameters that affect the detector response: the strip-to-strip crosstalk, the strip channel noise, the strip pedestals, and the strip channel electronic gains.

Calibration constants are updated for use in reconstruction when significant improvement in the strip position resolution are observed in reconstruction tests. The reconstruction is most sensitive to changes in the gains and pedestals, and relatively insensitive to changes in the crosstalk calibration constants. During LS1, many CSCs components were upgraded and the new calibration constants were re-derived and updated.

The difference in the values of the gains for each of the 260k strip channels, before and after the LS1, is given in Fig. 2.14 where clusters of values localized around a given strip index value having similar values of relative change indicate correlation between changes on a given front-end board. A significant relative changes of the gains in ME1/1A and ME1/1B are expected because the front end electronics were changed (replacing CFEBs with DCFEBs and unganging ME1/1A readout) during LS1. For ME4/2, the chambers were newly installed and the old calibration constants are just dummy values, hence significant relative changes are also observed.

2.2.7 Triggers

The nominal operation of the LHC has a bunch crossing frequency of 40 MHz (bunch spacing of 25 ns) and an instantaneous luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. With these conditions, about one billion proton-proton interactions per second (see Fig 2.15), or equivalently, approximately 20 simultaneous pp collisions per bunch crossing will occur. It is impossible to store and

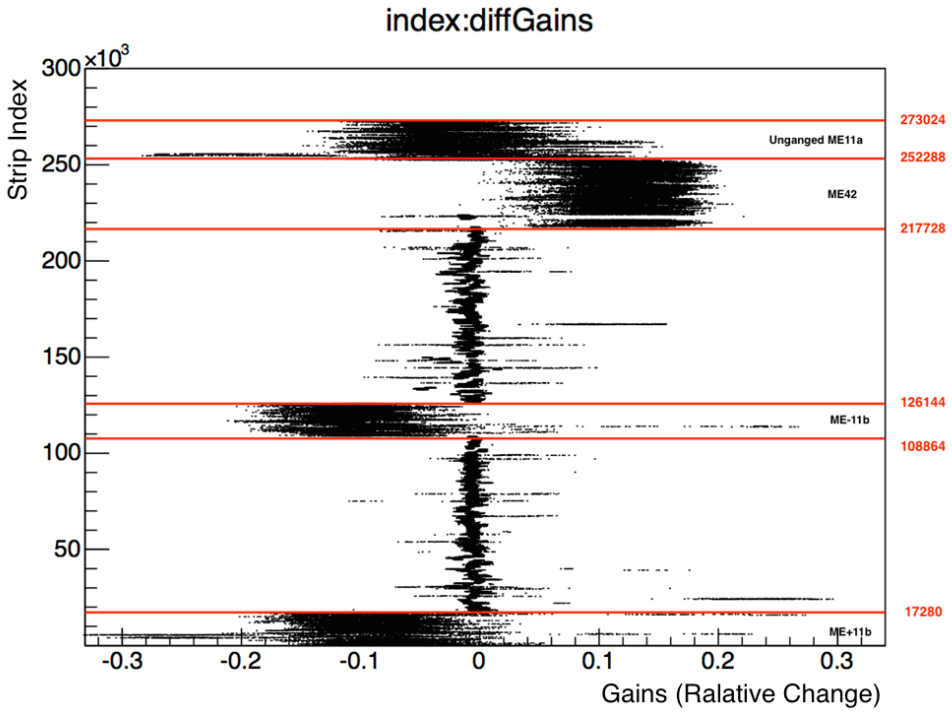


FIGURE 2.14: Relative changes in the strip channel electronic gains of CSC before and after the LS1. The vertical axis indicates the strip index which increases sequentially with endcap, station, ring, and chamber.

process such large amount of data. In addition, interesting events such as the production of W or Z boson occur rarely. Therefore, a system dedicated to the selection of potentially interesting events is needed. This system reduces the event rate to be recorded for subsequent analysis. This task is handled by the CMS trigger system whose function is divided into two steps: the Level-1 (L1) Trigger and the High-Level Trigger (HLT). The reduction in rate is designed to be at least a factor of 10^6 for the combined L1 Trigger and HLT.

The L1 Trigger consists of custom-designed, largely programmable electronics utilizing FPGA technology. In addition, ASICs technology and programmable memory lookup tables (LUT) are also used in the situations where the requirements in speed, density and radiation resistance are critical. Some of the electronics are mounted on the detector while the remaining are housed in the underground control room located at a distance of approximately 90 m from the experimental cavern. The Level-1 trigger automatically processes coarse data from

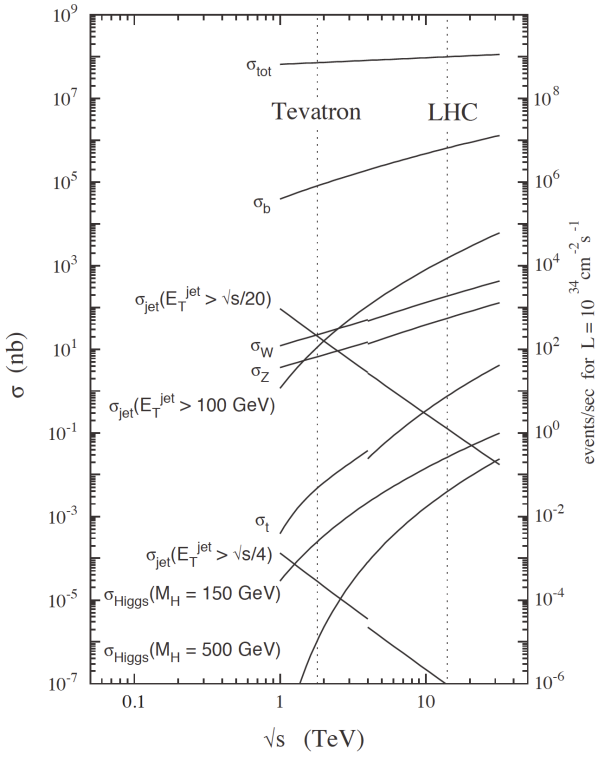


FIGURE 2.15: Cross sections for various processes in the proton-proton collisions at center of mass energies relevant to LHC physics [N. Rompotis].

the calorimeters and muon system looking for simple signature of interesting physics. It reduces the event rate down to about 30 kHz and passes these data to the HLT.

The HLT is a software system run on a farm of commercial processors. The HLT utilizes complete read-out data from all subdetectors and performs sophisticate calculations to ensure that only the interesting events will be selected and finally written to permanent storage for offline analysis. It is designed to reduce the output event rate to about 100 Hz.

2.3 Event Simulation

The structure of the events recorded at a colliding experiment is very complicated. To obtain precise information of what is produced during the collisions and inside the detector, the use of simulations is necessary. Such simulations are achieved via Monte Carlo (MC)

techniques in which the calculation of production cross sections is implemented and the particles produced in the interaction are made to interface with a realistic simulation of the detector response.

In general, the generation of an event follows multiple successive steps starting from the matrix element (ME) calculation of the underlying hard process with a given final state parton multiplicity and ending with the hadronization of the final state partons.

After the events are generated, a full simulation of the detector response and a reconstruction of that response is applied so that they can be directly compared with the reconstructed collision data. This procedure is achieved by using a simulation of the CMS detector response based on the GEANT4 [48] simulation toolkit. The details about GEANT4 will not be discussed further, it is advised to review the reference given above.

The following gives a brief description of each successive step in the event simulation of pp collisions at the LHC. Fig. 2.16 represents a pp collision generated by a typical event generator and it depicts how these steps fit into the simulation chain.

- **Hard parton-parton interaction**: This is a calculation of the fully differential cross section for the colliding partons for the process under consideration. It is computed by the matrix element and the corresponding phase space integration. It may also include full polarization treatment of initial, intermediate and final state particles (bosons, leptons, partons). The calculation includes n-final state particles and is computed at LO, NLO or NNLO. In particular, LO generators called the multi-leg generators incorporate the real emission amplitudes from higher order calculations by means of a larger number of final state partons.

- **The colliding protons' structure**: The hadron-hadron differential cross section is obtained from the partonic cross section calculated above, convoluted with the appropriate parton distribution function (PDF) which accounts for the probabilities for

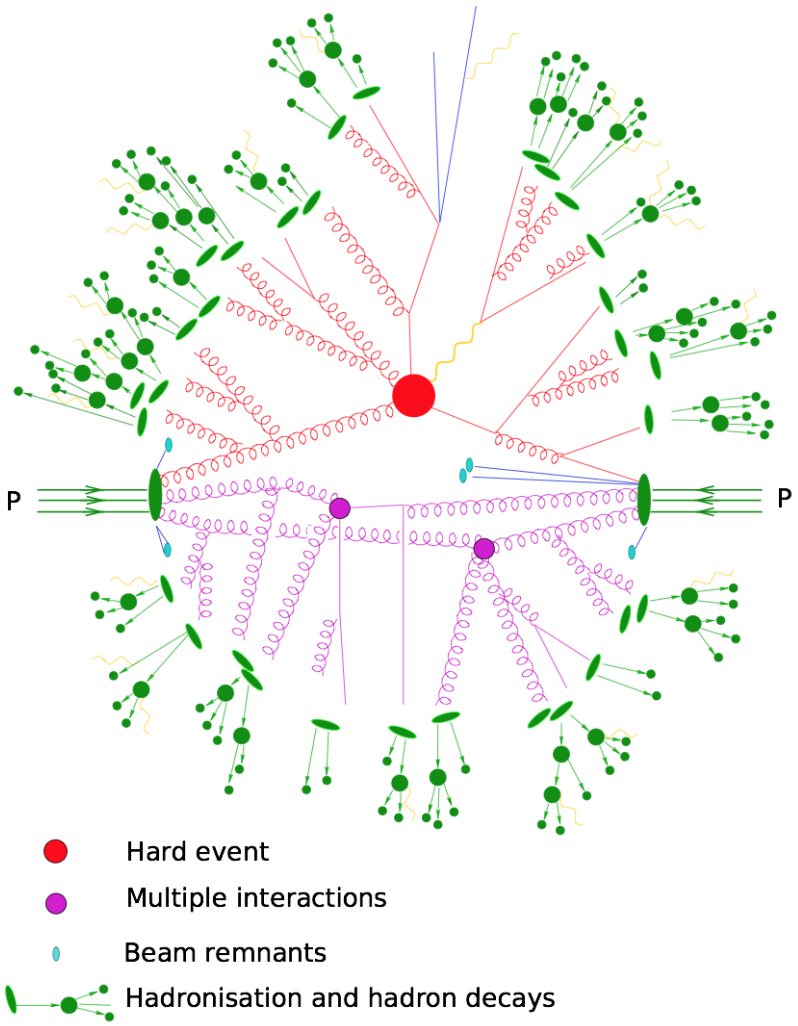


FIGURE 2.16: The schematic representation showing typical components in the process of event simulation in pp collision. Specifically, in this example diagram, two partons (gluon, quark, or antiquark) from the incoming protons interact via the hard scattering process (red blob). A gluon, a quark (or antiquark) and a boson are produced and subsequently decay into a quark-antiquark pairs. Before the incoming partons interact, additional radiation is emitted (initial state radiation). Similarly, additional radiation is emitted from final state partons (final state radiation). The softer multiple interactions (purple blobs) are also illustrated. In addition, the fragments of the initial hadrons are also taken into account (cyan). Then, the quarks and gluons produced during the interactions form into hadrons by hadronization. Finally, the hadrons can decay further into more stable particles (green) (This figure was adapted from [49] by Leonard A.).

certain initial state parton configurations within the colliding protons. The precision of the PDF set (LO, NLO or NNLO) should match the precision of the matrix element calculation.

- **ISR and FSR**: Additional radiation, produced when partons are accelerated, before and after the hard interaction has to be taken into account. The radiation is in the form of high energy gluons called initial state (ISR) or final state radiation (FSR). The modeling and simulation of this radiation are performed by parton shower programs. It is typically computed at LL precision. Additionally electromagnetic radiation (photon radiation) of charged particles, in particular electrons and muons, are then simulated. This is also called ISR and FSR. For electrons, the FSR radiation as well as bremsstrahlung effects in the detector are crucial.

- **Short-life particles decays**: Short-life particles, such as W/Z bosons and π^0 decay and produce additional final state particles.

- **Hadronization**: The colored final state quarks and gluons form into color-neutral baryons and mesons due to the confinement effect. This non-perturbative process is simulated by dedicated hadronization models which are tuned to data.

- **Long-life particles decays**: Relatively-long-life particles such as τ leptons, B-hadrons or some baryons and mesons decays before reaching the detector. In general, decays are simulated for particles with lifetimes up to 30 ps. Particles with lifetime longer that this (for example electrons, muons, photons, charged pions, protons and neutrons) are consider stable and no decay is simulated.

- **Underlying event**: The partons of the colliding protons which do not participate in the hard interaction can undergo additional softer parton-parton interactions, resulting in multiple parton interactions (MPI). In addition, the protons broken up by the hard interaction are still colour-connected to the rest of the event. The description of the

proton remnant thus includes additional radiation and hadronization. It is therefore necessary to simulate MPI and proton remnant interactions in order to describe the complete structure of an event.

- **Pile-up**: These are softer proton-proton (pp) collisions that occur almost at the same time at which the hard interaction takes place. The frequency of these collisions increases with the number of protons in a bunch.

  In summary, the generation of an event can be described as follow. First, the calculation of the parton-level differential cross section for the considered process (hard interaction) is performed. Integrating over the phase space, with the appropriate PDF, the hadron-hadron differential cross section is calculated. Then the final state partons involved in the process are generated according to the known branching factors of decay. The additional particles from initial state radiation (ISR), final state radiation (FSR), short-life particles decays and multiple parton interactions (MPI) are produced based on appropriate splitting functions and models. Finally, final state stable partons are hadronized.

  Various MC event generators are available. Some of them can perform specific tasks such as total cross section calculations, and predictions of kinematic distributions. Some of them have capability to generate a complete event while others can only provide the calculation of the hard process (either at LO or NLO) and have to be interfaced with another MC to perform parton shower (PS) and complete the event generation.

  With the ME+PS approach, great care must be taken in order to avoid double counting. For an event with n+1 partons in the final state, all of the partons may have resulted from the soft radiation evolution, or they may have evolved from an initial n-parton configuration to which PS adds an additional emission. To avoid double counting,

a so-called matching scheme (matching prescription) is utilized. Common matching schemes are the CKKW, MLM, and FxFx matching schemes [50] [51].

Chapter 3

Event Reconstruction

3.1 Track and Vertex Reconstruction

The inner tracking system is the first detector system that interacts with the particles emitted from the collisions. In particular, electrically charged particles propagate as helices through the tracker layers generating multiple hits on the different layers of the tracking system. A trajectory is fitted to the tracker hits using an algorithm based on a Kalman filter [52] while taking into account the uncertainty of the hit positions and the effects of random multiple coulomb scattering and energy loss. The result is a reconstruction of a track corresponding to the trajectory of the charged particle.

The reconstructed tracks are subsequently clustered using a deterministic annealing algorithm [52]. These clusters are used to infer the positions of the vertices which correspond to hard collisions. Multiple vertices can be reconstructed in an event. The primary vertex is chosen to be the reconstructed vertex with the largest value of the sum of the squared (reconstructed) track transverse momenta.

3.2 Particle-Flow Algorithm

The particle-flow (PF) algorithm [53] combines information from various elements of the CMS detector in the reconstruction and identification of final state particles, such as photons, electrons, muons, charged hadrons and neutral hadrons, as individual PF objects. Combinations of PF objects are then used to reconstruct higher-level objects such as jets and missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

3.3 Muon Reconstruction

Muons are reconstructed using information from both the muon system and the inner tracking system. In the muon system, muon track segments are generated by requiring a certain number of hits in the layers within a single DT or CSC chamber. These segments are then fitted, including information from the RPCs, to form standalone-muon tracks. For each standalone-muon track, a matched tracker track, reconstructed independently in the inner tracking system, is found by comparing parameters of the two tracks projecting onto a common surface. Finally, a global-muon track is fitted combining hits from the tracker track and the standalone-muon track. This reconstruction approach is referred to as "global muon reconstruction (outside-in)" [54].

Further selection based on various muon identification variables can be applied. The selection is analysis dependent and is discussed in details in each analysis section. Another requirement that is applied in all analyses discussed in this thesis is the muon isolation requirement. The particle-flow relative isolation is adopted and it is defined as:

$$I_{\mathrm{PF}}^{\mathrm{rel}} = \frac{1}{p_{\mathrm{T}}^{\mu}} \left[\sum^{\mathrm{charged}} p_{\mathrm{T}} + \max\left(0, \sum^{\mathrm{neutral}} p_{\mathrm{T}} + \sum^{\gamma} p_{\mathrm{T}} - 0.5 \sum^{\mathrm{PU}} p_{\mathrm{T}}\right)\right], \tag{3.1}$$

where the sums run over charged hadrons originating from the primary vertex of the event, neutral hadrons, photons (γ), and charged hadrons not originating from the primary vertex, but from PU. The sums consider only PF candidates with direction within a cone defined by $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.4$ around the direction of the muon candidate track. The transverse momentum of the muon candidate is denoted by p_{T}^{μ}. Because neutral PU particles deposit on average half as much energy as charged PU particles, the contamination in the isolation cone from neutral particles coming from PU interactions is estimated as $0.5 \sum^{\mathrm{PU}} p_{\mathrm{T}}$ and it is subtracted in the definition of $I_{\mathrm{PF}}^{\mathrm{rel}}$.

3.4 Electron Reconstruction

Electrons are charged particle interacting with matter through electromagnetic interaction, hence leaving tracks in the silicon tracker. The mass of electrons is much lower than that of muons, so the radiative loss is much more significant such that electrons are normally stopped and deposit all of their energy inside the ECAL. Usually, the energy deposit from an electron spreads out over several crystals of the ECAL, as the electron creates electromagnetic showers. The reconstruction of an electron candidate, therefore, uses information from the silicon tracker and the ECAL, by matching tracks in the tracker to the energy deposited in the ECAL.

An ECAL "supercluster", a group of one or more clusters of energy deposits, is reconstructed through the "Hybrid" or the "multi-5×5" algorithm while an electron track is reconstructed using a Gaussian-Sum Filter (GSF) algorithm. A GSF track is then associated with an ECAL supercluster to build a GSF electron candidate. The association is done by extrapolating the track, through the magnetic field, from the innermost track position to the point of closest approach to the supercluster's energy-weighted mean position. The criterion for a valid association considering the compatibility in these positions is $|\Delta\eta| < 0.02$ and $|\Delta\phi| < 0.15$.

Requirements on electron identification criteria [55] based on the quality of the reconstruction may be further imposed. The particle-flow relative isolation requirement is also adopted for electrons and its definition is as follows;

$$I_{\mathrm{PF}}^{\mathrm{rel}} = \frac{1}{p_{\mathrm{T}}^{\ell}} \left[\sum^{\mathrm{charged}} p_{\mathrm{T}} + \max\left(0, \sum^{\mathrm{neutral}} p_{\mathrm{T}} + \sum^{\gamma} p_{\mathrm{T}} - \rho A_{\mathrm{eff}}\right)\right], \tag{3.2}$$

where the sums run over charged hadrons originating from the primary vertex of the event, neutral hadrons, and photons (γ) inside a cone of radius $\Delta R < 0.3$ around the direction of the electron. The isolation includes a correction for pileup effects, ρA_{eff}, where ρ is the average transverse momentum flow density, calculated using the jet area method [56], and A_{eff} is the geometric area of the isolation cone times an η-dependent correction factor that accounts for residual PU effects.

3.5 Jets and $E_{\mathrm{T}}^{\mathrm{miss}}$

Quarks or gluons are produced in proton-proton collisions either through hard interactions or through QCD radiations from the initial or final state partons. As a consequence of QCD confinement, they form a collection of color-neutral hadrons, and are observed as a cluster of particles called jets. Jets can be classified according to their origin as quark-initiated jets and gluon-initiated jets of which various characteristics are different [57]. Quark jets can be further categorized based on their quark flavor. Heavy-flavor jets originate from b or c quarks, while light-flavor jets originate from u, d, or s quarks.

Jets are reconstructed from the list of PF objects using the anti-k_{T} [58] clustering algorithm, as implemented in the FastJet package [59] with a size parameter R being the algorithm parameter whose value can be chosen.

The energy of the reconstructed jets has to be corrected for various effects including non-uniformity and non-linearity effects of the ECAL and HCAL energy response to neutral hadrons, for the presence of extra particles from PU, for the thresholds used in jet constituent selection, reconstruction inefficiencies, and possible biases introduced by the clustering algorithm. The jet energy corrections which are p_{T}- and η-dependent are extracted from the measurement of the p_{T} balance in dijet and γ+jet events [60, 61]. An additional residual η- and p_{T}-dependent calibration is applied to correct for the small differences between data and simulated jets.

The jet energy resolution is approximately 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV [61]. The jets in simulated events are smeared by an η-dependent factor to account for the difference in energy resolution between data and simulation [60].

The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is defined as the projection on the plane perpendicular to the beams of the negative vector sum of the momenta of all reconstructed PF objects in an event. Its magnitude is referred to as $p_{\mathrm{T}}^{\mathrm{miss}}$ or $E_{\mathrm{T}}^{\mathrm{miss}}$ [62], and it is a measure of the p_{T} of particles leaving the detector undetected. Corrections to the jet energies are propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$ by replacing the vector p_{T} sum of particles that were clustered as jets by the vector p_{T} sum of the jets including the jet energy corrections.

3.5.1 b Jet Tagging

The identification of jets originating from b quarks is crucial to both W+jets and double Higgs analyses studied in this thesis. The tagging of jets as b jets relies on the unique properties of B hadrons including: a long life-time hence the presence of a secondary vertex displaced from the primary vertex where the B hadron is produced; a high number of charged particles per decay hence a large number of tracks within a jet; the possibility of a semi-leptonic decay hence the presence of a nearby lepton. Several algorithms based on these

characteristics have been developed and are described in [63, 64, 65]. Each produces, as output, a numerical "discriminator" which can be used for jet selection according to a chosen balance between efficiency for tagging b-quark jets versus misidentification probability for light-flavor jets. Three working points have been defined: "loose" (L), "medium" (M) and "tight" (T) for 10%, 1% and 0.1% misidentification probabilities respectively.

The combined secondary vertex (CSV) algorithm and the combined MVA (CMVA) algorithm are used in the W+jets analysis and in the double Higgs analysis, respectively. The combined secondary vertex algorithm (CSV) combines information about the reconstructed secondary vertices together with lifetime information such as impact parameter significance or decay length, and information on jet kinematic properties. The CSV algorithm was further optimized and the new version is referred to as CSVv2. The CMVA algorithm combines the information from six different b jet identification discriminators [63] with a Boosted Decision Tree (BDT).

Differences in the b tagging efficiency in data and simulation, as well as differences in mistagging rates, can be corrected for, by applying data-to-simulation scale factors which were derived as a function of jet p_{T} [63, 64, 65].

Chapter 4

Analysis I: W+Jets Measurement

4.1 Measurement Strategy

The production of W+jets process, which is referred to as 'signal' in the context of this analysis, can be measured through leptonic decay channels of W boson as illustrated in the left diagram of Fig. 1.3. The signature of these signal events is characterized by an energetic isolated lepton, a neutrino leading to significant missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ in the detector, and additional jets. Hence, the first step is to collect the data events which have this signature. However, there are many processes which have final states similar to those of W+jets signals. These processes are referred to as backgrounds and are discussed further in Sec. 4.3. Therefore, the strategy is to confine the measurement in the phase space where the signal process is dominant and backgrounds are minute in order to maximize signal-to-background ratio. Such phase space can be achieved by imposing kinematic selection requirements, for example a minimum value of $E_{\mathrm{T}}^{\mathrm{miss}}$, on data events. The background yields are estimated either by MC simulation or with data control regions. The dominant background, especially for events with high jet multiplicity, is $\mathrm{t\bar{t}}$. To reduce the contamination of $\mathrm{t\bar{t}}$ background, jets originating from b quarks, so called b-tagged jets, are identified, and events with one or more

b-tagged jets are discarded. Then the comparison of data with simulated processes at the reconstruction level for the observables that are used for the cross section measurement are investigated. Next, the estimated backgrounds are subtracted from the data yielding a measurement of W+jets distribution which can be measured as function of various observables such as jet multiplicities and kinematic variables of interest. The resulting distributions are then corrected for the effect of detector response, finite experimental resolutions, acceptance and efficiencies, allowing for a direct comparison with theoretical predictions. The correction procedure is achieved by mean of regularized unfolding which is discussed in detail in Sec. 4.4. The unfolding procedure is also used to derive two additional corrections; (i) the correction for events where the muon candidates originated in the decay of tau leptons from W bosons decays and (ii) the correction for the efficiency lost in discarding events with b-tagged jet(s). Then, the corrected distributions are converted into differential cross-section and are compared to theoretical predictions.

4.2 Overview

In this thesis, differential cross sections for W+jets in pp collisions at $\sqrt{s} = 8$ TeV and 13 TeV are measured with data recorded by the CMS detector and correspond to an integrated luminosity of 19.6 fb^{-1} [66] and 2.2 fb^{-1} [67] respectively. The analysis focuses on the production of W+jets with the subsequent decay of the W boson into a muon and a neutrino. The corresponding decay channel of the W boson into an electron and a neutrino is not used in this analysis because a higher momentum threshold was applied to the electron when acquiring data, resulting in lower statistics such that incorporating the electron channel would not add much benefit to the analysis. The final-state topology, therefore, consists of one isolated muon with high transverse momentum p_{T}, significant $E_{\mathrm{T}}^{\mathrm{miss}}$, and jets.

For the sake of a clear presentation, the analysis using data from pp collisions at $\sqrt{s} = 8$ TeV and 13 TeV are discussed separately in Sec 4.6 and Sec 4.7 respectively. They are referred to as W+jets at 8 TeV analysis and W+jets at 13 TeV analysis for the rest of the thesis.

For the W+jets at 8 TeV analysis, the differential cross sections are reported as functions of jet multiplicity, p_{T} of the jet, and the scalar sum of jet transverse momenta (H_{T}) for different jet multiplicities. The differential cross sections as functions of angular correlation variables are also measured. Finally, the average number of jets per event, $\langle N_{\mathrm{jets}} \rangle$, is studied as functions of H_{T} and angular variables. The study extends the kinematic reach (in p_{T} , H_{T}, and jet multiplicity) of the 7 TeV CMS results [22] and expand the set of kinematic observables, beyond what has been achieved so far. This is motivated by the increased center-of-mass energy and the larger data sample. These allow to study events with higher jet multiplicities that are sensitive to higher-order processes and to more accurately study angular correlation variables that can probe how particle emissions are modeled by the MC generators used in the analysis of the LHC data, and by the most current NLO calculations. The observables studied are discussed in details in Sec. 4.6.4. The results of up to 7 jets on jet multiplicity distributions and up to inclusive 4 jets on observables are presented.

For the W+jets at 13 TeV analysis, the first W+jets differential cross section measurement using 13 TeV pp collision data are presented. The differential cross sections are reported as functions of the jet multiplicities up to 6 jets. The measurement as functions of the jet p_{T}, the jet $|y|$, H_{T}, and of the azimuthal correlations between the muon and the ith jet from the p_{T}-ordered list of jets in the event $\Delta\phi(\mu, j_i)$ up to multiplicities of four inclusive jets are also performed. In addition, the differential cross section is measured as a function of the angular distance between the muon and the closest jet $\Delta R(\mu,$ closest jet) for events with one or more jets of high p_{T}.

The measured cross sections are compared with several predictions including predictions from MC event generators and from fixed-order parton-level calculations detailed in Sec. 4.5.

4.3 Background Processes

To isolate signal events, we require the presence of an energetic isolated lepton, a significant $E_{\mathrm{T}}^{\mathrm{miss}}$, and additional jets. The other physical processes which may provide the same signature as W+jets in the experimental observation are referred to as backgrounds. The event selection described in details in 4.6.2 and 4.7.2 is established in order to reduce the contamination of major backgrounds while also maintain good statistics of W+jets signal processes. After the selection, a significant number of the events from background processes can still be present. The following backgrounds are considered:

- **QCD multi-jet processes**; these processes have a very high rate at the LHC. This non-negligible background contribution originates from hadronic interaction process. Particularly, the heavy mesons, formed by bottom and charm quarks ($b\bar{b}, c\bar{c}$), when are produced copiously in these events, decay semi-leptonically into leptons with large branching ratios, eventually providing a muon in observed events. The muons produced in semi-leptonic decays of bottom and charm hadrons are typically non-isolated and appear embedded in jets. However, there are instances when the jet is not reconstructed (e.g. below threshold) and the muons appear isolated. Because multijet processes have high rates of production, non-isolated muons in multijet processes can still lead to non-negligible background contributions to W+jets. Leptons that originate from the decay of charged pions (muons in this case) and kaons in flight are relatively soft and mostly cannot survive the selection thresholds. Since only neutrinos coming from semi-leptonic decays of b and c hadrons provide a true source of $E_{\mathrm{T}}^{\mathrm{miss}}$, resulting in small reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$, the contribution from multijet processes can be significantly reduced by imposing a requirement on $E_{\mathrm{T}}^{\mathrm{miss}}$ or M_{T}, the transverse mass of the muon and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, which is defined as $M_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) \equiv \sqrt{2p_{\mathrm{T}}^{\mu} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi)}$, where $\Delta\phi$ is the

difference in the azimuthal angle between the direction of the muon momentum and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ (see the definition in Sec. 3.5)

- Drell-Yan (Z/γ^*) +jets, where Z/γ^* decays into oppositely charged leptons of the same flavor, in which one of the leptons falls out of the fiducial volume of the detector and thus is not reconstructed. The contribution of Z bosons decaying into taus, with at least one of the taus decaying leptonically into a muon/electron, is considerably small.

- $\mathrm{t\bar{t}}$, where one W boson (from $t \to bW$) decays to a muon and a neutrino, while the remaining decay products are reconstructed as four jets, two of which originate from the b-quarks. This contribution is reduced with a veto on the presence of b-jets.

- Single top quark production, which is similar to $\mathrm{t\bar{t}}$ +jets, but with a lower number of jets,

- Dibosons (ZZ/WZ/WW) +jets, where the W or Z decay leptonically but only one lepton is left inside the detector acceptance.

4.4 Unfolding

In order to compare the results of measurements of physics processes, such as W+jets production, with different experiments or theoretical predictions, the detector specific effects need to be removed from the measured distributions. In other words, the distributions have to be corrected for the effect of detector response, finite experimental resolutions, acceptance and efficiencies. The correction procedure is called "unfolding". Several methods are available to perform unfolding of measured distributions. In this analysis, the procedure is performed using the iterative d'Agostini method [68] implemented in the RooUnfold toolkit [69].

4.4.1 Iterative Bayes theorem

The iterative Bayes theorem is summarized here. The corrected number of events $\hat{n}(i)$ in a particular bin i for a distribution is given by

$$\hat{n}(i) = \sum_{j=1}^{n_{\text{bins}}} M(i_{\text{gen}}|j_{\text{reco}}) \cdot n_{\text{obs}}(j).$$

This formula gives the solution in terms of $n_{\text{obs}}(j)$ the number of events observed in bin j in the observed reconstructed distribution, and the unfolding matrix $M(i_{\text{gen}}|j_{\text{reco}})$

$$M(i_{\text{gen}}|j_{\text{reco}}) = \frac{R(j_{\text{reco}}|i_{\text{gen}})P_0(i_{\text{gen}})}{[\sum_{l=1}^{n_{\text{bins}}} R(l_{\text{reco}}|i_{\text{gen}})] \cdot [\sum_{l=1}^{n_{\text{bins}}} R(j_{\text{reco}}|l_{\text{gen}})P_0(l_{\text{gen}})]}.$$

The term $0 < \sum_{l=1}^{n_{\text{bins}}} R(l_{\text{reco}}|i_{\text{gen}}) \leq 1$ is the efficiency of observing an event generated in bin i. The so-called (normalized) response matrix $R(j_{\text{reco}}|i_{\text{gen}})$ represents the probability to observe in bin j an event generated in bin i. The knowledge of the response matrix is obtained from MC simulation. $P_0(i_{\text{gen}})$ is the initial probability for the event to be generated in bin i. The general theory does not strictly impose the choice of the initial probability but suggests that it is chosen based on the best knowledge of the process under study. In this analysis, it is chosen to mimic the generated level probability distribution, hence denoted by the subscript *gen*.

These expressions can be rewritten in terms of the smearing matrix S.

$$\hat{n}(i) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_{\text{bins}}} S(i_{\text{gen}}|j_{\text{reco}}) \cdot n_{\text{obs}}(j).$$

Where the efficiency is defined as $\epsilon_i \equiv \sum_{l=1}^{n_{\text{bins}}} R(l_{\text{reco}}|i_{\text{gen}})$ and the smearing matrix S is defined as:

$$S(i_{\text{gen}}|j_{\text{reco}}) = \frac{R(j_{\text{reco}}|i_{\text{gen}})P_0(i_{\text{gen}})}{\sum_{l=1}^{n_{\text{bins}}} R(j_{\text{reco}}|l_{\text{gen}})P_0(l_{\text{gen}})}.$$

The iteration procedure is to replace the initial probability distribution $P_0(i_{\text{gen}})$ with the new one obtained, $\hat{P}(i)$. The calculation of $\hat{P}(i)$ is straightforward and uses the unfolded result $\hat{n}(i)$.

$$\hat{N}_{\text{gen}} = \sum_{i=1}^{n_{\text{bins}}} \hat{n}(i)$$

$$\hat{P}(i) = \frac{\hat{n}(i)}{\hat{N}_{\text{gen}}}$$

Note that the only quantity which is updated is $P_0(i_{\text{gen}})$. The response matrix is not updated. Therefore, if there are ambiguities concerning the choice of this matrix, one has to estimate them.

The iteration procedure stops when a stable solution is obtained. Several approaches can be used to justify the stable solution. One approach is to use a χ^2 comparison between $\hat{n}_{\text{current}}(i)$ and $\hat{n}_{\text{previous}}(i)$. The approach adopted in this analysis is described in Sec 4.4.2. The error calculation for the unfolded distribution is described in [69].

4.4.2 Unfolding Procedures

As described in 4.1, a measurement of W+jets distribution at reconstruction level is obtained by subtracting the estimated backgrounds from data. Then a response matrix, which defines the event migration probability between the particle-level and reconstructed quantities, is constructed using generator and reconstruction levels of W+jets simulated events, respectively.

At the generator level, the events are required to pass the same kinematic selection used at the reconstruction level. This generator-level selection defines the fiducial phase space of the measurements. The generator level quantities refer to the stable leptons, a muon and a neutrino, from the decay of the W boson and to jets built from stable particles excluding neutrinos, using the same algorithm as for the measurement. Particles are considered stable if their decay length $c\tau$ is greater than 1 cm. The muons are 'dressed' by recombining the bare muons and all of the radiated photons in a cone of $\Delta R < 0.1$ around the muon to account for the FSR effects. At the generator level, the M_{T} of the W boson is calculated using the dressed muon and the neutrino.

The response matrix, the reconstructed-level distribution and the corresponding generator-level distribution contain information about the migration probability between the reconstructed and generated quantities as well as the information used to determine the effect of inefficiencies. Events that pass the reconstructed-level selection but are absent from the generator-level selection due to migrations across neighboring bins are estimated, and these events are subtracted from the measured distribution.

The unfolding procedure is regularized by choosing the number of iterations optimized by folding the unfolded distributions with the response matrix and comparing to the initial measured distributions. The number of iterations is determined when the distribution obtained by folding the unfolded distribution becomes compatible with the initial measured distribution based on a χ^2 comparison. A minimum of four iterations is required to avoid biasing the unfolded results towards the simulated sample used to construct the response matrix. This unfolding procedure described is applied separately to each measured differential cross section.

4.5 Theoretical Predictions

Theoretical predictions can be broadly divided into two categories. The first category is the prediction based on matrix element (ME) calculation plus parton shower (PS) MC simulation. The second category is the fixed-order calculation at the level of ME patrons.

4.5.1 ME+PS MC Event Generators

The details of MC event simulation are discussed in Sec. 2.3. The summary of the theoretical predictions based on ME+PS MC event generator considered in this analysis is given below.

MadGraph5+pythia MadGraph 5 [70] is a MC generator which provides LO ME calculations of the underlying hard process. The output is interfaced with pythia [71], another MC generator which has the capability to implement the remaining steps such as PS and hadronization. Implementations of ME+PS use the MLM matching scheme.

MadGraph5_aMC@NLO+pythia MadGraph5_aMC@NLO [72] is a recently released MC generator providing the computation of tree-level and NLO differential cross sections. It is interfaced with pythia for the PS, for comparison with the data. The merging of parton showers and MEs is done with the FxFx merging scheme [51].

sherpa 2 Simulation of High Energy Reactions of Particles (sherpa) [73, 74, 75] is a generator that allows for complete hadronic final states in high-energy particle collisions. sherpa 2 allows for NLO calculations, and for the case of W+jets with up to four jets, with the 0, 1 and 2 jets multiplicities at NLO in QCD accuracy, and the additional 3 and 4 jets multiplicities with LO accuracy.

4.5.2 Fixed-Order Theoretical Calculations

BlackHat+sherpa The BlackHat+sherpa [76] calculation yields fixed-order NLO predictions for W + n jets at the level of ME partons, where n = 1–4.

N_{jetti} NNLO This provides a calculation of W +1-jet production at NNLO in perturbative QCD based on the N-jettiness subtraction scheme (N_{jetti} NNLO) [77, 78].

4.6 W+Jets 8 TeV Analysis

4.6.1 Samples and Dataset

4.6.1.1 Data Samples

The data used in this analysis is from the pp collision at $\sqrt{s}$ = 8 TeV collected in 2012. The analysis uses only the data with the validated luminosity sections of which information was recorded in a JSON formatted file listed in Table 4.2. The data corresponds to an integrated luminosity of 19.6 fb^{-1}. Events come from the single muon (`SingleMu`) primary dataset. Events are selected for analysis if they pass the single muon trigger (listed in Table 4.2) requiring one isolated muon with $p_T > 24$ GeV and $|\eta| < 2.1$. This analysis has been done using the `CMSSW_5_3_11` software release.

Table 4.1: Datasets used in W+jets 8TeV analysis.

Name	Run-Range
/SingleMu/Run2012A-22Jan2013v1/AOD	190456-193621
/SingleMu/Run2012B-22Jan2013v1/AOD	193834-196531
/SingleMu/Run2012C-22Jan2013v1/AOD	198022-203742
/SingleMu/Run2012D-22Jan2013v1/AOD	203777-208686

TABLE 4.2: JSON file and the single muon trigger used in the W+jets at 8 TeV analysis.

JSON File
Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt

Single Muon Triggers
HLT_IsoMu24_eta2p1

4.6.1.2 Monte Carlo Samples

The backgrounds considered in this analysis include $t\bar{t}$, Drell-Yan (Z/γ^*) +jets, single top quark, diboson (ZZ/WZ/WW) +jets, and QCD multijet production. The details of how the final states of these backgrounds mimic those of signal are discussed in Chapter 4.3. The background processes are estimated from MC simulations, except for the QCD multijet background, which is estimated from control data samples, discussed in details in Chapter 4.6.3.

The simulated samples of $t\bar{t}$ and Z/γ^*+jets events are generated with MADGRAPH version 5.1.1; the single top quark samples (s-, t-, and tW-channel production) are generated with POWHEG version 1 [79, 80, 81, 82]; and the diboson samples (WW, WZ, or ZZ) are generated with PYTHIA 6.424 using the Z2* tune. The simulations with MADGRAPH and PYTHIA use the CTEQ6L1 PDFs, and the simulations with POWHEG use the CTEQ6M PDFs. The Z/γ^*+jets sample is normalized to the NNLO inclusive cross section calculated with FEWZ 3.1 [83]. Single top quark and diboson samples are normalized to NLO inclusive cross sections calculated with MCFM [84, 85, 86, 87]. The $t\bar{t}$ contribution is normalized to the predicted cross section at NNLO with next-to-next-to-leading-logarithm accuracy [88].

The W+jets signal process is simulated with the ME generator MADGRAPH 5.1.1 [70] interfaced with PYTHIA 6.426 using the Z2* tune for parton showering and hadronization. This sample of events, denoted MADGRAPH5+PYTHIA6 (denoted as MG5+PY6 in the figure legends), is generated with the CTEQ6L1 parton distribution function (PDF) set [89]

and is normalized to the inclusive NNLO cross section calculated with FEWZ 3.1 [83]. The MADGRAPH5+PYTHIA6 calculation includes the production of up to four partons at LO. The jets from matrix elements are matched to parton showers following the k_T-jet MLM prescription [50], where partons are clustered using the k_T algorithm [59] with a distance parameter of 1. The merging of parton showers and matrix elements with the MLM scheme uses a matching scale of 20 GeV. The factorization and renormalization scales for the 2→2 hard process in the event are chosen to be the transverse mass of the W boson produced in the central process. The k_T computed for each QCD emission vertex is used as renormalization scale for the calculation of the strong coupling constant α_S of that vertex.

Signal and background simulated events are generated and passed through detector simulation based on GEANT4 [48] description of CMS. Each simulated sample is normalized to the integrated luminosity of the data sample. Simulations also include additional collisions in the same or adjacent bunch crossings (pileup, PU). To model PU, minimum-bias events generated with PYTHIA6 using the Z2* tune [90] are superimposed on the simulated events, matching the multiplicity of PU collisions observed in data, which has an average value of approximately 21.

All MC samples in this analysis were produced during the `Summer2012` official production. They are listed in Table 4.3

Additional event generators are used for the W+jets process, when comparing the measurements with theoretical predictions. These MC generators are described in Sec. 4.5 and Sec. 4.6.7.

4.6.2 Event Selection

Events are required to pass the single muon trigger *HLT_IsoMu24_eta2p1* which requires one isolated muon with $p_T > 24$ GeV and $|\eta| < 2.1$. The simulated events are required to

TABLE 4.3: MC samples used in the W+jets at 8 TeV analysis.

Name
/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball v1&v2
/W1JetsToLNu_TuneZ2Star_8TeV-madgraph
/W2JetsToLNu_TuneZ2Star_8TeV-madgraph
/W3JetsToLNu_TuneZ2Star_8TeV-madgraph
/W4JetsToLNu_TuneZ2Star_8TeV-madgraph
/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball
/DYJetsToLL_M-10To50_TuneZ2Star_8TeV-madgraph
/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola
/ZZ_TuneZ2star_8TeV_pythia6_tauola
/WZ_TuneZ2star_8TeV_pythia6_tauola
/WW_TuneZ2star_8TeV_pythia6_tauola
/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola
/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola
/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola
/T_s-channel_TuneZ2star_8TeV-powheg-tauola
/T_t-channel_TuneZ2star_8TeV-powheg-tauola
/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola

pass an emulation of the trigger requirements applied to the data. Trigger efficiencies in the simulation are corrected for differences with respect to the efficiencies in data.

The final-state particles in W+jets events are identified and reconstructed with the particle-flow (PF) algorithm [91, 92], which optimally combines the information from the various elements of the CMS detector.

Muons are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and to be reconstructed in the HLT fiducial volume $|\eta| < 2.1$. The track associated with a muon candidate is required to have hits in at least six strip tracker layers, at least one pixel hit, segments from at least two muon stations, and a good quality global fit with χ^2 per degree of freedom < 10. To reject muons from cosmic rays, the transverse impact parameter of the muon candidate with respect to the primary vertex is required to be less than 2 mm, and the longitudinal distance of the tracker track from the primary vertex is required to be less than 5 mm. In order to reduce the contamination due to muons that do not originate from the decay of a W boson, an isolation requirement

(defined in Sec. 3.3) of $I_{\mathrm{PF}}^{\mathrm{rel}} \leq 0.12$ is imposed. Trigger efficiency corrections, as well as muon identification and isolation efficiency corrections, are applied to the simulation as a function of p_{T} and η on an event-by-event basis and are generally less than 4% and 2.5%, respectively.

Jets and transverse missing energy $E_{\mathrm{T}}^{\mathrm{miss}}$ are also reconstructed using the PF algorithm. Jets are reconstructed using the anti-k_{T} [58, 59] algorithm with a distance parameter of 0.5. Reconstructed jet energies are corrected with p_{T}- and η-dependent correction factors

Jets are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$, $|\eta| < 2.4$, and a spatial separation from muon candidates of $\Delta R > 0.5$. In order to reduce the contamination from PU, jets are required to be matched to the same primary vertex as the muon candidate.

The primary background process for the measurement of W+jets at high jet multiplicities (4 or more) is $\mathrm{t\bar{t}}$ production. To reduce the $\mathrm{t\bar{t}}$ contamination, a veto is applied that removes events containing one or more b-tagged jets. The tagging criteria used for this veto are based on the combined secondary vertex algorithm (CSV) [65]. Differences in the b tagging efficiency in data and simulation, as well as differences in mistagging rates, are corrected using scale factors [65] determined as a function of p_{T} in multijet and $\mathrm{t\bar{t}}$ events. Specifically, the tagging efficiency in simulation is decreased by randomly untagging b-tagged jets such that the data and simulated efficiencies are matched. Additionally, a small adjustment to the mistagging rates is performed by randomly tagging untagged jets in simulated events such that the data and simulated mistagging rates agree within uncertainties.

In order to select a $\mathrm{W}(\rightarrow \mu\nu)$+jets sample, events are required to contain exactly one muon satisfying the muon selection criteria described above and one or more jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. Events containing additional muons with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ are vetoed. Events are required to have $M_{\mathrm{T}} > 50\,\mathrm{GeV}$.

4.6.3 Background Estimation

The multijet background is estimated using a control data sample with an inverted muon isolation requirement. In the control data sample, the muon misidentification rate for multijet processes is estimated in a multijet-enriched sideband region with $M_{\mathrm{T}} < 50$ GeV, and the shape template of the multijet distribution is determined in the region with $M_{\mathrm{T}} > 50$ GeV. The muon misidentification rate is then used to rescale the multijet shape template. This method of estimation was used in the measurement of the W+jets production cross section at 7 TeV, and it is described in detail in Ref. [22].

The dominant source of background comes from the $\mathrm{t\bar{t}}$ process, which is reduced by the application of the b jet veto described in Section 4.6.2. For jet multiplicities of 1 to 7, the b jet veto rejects 62–88% of the predicted $\mathrm{t\bar{t}}$ background, while eliminating 4–22% of the predicted W+jets signal.

4.6.4 Measured Observables

Fiducial cross sections are measured as a function of jet multiplicity, inclusively and exclusively, as a function of jet p_{T} and $|\eta|$, and as a function of H_{T}. In terms of angular correlations between jets, cross sections are measured as a function of the difference in rapidity $\Delta y(j_i, j_k)$, and of the difference in azimuthal angle $\Delta\phi(j_i, j_k)$, between the ith and kth jets from the p_{T}-ordered list of jets in the event. Cross sections are also measured as a function of the differences in rapidity and in azimuthal angle between rapidity-ordered jets, most notably $\Delta y(j_F, j_B)$ and $\Delta\phi(j_F, j_B)$, the differences between the most forward and the most backward jet in the event. Cross sections are measured as a function of $\Delta R(j_1, j_2) = \sqrt{\Delta\phi(j_1, j_2)^2 + \Delta y(j_1, j_2)^2}$ between p_{T}-ordered jets. The dependence of the cross section on the invariant mass of the two leading jets for different jet multiplicities is also examined. The difference in azimuthal angle between the muon and the leading jet

is measured for different jet multiplicities. The dependence of $\langle N_{\text{jets}} \rangle$ on H_{T} and on both $\Delta y(j_1, j_2)$ and $\Delta y(j_F, j_B)$ is studied for different jet multiplicities.

Before correcting for detector effects and determining the cross section values, we compare the kinematic distributions reconstructed in data with the predictions for the simulated W+jets signal and the simulated background processes. The comparison of reconstructed data and simulated signal and background processes is shown in Fig. 4.1 for the inclusive jet multiplicity. The uncertainty band represents the total statistical and systematic uncertainty including uncertainties in the jet energy scale and resolution, the muon momentum scale and resolution, the integrated luminosity, the pileup modeling, the normalization of the background processes, the modeling of the W b contribution in the signal simulation, and the reconstruction, identification, and trigger efficiencies.

The number of events in each bin of exclusive reconstructed jet multiplicity for both data and simulated signal and backgrounds is listed in Table 4.4. The predicted total yields agree well with the data yields for all the values of jet multiplicity.

TABLE 4.4: Number of events in data and simulation as a function of exclusive reconstructed jet multiplicity. The purity is the number of simulated signal events (W +jets) divided by the total number of simulated signal and background events (Total). The ratio is the total number of simulated signal and background events divided by the number of data events.

N_{jets}	0	1	2	3	4	5	6	$\geq$7
WW+jets	18 093	24 420	13 472	3057	515	77	12	1
WZ+jets	8125	6799	4153	1042	183	30	4	0
ZZ+jets	932	669	384	96	18	3	0	0
QCD multijet	570 722	228 188	37 154	6734	1076	171	40	9
Single top quark	6438	14 386	9838	3444	877	196	34	7
Z/γ+jets	1 935 191	265 387	51 613	9570	1697	281	48	6
$t\bar{t}$	1504	7576	16 052	17 377	10 090	3487	1000	288
W+jets	54 617 816	6 999 393	1 320 381	222 457	37 822	5857	860	139
Total	57 158 821	7 546 818	1 453 047	263 777	52 278	10 102	1998	450
Purity	0.96	0.93	0.91	0.84	0.72	0.58	0.43	0.31
Data	57 946 098	7 828 967	1 517 517	279 678	54 735	10 810	2058	441
Ratio	0.99	0.96	0.96	0.94	0.96	0.93	0.97	1.02

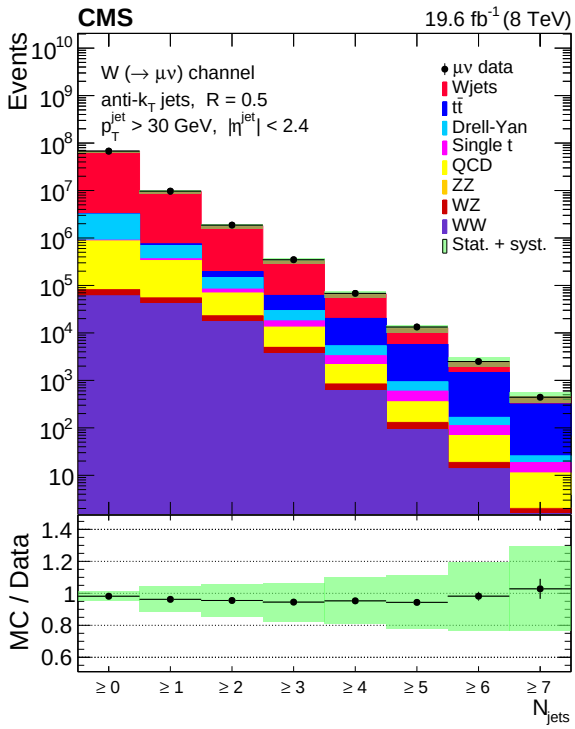


FIGURE 4.1: Distribution of inclusive jet multiplicity, for reconstructed data (points) and simulated signal and backgrounds (histograms). The ratio of simulated and measured data events is shown below the distribution. The data points are shown with statistical error bars. The error band represents the total statistical and systematic uncertainty.

4.6.5 Unfolding

The measured W+jets signal distributions are obtained by subtracting the data by the simulated backgrounds and the estimated QCD multijet background. These background-subtracted distributions are corrected back to the particle level by unfolding discussed in Sec. 4.4.

A response matrix, which defines the event migration probability between the particle-level and reconstructed quantities is constructed using the W+jets MC samples from MADGRAPH5+PYTHIA6, of which both reconstructed and generator levels information are utilized.

The generator-level selection defines the fiducial phase space of the measurements and is identical to the selection applied to the reconstructed objects, including the requirement of exactly one muon with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.1$, jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$, and $M_{\mathrm{T}} > 50\,\mathrm{GeV}$. The generator-level $E_{\mathrm{T}}^{\mathrm{miss}}$ is determined using the neutrino from the decay of the W boson. The momenta of all photons in a cone of $\Delta R < 0.1$ around the muon are added to that of the muon in order to take into account final-state radiation. The generator-level jets are clustered using the anti-k_{T} [58] algorithm with a distance parameter of 0.5. The jet clustering algorithm uses all particles after decay and fragmentation, excluding neutrinos. The b jet veto explained in Section 4.6.2 is treated as an overall event selection requirement, and the cross section is corrected by the unfolding procedure to correspond to W boson production in association with jets of any flavor. The contribution from $\mathrm{W} \to \tau\nu$ decays resulting in a muon in the final state is estimated to be small ($\sim$1% of selected signal sample), and it is therefore not considered as part of the signal definition at the generator level.

4.6.6 Systematic Uncertainties

The sources of systematic uncertainties considered in this analysis are described below.

4.6.6.1 Systematic Variations

Mostly, the systematic uncertainties are obtained by repeating all of the analysis steps (including backgrounds subtraction and unfolding) with systematic variations (up/down) corresponding to their sources. Then, the unfolded results with the variations are compared to the unvaried result and the difference between them is taken to be the estimation of systematic uncertainty. The systematic variations considered are described below. Note that all the jet-related systematic variations are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$.

Jet energy scale correction (JES) This uncertainty is estimated by shifting the momentum p_{T} of the jets in data up and down. The magnitude of the shifting for each individual jet is equal to the uncertainty in the jet energy correction assigned by a p_{T} and η dependent value [61]. These scale factors range from 1% to 5% for jets passing our selection criteria. Scaling the value of p_{T} for each individual jet should affect the $E_{\mathrm{T}}^{\mathrm{miss}}$, therefore the $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated. This affects the value of $M_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ which is also one of the event selection criteria. The difference in the unfolded results (up/down) compared to the unvaried one is the estimation of the systematic uncertainty due to JES.

Jet energy resolution correction (JER) When the measured signal is converted to particle level, the effect of energy resolution of the detector is taken care of by the unfolding. Before doing so, MC backgrounds are subtracted from data, and a response matrix is constructed with the MC signal sample. Since energy resolution in the reconstructed data and MC are different, specifically it is worse in data than in MC, the jets in MC need to be smeared to describe the data. The smearing is done for each generator-level jet, with a matched reconstructed jet being defined as the closest reconstructed jet within a cone of $\Delta R < 0.5$. Then, the difference in jet p_{T} between the matched reconstructed jet and the generator-level jet is increased by an η dependent scale factor [61]. The uncertainty on these factors need to be considered. The effect of this is assessed by scaling the jets in W+jets sample with two additional sets of scale factors that correspond to varying the factors up and down by one sigma and evaluating the impact of these new sets. As in the case of JES, the changes in jet p_{T} are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$.

Background cross sections (BG) This systematic uncertainty contribution is determined by varying the background cross sections within their uncertainty. All backgrounds are varied up (down) simultaneously, instead of individually. The uncertainties in the cross section are theoretically predicted [84, 85, 86, 87, 93]. They are 7% for $t\bar{t}$, 6% for ZZ and

WZ, 8% for WW, and 4% for Z+jets for the region $M_{\mu\mu} > 50$ GeV. For single-top processes, the uncertainties are 6% for s- and t-channel and 9% for the tW-channel.

Pile up (PU) The uncertainty in the modeling of PU in simulation is assigned by varying the inelastic cross section by ±5% [94].

b-tagging efficiency (b tagging) The uncertainties on the data/MC scale factors of the b-tagging efficiencies are considered. As described before, the nominal values of the scale factor are applied to MC in the standard analysis in order to correct for the difference between the b-tagging efficiencies of data and MC. Therefore, this systematic uncertainty is assessed by adjusting the scale factors up and down according to their uncertainties found in the official payload [65]. The entire analysis is performed with these variations and the final unfolded results are compared to the results of the standard analysis. The difference is taken to be the systematic uncertainty.

Muon identification, isolation, and trigger (LepSF) The systematic uncertainties on the data/MC scale factors for muon identification, isolation, and trigger efficiency corrections are considered. This is studied by simultaneously varying the scale factors for muon identification/isolation/trigger according to the recommended uncertainties. The uncertainties are generally less than 3%.

Muon momentum scale (MES) This uncertainty is studied by varying the reconstructed p_{T} of muons in MC up and down by 0.2%. The modification is performed on all muons passing all identification criteria except for the requirement that $p_{\mathrm{T}} > 25$ GeV. The value of 0.2% is the conservative value for muons with $p_{\mathrm{T}} < 200$ GeV (without a correction) [54]. Since the effect of these variations comes from the muons at the 25 GeV threshold,

using this conservative value is sufficient to assess this uncertainty. The resulting uncertainty is very small, ranging from 0.049% to 1.5% in jet multiplicity distribution.

Muon momentum resolution (MER) This uncertainty is assessed by smearing the momentum of muons in MC by a nominal value of 0.6% [54], i.e. the reconstructed p_{T} is simply modified by a random Gaussian value with a width of 0.006 $\times p_{\mathrm{T}}$. The uncertainty in muon momentum leads to a very small effect (mostly $< 1\%$).

4.6.6.2 Other Sources of Systematic Uncertainties

Other sources of systematic uncertainties include the finite number of MC events in the unfolding response matrix (MC), different generator used to construct the response matrix for unfolding, and the uncertainty in the integrated luminosity.

Response statistics (MC) The unfolding implemented in RooUnfold does not take account the uncertainty due to finite statistics of the MC used in constructing the response matrix. This uncertainty is obtained by doing toy MC experiments and randomly smearing each of the bin of the response matrix within its statistical uncertainty. Two thousand randomly smeared matrices are used to obtain a set of 2000 unfolded results. The standard deviation of the results is the estimation of the uncertainty.

Unfolding using reweighed response matrix (Resp) The use of MadGraph MC in filling the response matrix might result in a bias since it does not reproduce the shape of observed data distributions well and might therefore biased the unfolded result towards the generator predictions Note that this uncertainty is not applicable to jet multiplicity distribution. The uncertainty is assessed by weighting simulation to match the data in order to produce a reweighted response matrix. The procedure is described below.

In addition to the nominal binning, a finer binned histogram is filled. Using the fine binning, the weight factor f_{w_i} for each bin is calculated as the ratio between the pre-unfolded data and the reconstructed MC prediction

$$f_{w_i} = \frac{N_{Data_i}}{N_{MC_i}} \times \frac{N_{MC}}{N_{Data}}$$

where N_{Data_i} is the number of events in bin i of the background-subtracted data distribution, N_{MC_i} is the number of events in bin i of the Monte Carlo, and N_{Data} and N_{MC} are the total number of events in the corresponding distributions.

Then, each element of the fine-binned response matrix is multiplied by f_{w_i}

$$R_{new\,ij} = f_{w_i} \times R_{ij}$$

where index i (j) refers to the reconstructed level index (generated level index).

This reweighed response matrix is then re-binned to the nominal binning and it is used to unfold the background-subtracted data. Finally, the difference between the unfolded result obtained using reweighed response matrix and the one obtained using the main method is an estimation of systematic uncertainty.

$t\bar{t}$ background modeling The uncertainty in the $t\bar{t}$ modeling is assessed by studying a $t\bar{t}$-enriched control region created by removing the b jet veto and instead requiring two or more b-tagged jets. Data-driven scale factor which rescales the Monte Carlo to match the data can be obtained from this control region using the formula:

$$SF_i = \frac{N_{Data_i} - N_{otherMC_i}}{N_{t\bar{t}_i}}$$

where $N_{otherMC_i}$ is the sum of QCD multijet estimation and all Monte Carlo estimations (W+jets signal and backgrounds) except for $t\bar{t}$.

The scale factors are applied to $t\bar{t}$ in the signal region and the difference between the unfolded results with and without $t\bar{t}$ rescaling is taken as an the uncertainty on the modeling of $t\bar{t}$. This study is not relevant for the 1-jet bin, since 2 bjets are required, but $t\bar{t}$ is not a strong contribution in the 1-jet bin.

The normalization of Wb This uncertainty is due to the normalization of the Wb content in the simulation. First, the agreement between data and simulation in a control region defined by requiring exactly one b-tagged jet is investigated. The normalization of the Wb process is found to be underestimated by a factor of 1.3. Then, in the signal region, we reweigh the Wb events in the W+jets MC by this factor. The difference between the unfolded results with and without reweighing the Wb events is taken as the estimate of uncertainty on the modeling of Wb.

Integrated luminosity We assign a 2.6 % uncertainty on the measured integrated luminosity [95].

4.6.6.3 Statistical uncertainty

The statistical uncertainty on the background-subtracted data is propagated through the unfolding procedure. The calculation described on Section 4.6.5 is automatically handled by RooUnfold. This is cross-checked by a set of pseudo experiments that repeat the unfolding procedure. In the pseudo experiments we generated 2000 randomly smeared background-subtracted data distributions. The smearing uses a Gaussian function with a standard deviation equal to the uncertainty of the original distributions. These are unfolded. The standard deviations for every bin of the unfolded results are calculated and compared with

the uncertainties obtained from RooUnfold. The comparison shows that two methods of handling the statistical uncertainty are equivalent.

The effect of the systematic uncertainties in the measured cross section as a function of jet multiplicity is illustrated in Fig. 4.2, and in Table 4.5 for jet multiplicities of 0–7. The statistical uncertainty is also included. The total uncertainty is the sum in quadrature of all contributions.

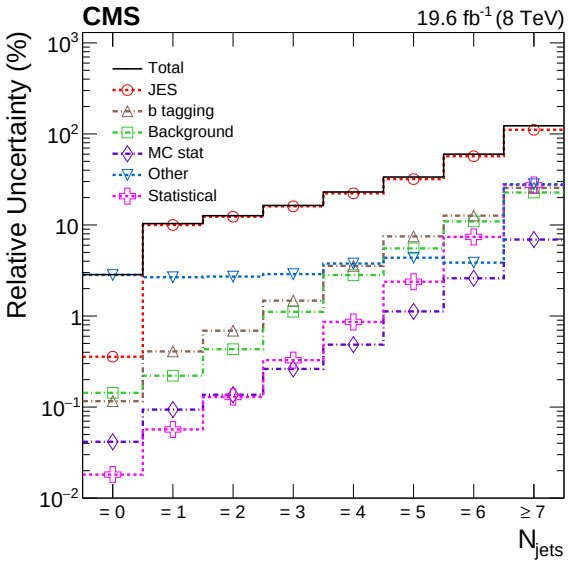


FIGURE 4.2: Systematic uncertainties in the measured cross section as a function of jet multiplicity, including uncertainties due to jet energy scale (JES), background normalization, b tagging efficiency, finite number of simulated events used to construct the response matrix (MC stat), and other systematic uncertainties mentioned in Section 4.6.6. The largest contribution to the other uncertainties is the uncertainty on the integrated luminosity, which is 2.6%. Statistical uncertainty of the unfolded data and total uncertainty in the measured cross section are also shown.

4.6.7 Results

The measured W($\rightarrow \mu\nu$)+jets fiducial cross sections are shown in Figs. 4.3–4.15 and compared to the following theoretical predictions.

TABLE 4.5: Uncertainties in the measured cross section as a function of jet multiplicity, including uncertainties due to the statistical uncertainty of unfolded data (Stat), the jet energy scale (JES), pileup modeling (PU), background normalization (BG), the jet energy resolution (JER), trigger efficiency and muon identification (LepSF), b tagging efficiency, muon momentum scale (MES) and resolution (MER), the normalization of the Wb content in the signal simulation (Wb), the $t\bar{t}$ modeling, a finite number of simulated events used to construct the response matrix (MC stat), and integrated luminosity (Int Lumi).

	$N_{\text{jets}} = 1$	$N_{\text{jets}} = 2$	$N_{\text{jets}} = 3$
Total(%)	10	13	16
Stat(%)	0.057	0.13	0.33
JES(%)	10	12	16
PU(%)	0.025	0.26	0.35
BG(%)	0.22	0.43	1.1
JER(%)	0.43	0.23	0.29
LepSF(%)	0.35	0.50	0.72
b tagging(%)	0.41	0.69	1.5
MES(%)	0.20	0.18	0.17
MER(%)	0.015	0.0016	0.017
Wb(%)	0.062	0.22	0.38
$t\bar{t}$ (%)	0.014	0.38	0.83
MC stat(%)	0.094	0.14	0.26
Int Lumi(%)	2.6	2.6	2.6

- **MADGRAPH5+PYTHIA6 :** LO MC generator (described in Section 4.6.1.2). This is also the MC used to derive the response matrix.

- **MADGRAPH5_AMC@NLO+PYTHIA8 :** This is an NLO prediction provided by MADGRAPH5_AMC@NLO version 2.2.1 [72], a MC generator with up to three final-state partons, with ME computation for up to two jets at NLO accuracy, which uses the NNPDF3.0 PDF set [96]. The generator is interfaced with PYTHIA 8 [97] for parton showering and hadronization. The merging of parton shower and ME is done with the FxFx merging scheme [51] and the merging scale is set at 30 GeV. The NNPDF2.3 PDF set [98] and the CUETP8M1 tune [99] are used in PYTHIA8. Using the weighting methods available in the generator [100], PDF and scale uncertainties are assigned to the MADGRAPH5_AMC@NLO+PYTHIA8 predictions by considering the NNPDF3.0 PDF uncertainties, and by independently varying the factorization and renormalization

scales by a factor of 0.5 or 2, excluding the combinations where one scale is varied by a factor of 0.5 and the other one by a factor of 2. The prediction is denoted as MG5_aMC+PY8 in the figure legends.

- **SHERPA 2 :** This is an NLO prediction provided by SHERPA version 2.1.1, a multileg NLO MC generator with parton showering interfaced with BLACKHAT [101, 102] for the one-loop corrections. This sample of events is produced with the CT10 PDF set. The corresponding sample is denoted SHERPA 2. The SHERPA 2 matrix element calculations include the production of up to four parton jets, with NLO accuracy for up to two jets and LO accuracy for three and four jets. The merging of parton showers and MEs is done with the MEPS@NLO method [103, 104] and the merging scale set at 20 GeV. The predictions from MADGRAPH5+PYTHIA6 and SHERPA 2 are shown with statistical uncertainties only.

- **BLACKHAT+SHERPA:** [105] The calculation yields fixed-order NLO predictions for 8 TeV W + n jets at the level of ME partons, where n = 1–4. The choice of renormalization and factorization scales for BLACKHAT+SHERPA is $\hat{H}'_{\mathrm{T}}/2$, where $\hat{H}'_{\mathrm{T}} = \sum_m p_{\mathrm{T}}^m + E_{\mathrm{T}}^{\mathrm{W}}$, the sum running over final-state partons, and $E_{\mathrm{T}}^{\mathrm{W}}$ being the transverse energy of the W boson. A nonperturbative correction is applied to the BLACKHAT+SHERPA distributions to account for the effects of multiple-parton interactions and hadronization. This correction is determined with MADGRAPH 5.1.1 interfaced with PYTHIA 6.426 with and without hadronization and multiple-parton interactions. The nonperturbative correction factor is mostly in the range 0.90–1.20. A PDF uncertainty is assigned to the predictions of BLACKHAT+SHERPA by considering the error sets of CT10 PDFs. A factorization and renormalization scale uncertainty is also assigned to BLACKHAT+SHERPA predictions, as determined by varying the scales simultaneously by a factor of 0.5 or 2.

- **N_{jetti} NNLO :** [77, 78] An NNLO calculation of W +jet production in perturbative QCD (N_{jetti} NNLO) is also used for comparisons with certain measured distributions (leading jet p_{T}, H_{T}, and $|\eta|$, Figs. 4.4, 4.5, and 4.8) for $N_{\text{jets}} \geq 1$. The CT14 NNLO PDF set is used in the calculation. A nonperturbative correction is applied to this prediction, as in the case of BlackHat+sherpa, as well as an additional correction factor of about 1.01 due to the effect of final-state radiation from the muon. A factorization and renormalization scale uncertainty is assigned to this prediction, as determined by varying the central scale $\sqrt{m_{\ell\nu}^2 + \left(\sum_{\text{jet}} p_{\text{T}}^{\text{jet}}\right)^2}$ by a factor of 0.5 or 2.

The 8 TeV data sample allows us to determine the cross sections for jet multiplicities up to 7 and to study the fiducial cross sections as functions of most kinematic observables for up to four jets.

Jet multiplicity

The measured exclusive and inclusive jet multiplicity distributions, shown in Fig. 4.3, are in agreement with the predictions of the MadGraph5_aMC@NLO+Pythia8 generators and with the calculation of BlackHat+sherpa. For multiplicities above 5, sherpa 2 starts to deviate upward from the measurement.

Jet p_{T} and H_{T}

The cross sections differential in jet p_{T} for inclusive jet multiplicities from 1 to 4 are shown in Fig. 4.4. The jet p_{T} and H_{T} distributions are sensitive to the effects of higher-order processes. The current results extend to 1.0 and 1.5 TeV in the leading-jet p_{T} and H_{T} distributions, respectively, for at least one jet. The predictions from BlackHat+sherpa

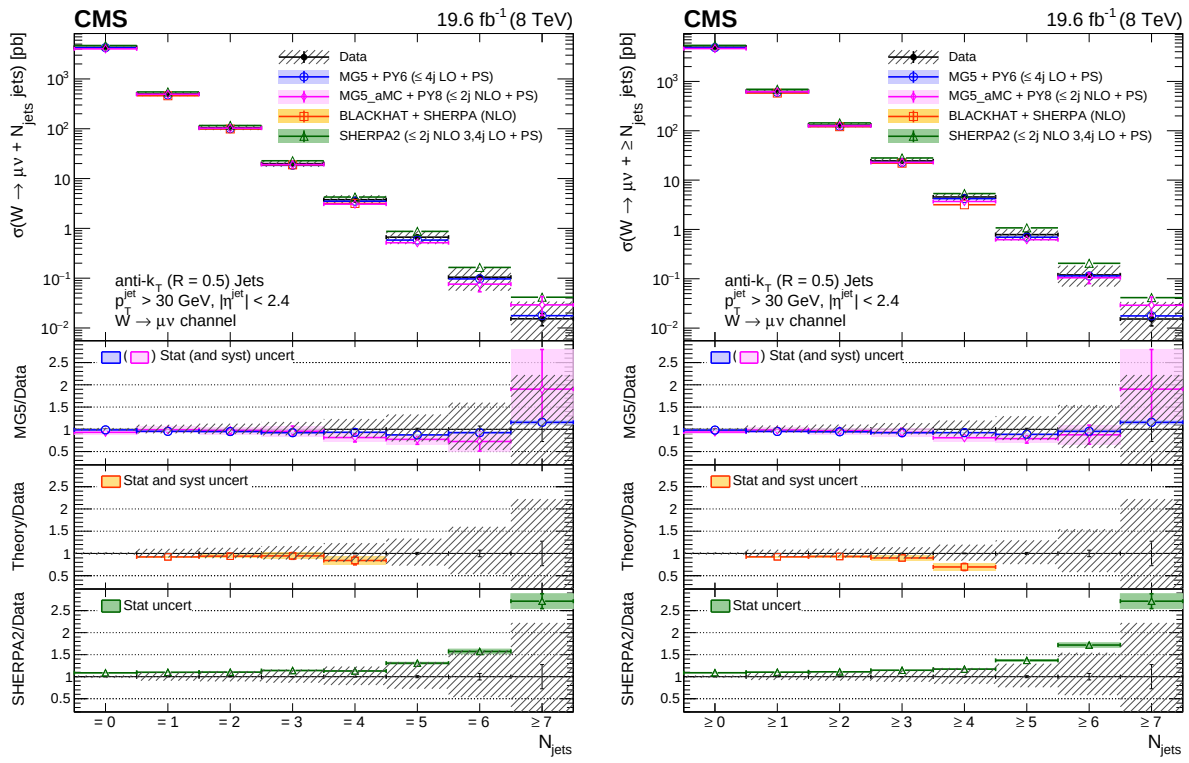


FIGURE 4.3: Measured cross section versus exclusive (left) and inclusive (right) jet multiplicity, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions), for which we currently have predictions only up to W + 4 jets. Black circular markers with the gray hatched band represent the unfolded data measurement and its total uncertainty. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

(jets 1 through 4) are in agreement with the measured distributions within the systematic uncertainties. The predictions from MADGRAPH5+PYTHIA6 show reasonable agreement with data, with the largest discrepancy being an overestimate of up to 20% for the leading and second-leading jet p_{T} distributions in the intermediate-p_{T} region. In comparison to the corresponding measurements of the leading and second-leading jet p_{T} spectra made by CMS with 7 TeV data [22], we observe a smaller slope in the ratio of the MADGRAPH5+PYTHIA6 prediction to the measurement. The predictions from MADGRAPH5_aMC@NLO+PYTHIA8 are in agreement with data within uncertainties. The NNLO prediction for at least one jet agrees with the unfolded jet p_{T} cross section within the systematic uncertainties. At low p_{T} values (below 50 GeV), the predictions for the first-, second-, and third-leading jet p_{T} from

SHERPA 2 overestimate the data.

The H_{T} distributions for inclusive jet multiplicities of 1 to 4 are shown in Fig. 4.5. The H_{T} distributions are best modeled by the NNLO prediction for an inclusive jet multiplicity of 1, and by MADGRAPH5+PYTHIA6 and MADGRAPH5_aMC@NLO+PYTHIA8 for inclusive jet multiplicities of 1 and 2. For higher jet multiplicities, the MADGRAPH5+PYTHIA6 and MADGRAPH5_aMC@NLO+PYTHIA8 predictions underestimate the data at low values of H_{T} (below 200 GeV). The SHERPA 2 predictions for H_{T} consistently overestimate the data for all inclusive jet multiplicities and display a harder H_{T} spectrum. The BLACKHAT+SHERPA prediction underestimates the data H_{T} distribution for $N_{\mathrm{jets}} \geq 1$, as expected because the NLO prediction for H_{T} for $N_{\mathrm{jets}} \geq 1$ is a fixed-order prediction with up to two real partons, and contributions from higher jet multiplicities are missing [106].

Dijet p_{T} and $M(j_1, j_2)$

The dijet p_{T} and dijet invariant mass spectra for inclusive jet multiplicities of 2, 3, and 4 are shown in Figs. 4.6 and 4.7. Dijet quantities are based on the two leading jets in the event, and they constitute an important test of the modeling of p_{T} correlations among jets, whose correct accounting is crucial for searches for physics beyond the SM in dijet final states. All of the predictions agree reasonably well with data, but SHERPA 2 consistently overestimates the data for high values of dijet p_{T} and invariant mass, particularly in the dijet p_{T} spectrum for $N_{\mathrm{jets}} \geq 2$. The MADGRAPH5_aMC@NLO+PYTHIA8 prediction also underestimates the data for values of the invariant mass below 200 GeV in the inclusive four-jet distribution.

Jet pseudorapidity $|\eta|$

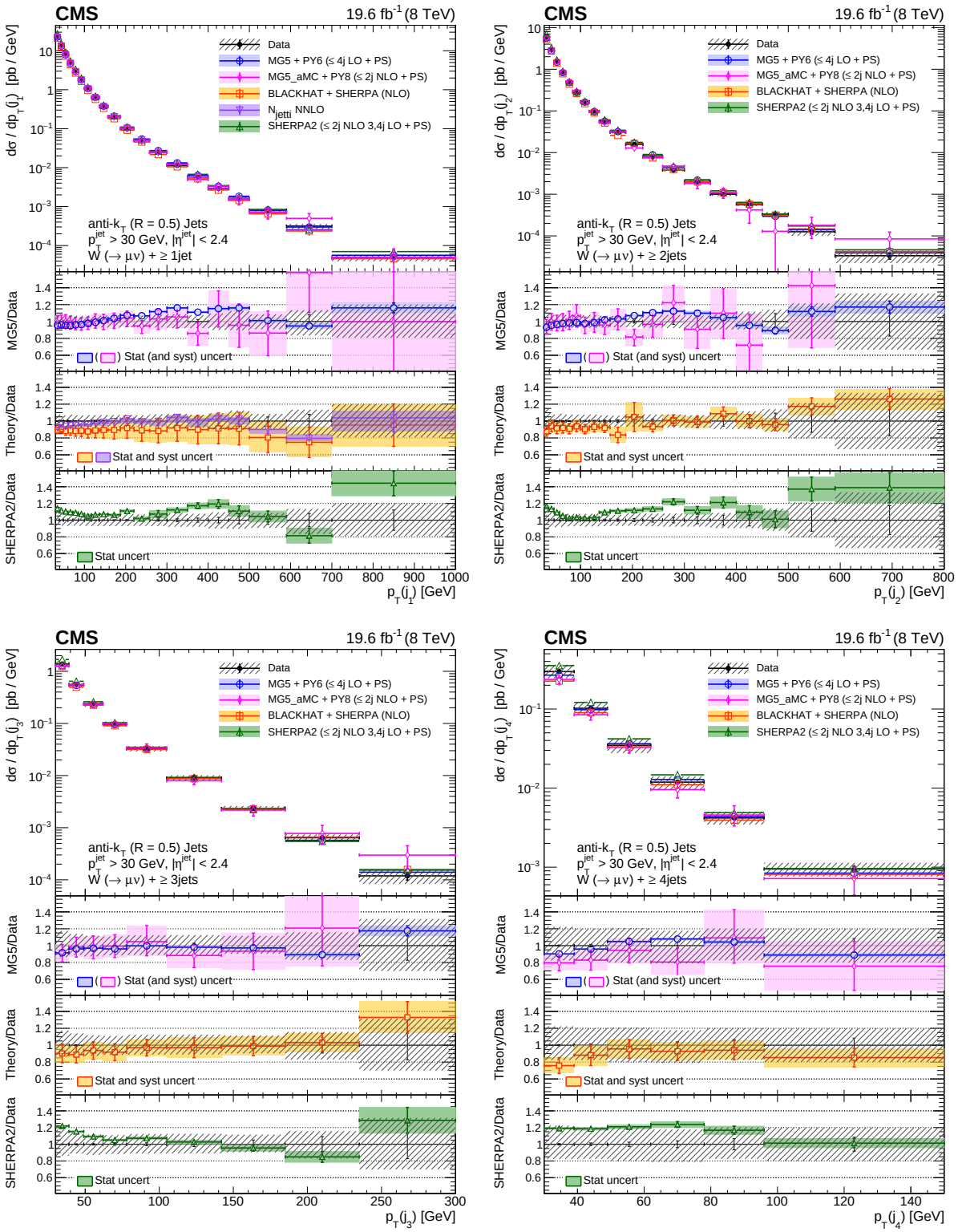


FIGURE 4.4: Cross sections differential in the transverse momenta of the four leading jets, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, BLACKHAT+SHERPA, and NNLO inclusive one-jet production (indicated as N_{jetti} NNLO). The BLACKHAT+SHERPA and NNLO predictions are corrected for hadronization and multiple-parton interaction effects. Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

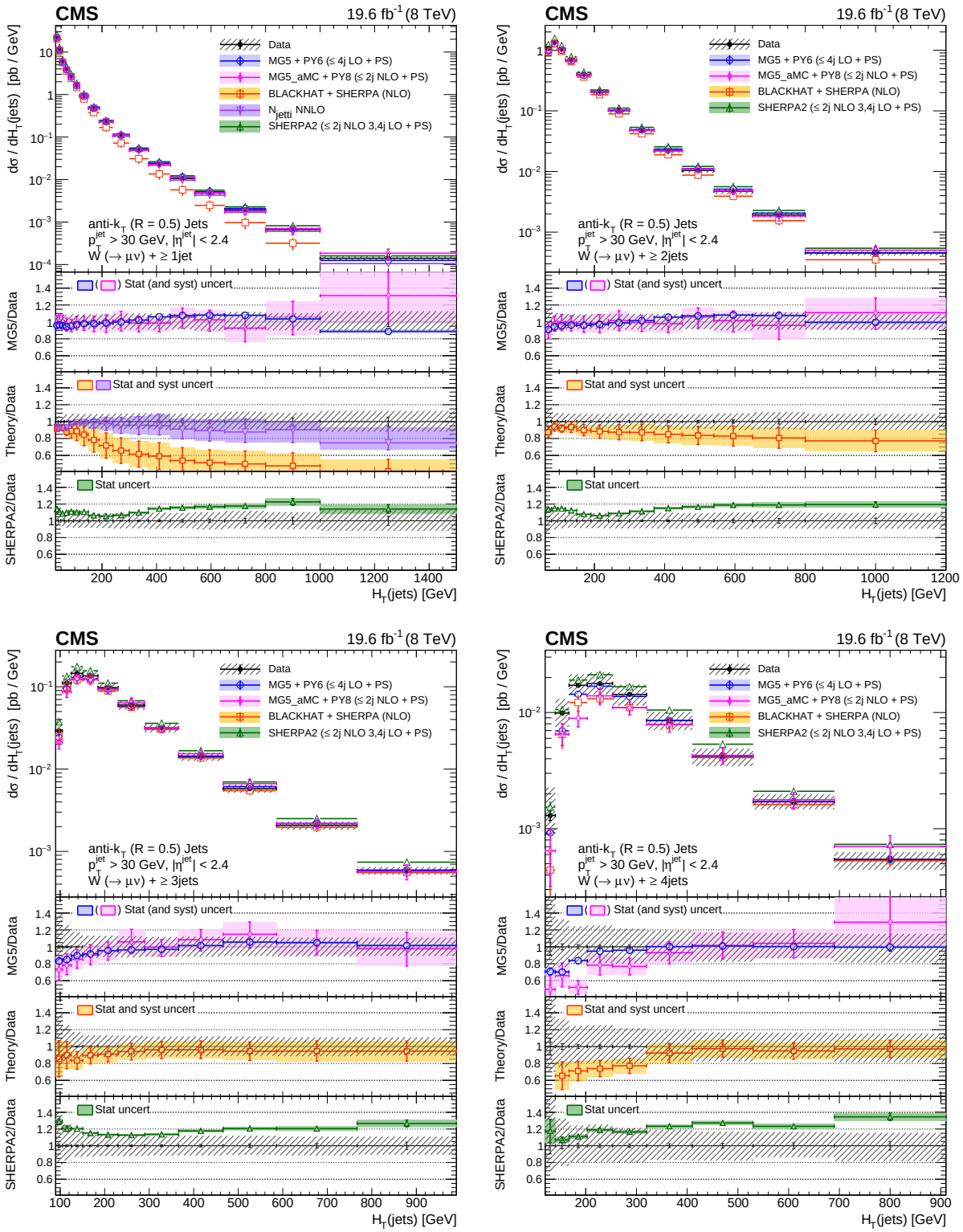


FIGURE 4.5: Cross sections differential in H_T for inclusive jet multiplicities 1–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, BLACKHAT+SHERPA, and NNLO inclusive one-jet production (indicated as N_{jetti} NNLO). The BLACKHAT+SHERPA and NNLO predictions are corrected for hadronization and multiple-parton interaction effects. Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

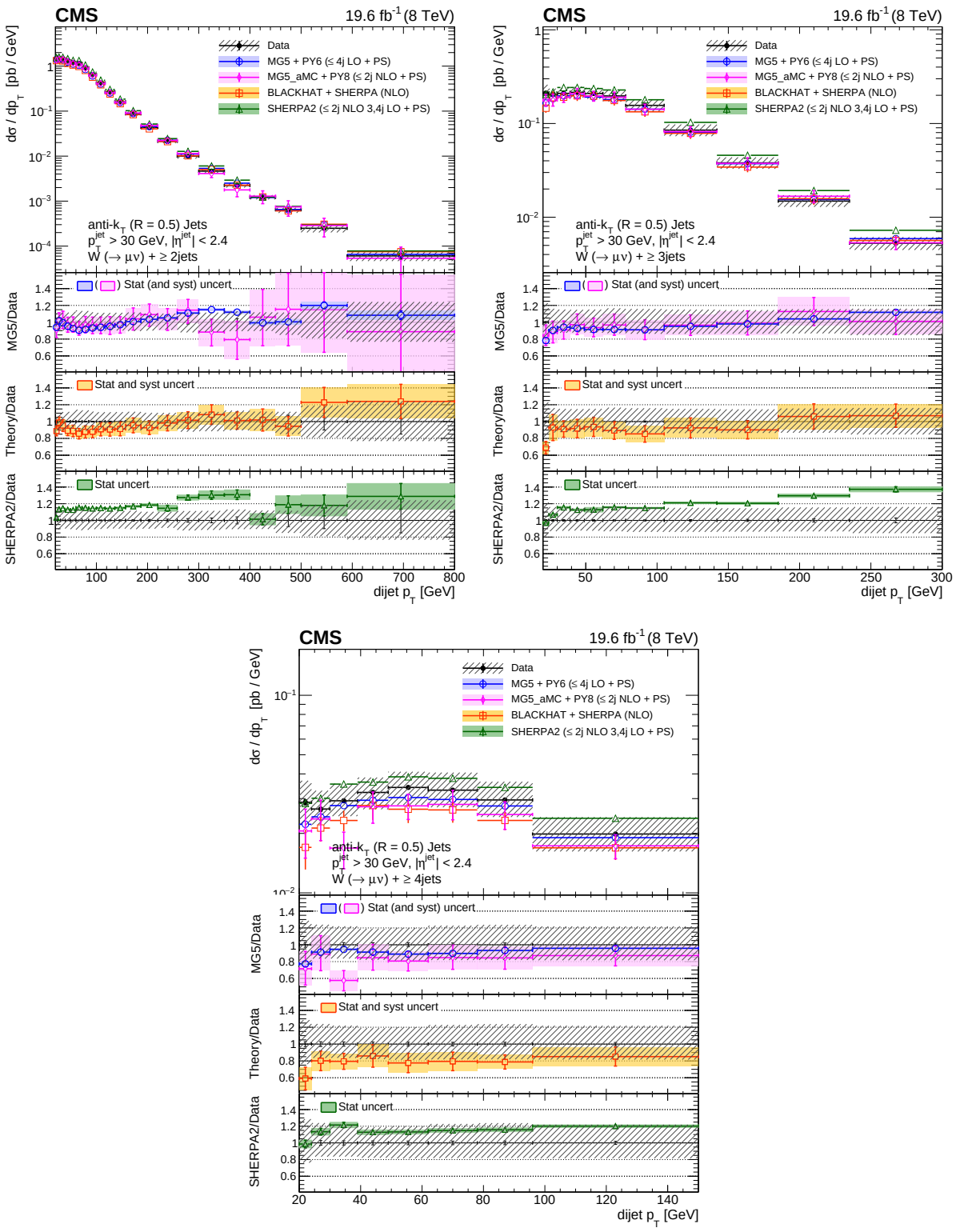


FIGURE 4.6: Cross sections differential in dijet p_{T} (calculated from the two leading jets) for inclusive jet multiplicities 2–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

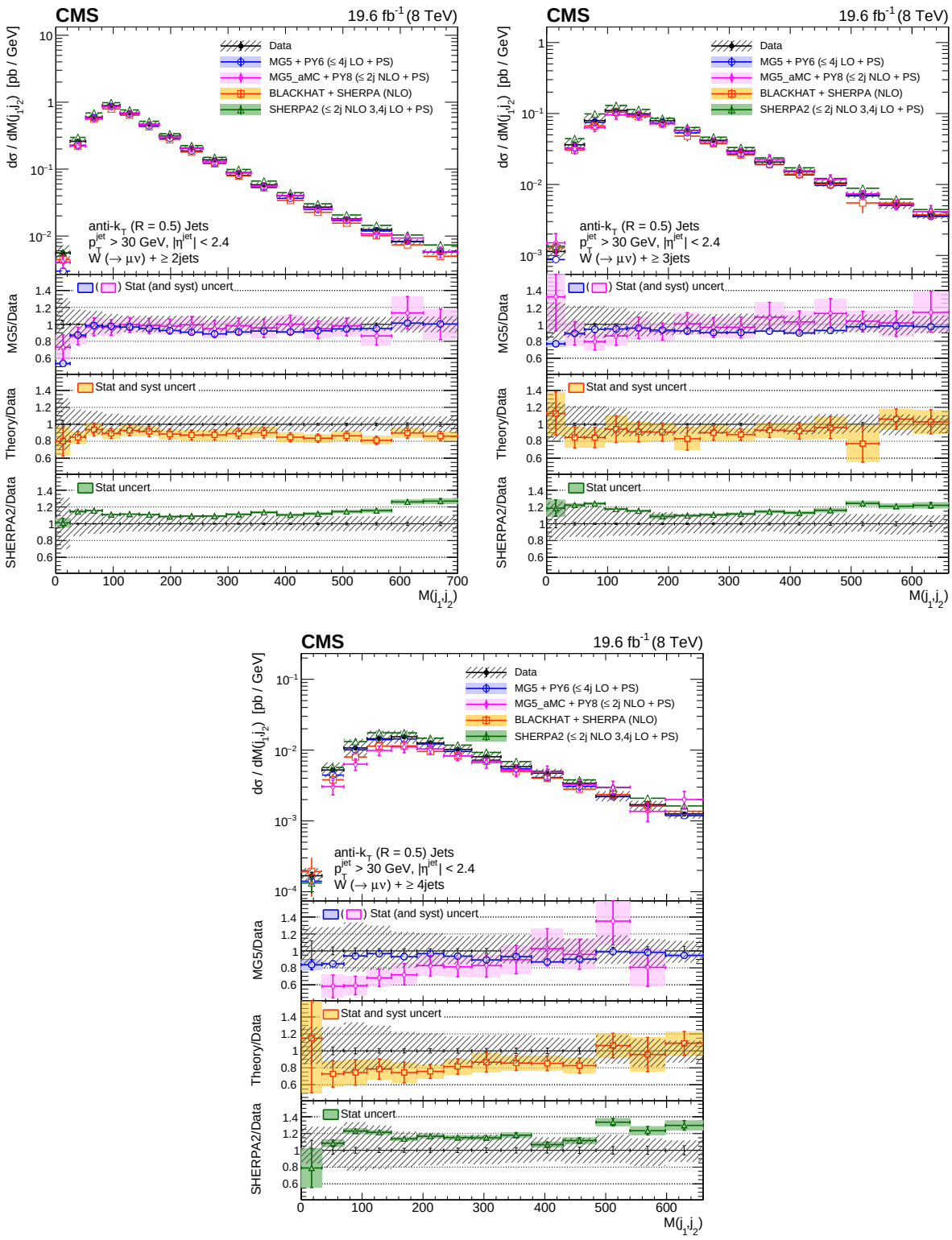


FIGURE 4.7: Cross sections differential in dijet invariant mass (calculated from the two leading jets) for inclusive jet multiplicities 2–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

The pseudorapidity distributions for the four leading jets in each event are shown in Fig. 4.8. The cross sections are best predicted by MadGraph5+pythia6 and MadGraph5_aMC@NLO +pythia8. All predictions agree with the data, with some variations in the overall normalization and a slight underestimation for large values of $|\eta|$.

The angular correlation between jets $\Delta y(j_n, j_m)$, $\Delta\phi(j_n, j_m)$, and $\Delta R(j_1, j_2)$

The distributions of the rapidity difference and the azimuthal angles between p_{T}-ordered and rapidity-ordered jets are shown in Figs. 4.9–4.12. The measurement of the rapidity difference between p_{T}-ordered jets is shown for different jet pairings: the two leading jets $\Delta y(j_1, j_2)$ and the first- (second-) and third-leading jets $\Delta y(j_1, j_3)$ ($\Delta y(j_2, j_3)$). The measurement of the rapidity difference between rapidity-ordered jets makes use of the most forward and most backward jets, $\Delta y(j_{\mathrm{F}}, j_{\mathrm{B}})$. The quantities $\Delta y(j_1, j_2)$ and $\Delta y(j_{\mathrm{F}}, j_{\mathrm{B}})$ are studied for inclusive jet multiplicities of 2 to 4, while $\Delta y(j_1, j_3)$ and $\Delta y(j_2, j_3)$ are studied for $N_{\mathrm{jets}} \geq 3$. A study of the rapidity difference between the two leading jets is helpful in testing the wide-angle soft parton radiation and the implementation of parton showering. The measurement of the rapidity differences between the forward/backward jets is also instrumental in understanding QCD radiation and wide-angle particle emission. The distribution of the azimuthal angle difference is sensitive to higher-order processes and is shown for p_{T}-ordered and rapidity-ordered jets for an inclusive multiplicity of 2. Overall, the predicted distributions of the rapidity difference between p_{T}-ordered jets are in agreement with the measurements, with MadGraph5+pythia6 and BlackHat+sherpa underestimating the data for $|\Delta y|$ values above 2. A similar observation can be made for the rapidity difference between the most forward and most backward jets. This behavior is also reflected in the $\Delta R(j_1, j_2)$ measurement, shown in Fig. 4.13. All predictions for the azimuthal angle difference between jets are in agreement with data, with some variations in the overall normalization.

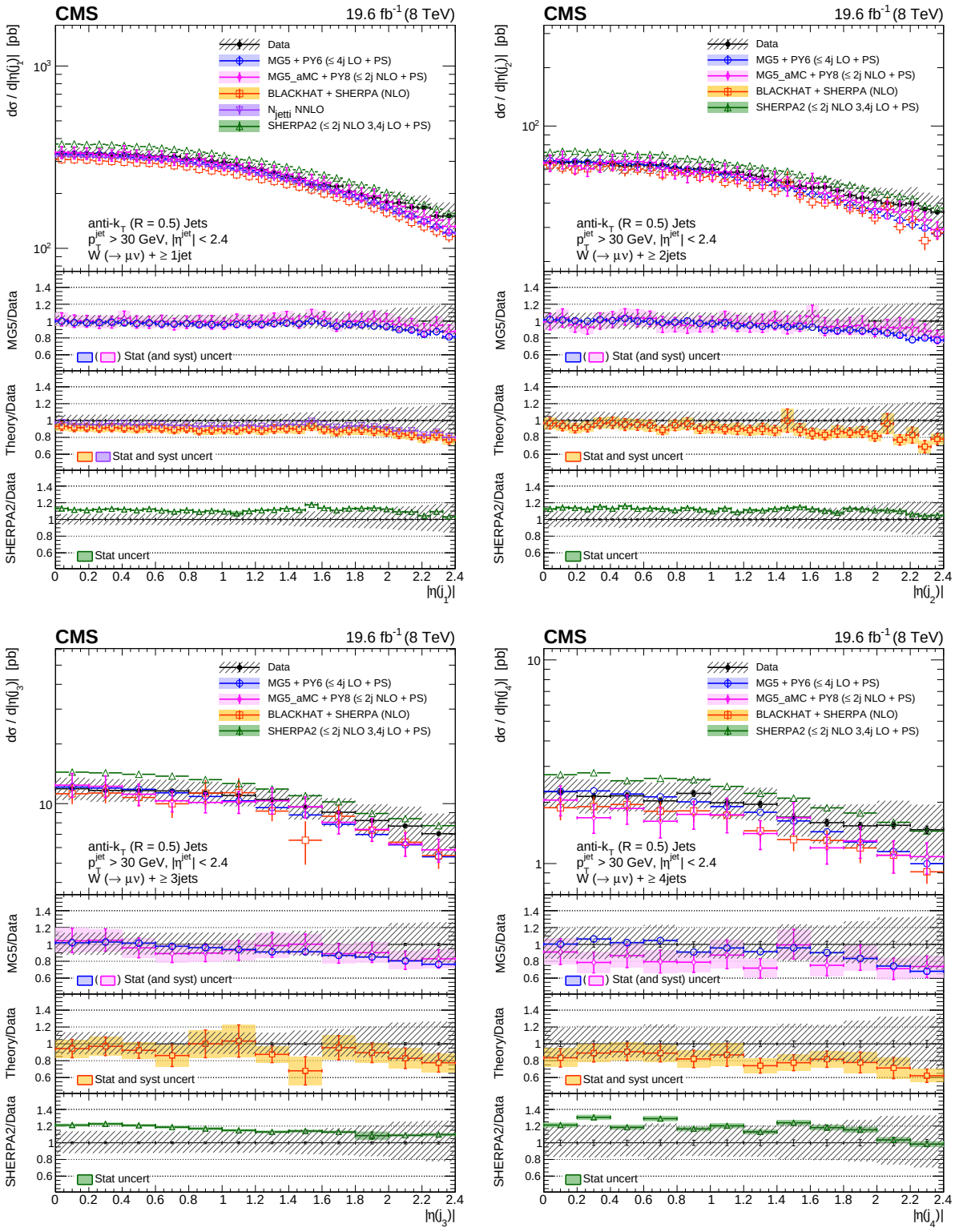


FIGURE 4.8: Cross sections differential in the pseudorapidities of the four leading jets, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, BLACKHAT+SHERPA, and NNLO inclusive one-jet production (indicated as N_{jetti} NNLO). The BLACKHAT+SHERPA and NNLO predictions are corrected for hadronization and multiple-parton interaction effects. Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

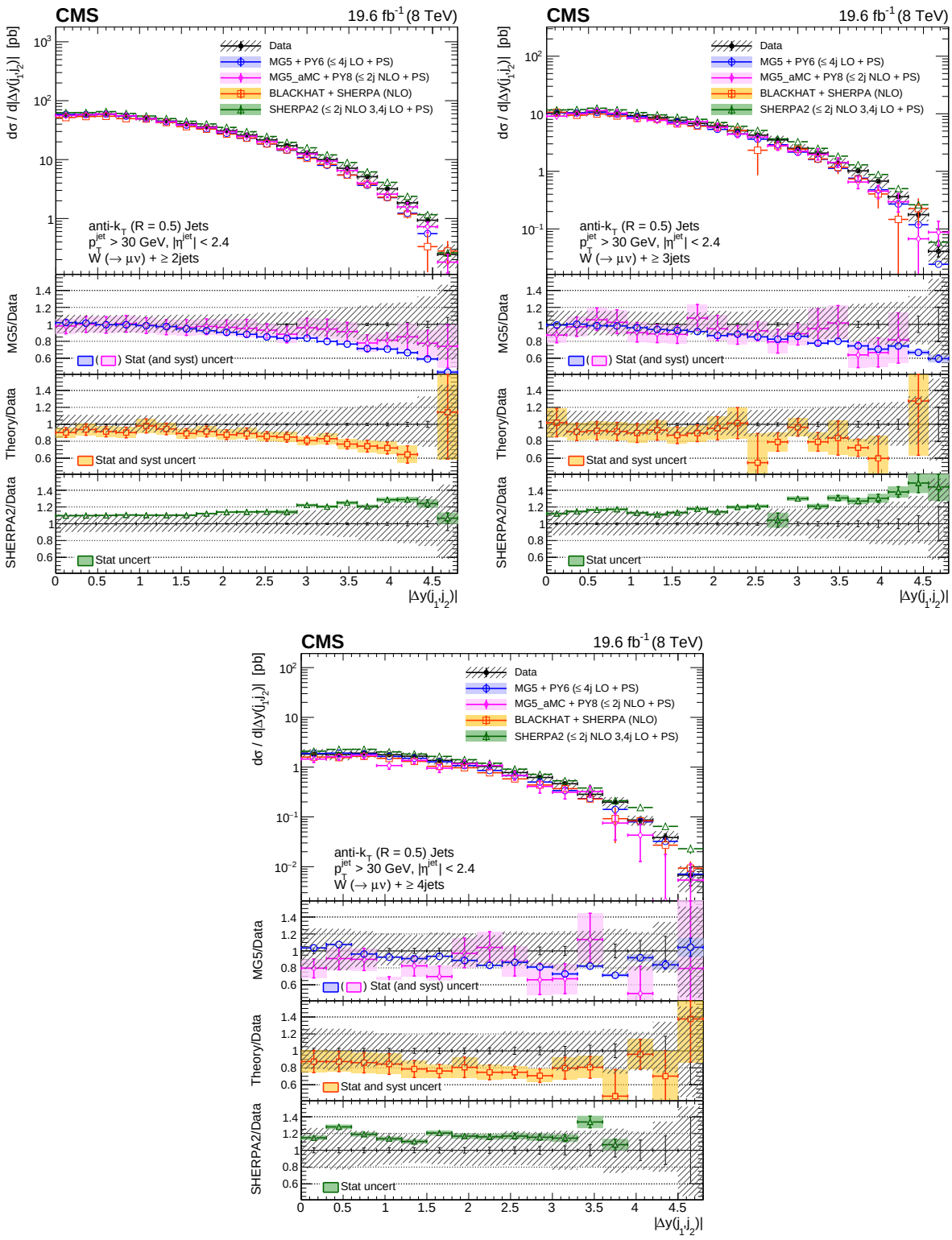


FIGURE 4.9: Cross sections differential in $\Delta y(j_1, j_2)$ for inclusive jet multiplicities 2–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

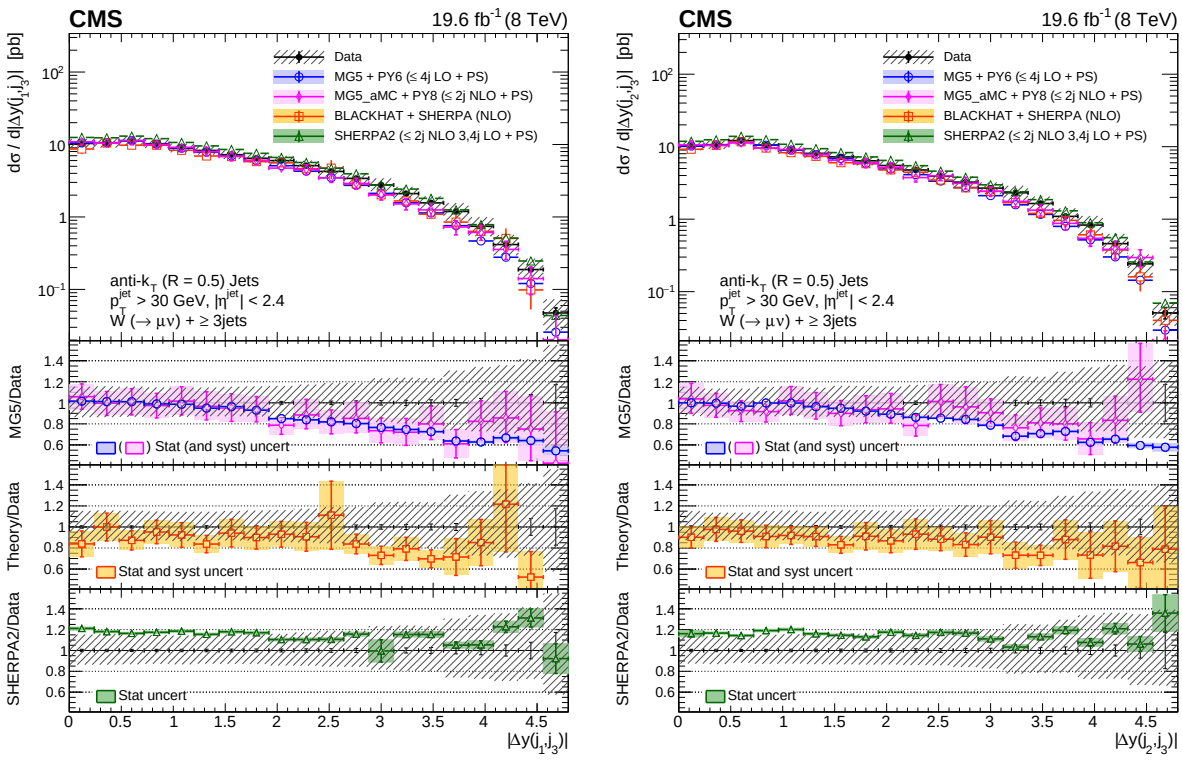


FIGURE 4.10: Cross sections differential in $\Delta y(j_1, j_3)$ (left) and $\Delta y(j_2, j_3)$ (right) for an inclusive jet multiplicity of 3, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

The azimuthal angle between the muon and jet $\Delta\phi(j_n, \mu)$

The distributions of the azimuthal angle between the muon and the nth leading jet, for inclusive jet multiplicities 1 to 4, are shown in Fig. 4.14. Overall, the predictions are in agreement with the measurements, except for BLACKHAT+SHERPA, which disagrees with the data at low values of the $\Delta\phi$ for an inclusive jet multiplicity of 1.

The average number of jets $\langle N_{\mathbf{jets}} \rangle$

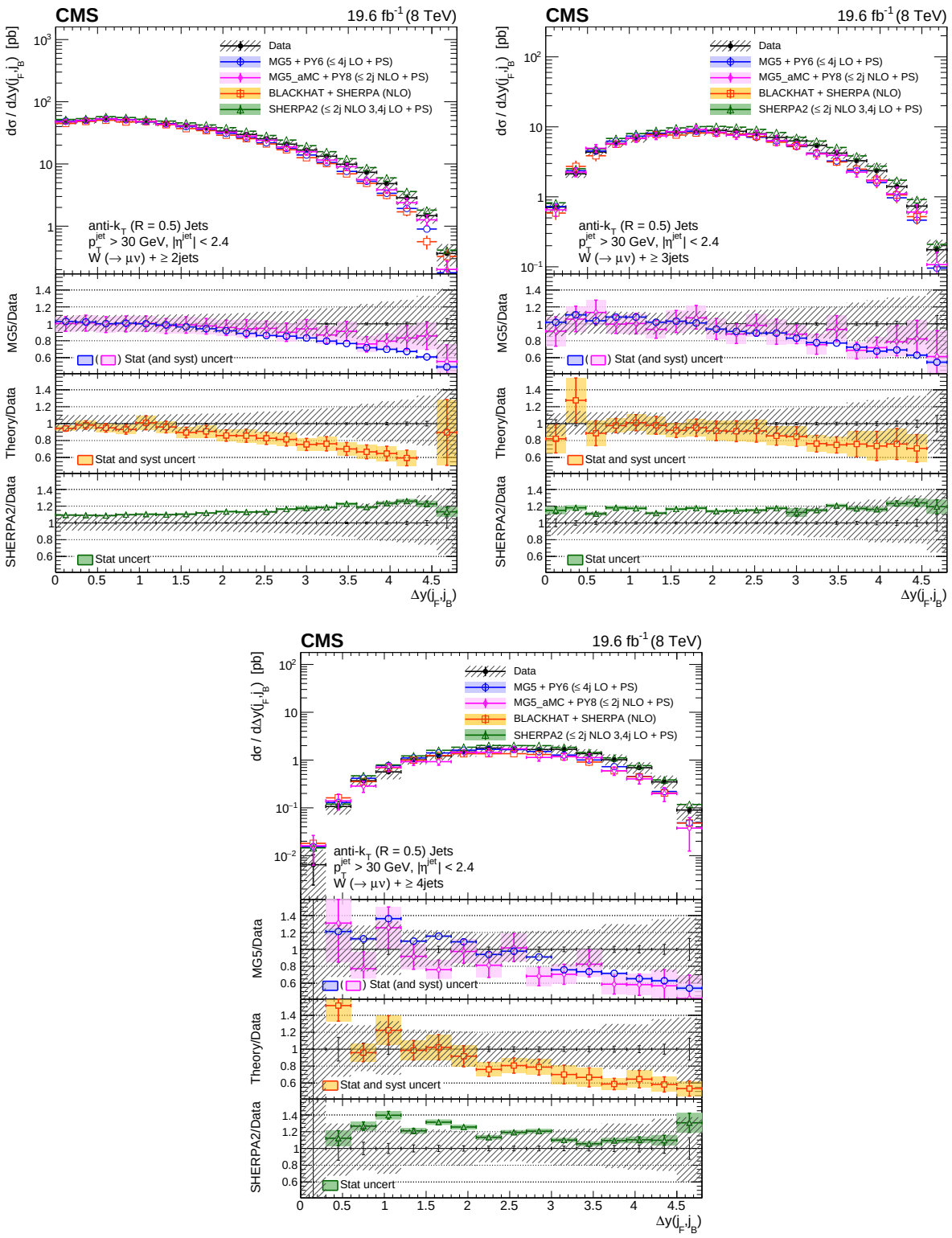


FIGURE 4.11: Cross sections differential in $\Delta y(j_F, j_B)$ for inclusive jet multiplicities 2–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

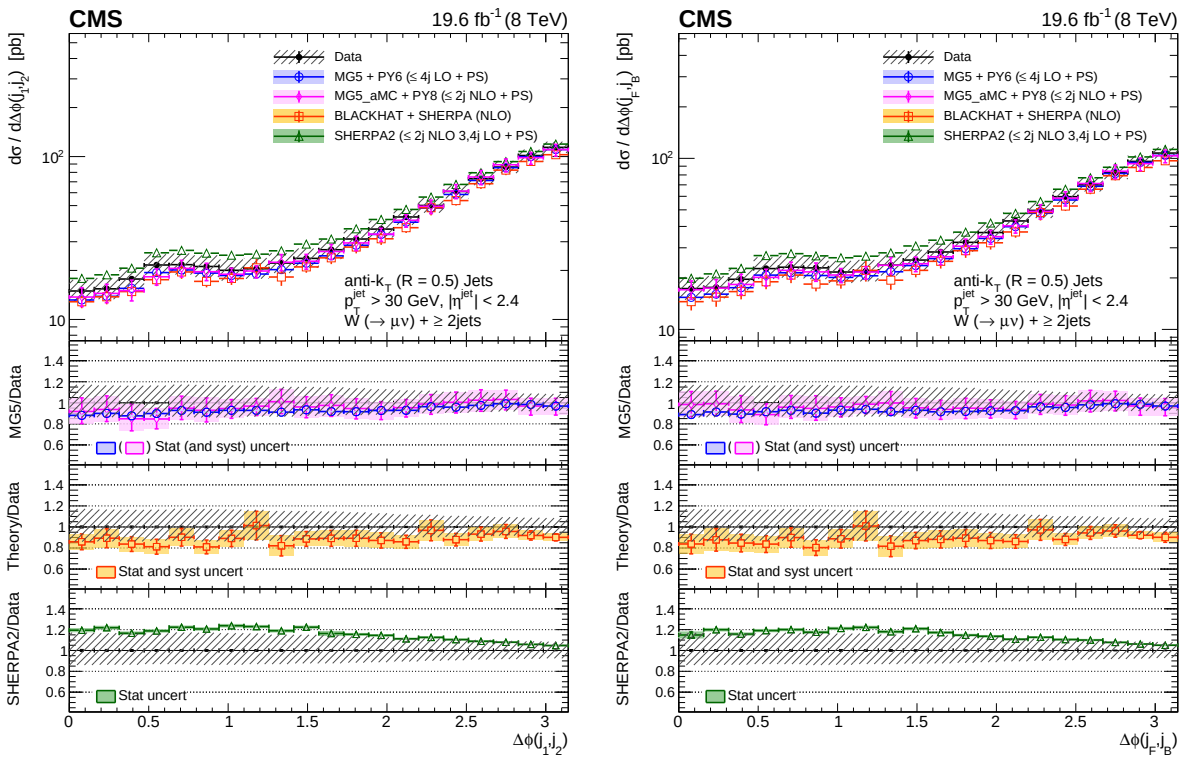


FIGURE 4.12: Cross sections differential in $\Delta\phi(j_1, j_2)$ (left) and $\Delta\phi(j_F, j_B)$ (right) for an inclusive jet multiplicity of 2, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

Finally, the average number of jets, $\langle N_{\text{jets}} \rangle$, is shown as a function of H_{T}, $\Delta y(j_1, j_2)$, and $\Delta y(j_{\text{F}}, j_{\text{B}})$ in the inclusive two-jet events in Fig. 4.15. In the high-H_{T} region, which is particularly sensitive to higher-order processes, the average number of jets plateaus around a value of 3.5. Although MADGRAPH5+PYTHIA6 tends to underestimate $\langle N_{\text{jets}} \rangle$ and SHERPA 2 tends to overestimate it, the deviations are not significant and both generators appear to adequately reproduce the data. Good agreement is observed between the data and all predictions for the dependence of $\langle N_{\text{jets}} \rangle$ on the p_{T}-ordered and rapidity-ordered rapidity differences. These measurements provide an important test of the implementation and modeling of wide-angle gluon emission in the MC generators and NLO calculations. Overall, the accuracy of the predictions for $\langle N_{\text{jets}} \rangle$ is much better than was found at the Tevatron [20].

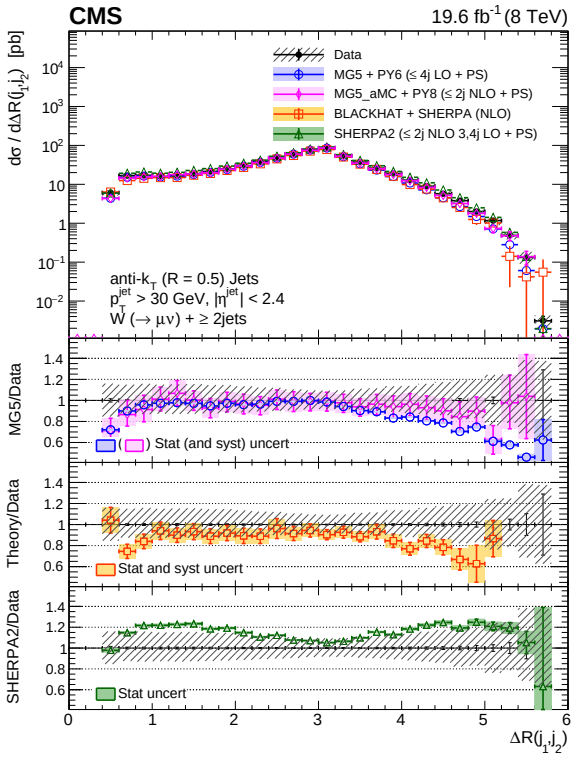


FIGURE 4.13: Cross section differential in $\Delta R(j_1, j_2)$ for an inclusive jet multiplicity of 2, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

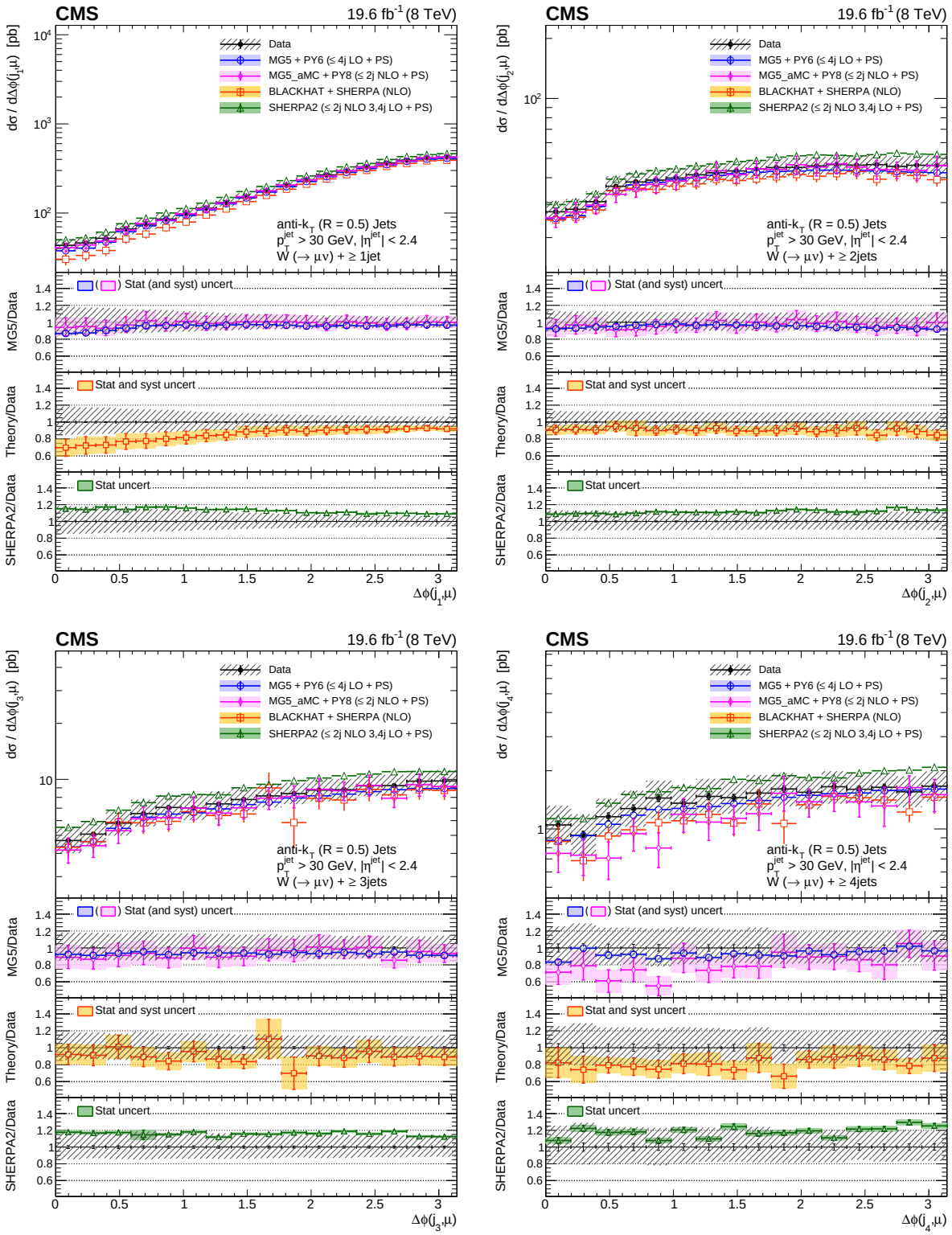


FIGURE 4.14: Cross sections differential in $\Delta\phi(j_n, \mu)$ for inclusive jet multiplicities $n = 1$–4, compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

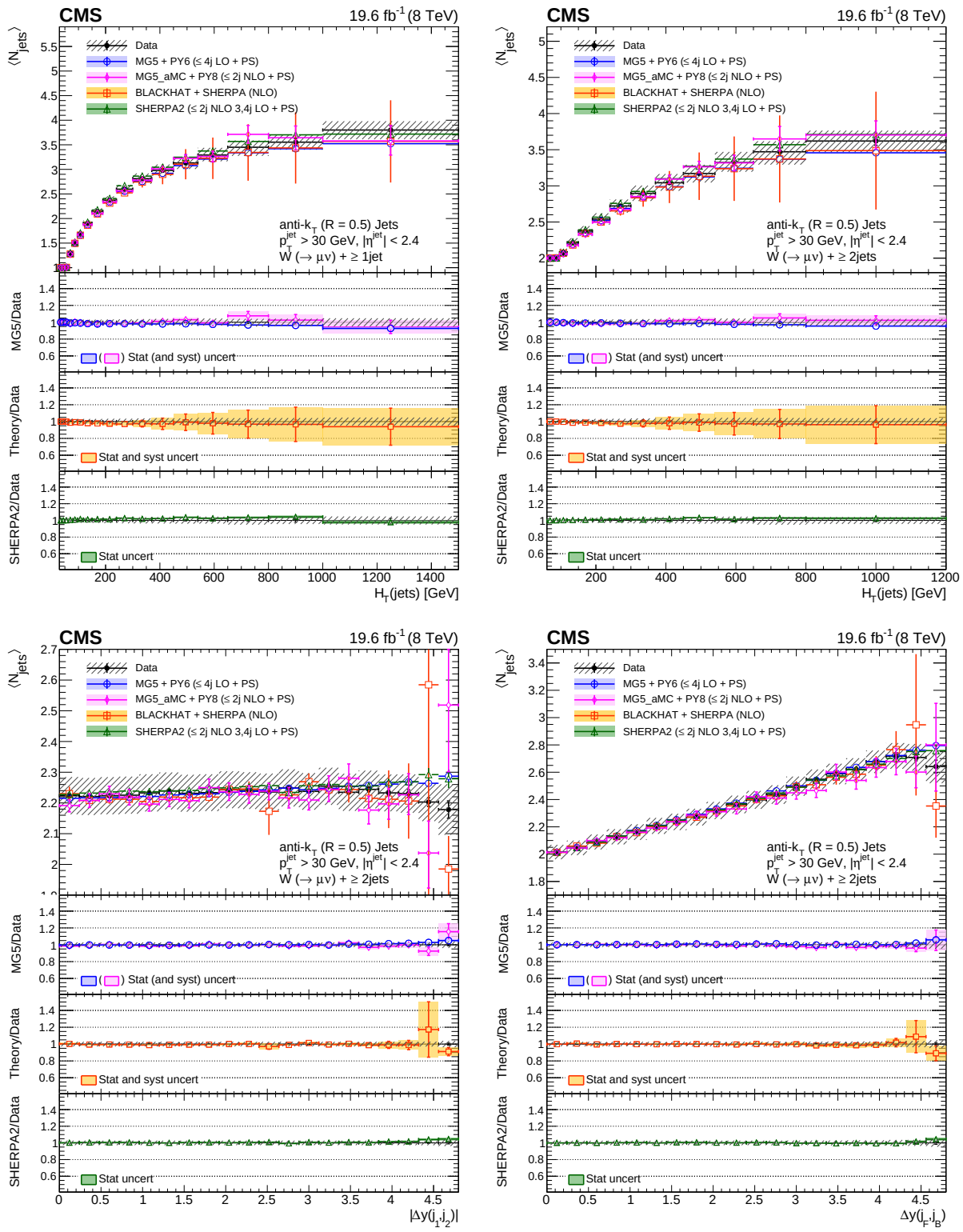


FIGURE 4.15: Average number of jets $\langle N_{\text{jets}} \rangle$ as a function of H_{T} for inclusive jet multiplicities 1–2 (top row) and as a function of $\Delta y(j_1, j_2)$ and $\Delta y(j_F, j_B)$ for an inclusive jet multiplicity of 2 (bottom row), compared to the predictions of MADGRAPH, MADGRAPH5_aMC@NLO, SHERPA 2, and BLACKHAT+SHERPA (corrected for hadronization and multiple-parton interactions). Black circular markers with the gray hatched band represent the unfolded data measurements and their total uncertainties. Overlaid are the predictions together with their uncertainties. The lower plots show the ratio of each prediction to the unfolded data.

4.7 W+Jets 13 TeV Analysis

4.7.1 Samples and Dataset

Similar to the W+jets at 8 TeV analysis, signal and background processes are produced and fully reconstructed using a simulation of the CMS detector based on GEANT4 (v9.4p03) [48], except for the QCD multijet background, which is estimated from control data samples. The processes of the W+jets signal and the Z/γ^*+jets background are generated by MADGRAPH5_aMC@NLO (v5.2.2.2) [72] with an NLO calculation. The FxFx jet merging scheme [51] is used in the MADGRAPH5_aMC@NLO generator. The $t\bar{t}$ background is generated at NLO with POWHEG (v2.0) [79, 80, 81]. The single top quark background processes are simulated either with POWHEG (v1.0) [107] or with MADGRAPH5_aMC@NLO depending on the particular channel. Among the diboson background processes, the WW production is generated with POWHEG (v2.0) [108], while the WZ and ZZ productions are generated using PYTHIA8 (v8.212) [71, 109]. The signal and background simulated samples, except for diboson production, are interfaced with PYTHIA8 for parton showering and hadronization. The CUETP8M1 tune [99] was used in PYTHIA8. The NNPDF 2.3 LO PDF [110, 111] and the NNPDF 3.0 NLO PDF [96] are used to generate background processes, where the former is used in PYTHIA8. The simulated processes include the effect of additional pp collisions in the same or adjacent bunch crossings (PU). The PU contribution is simulated as additional minimum bias events superimposed on the primary simulated events based on a distribution of the number of interactions per bunch crossing with an average of about 11 collisions, which is reweighted to match that observed in data.

4.7.2 Event Selection

Data events are retained if they pass an online trigger requirement (an OR of 2 single muon triggers: *HLT_IsoMu20* and *HLT_IsoTkMu20*) with a muon reconstructed in the online system with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, while the simulated events are required to pass an emulation of the trigger requirement.

Muon PF candidates are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ inside the acceptance of $|\eta| < 2.4$ and to satisfy the tight identification criteria [54]. In addition, a combined relative PF isolation requirement $I_{\mathrm{PF}}^{\mathrm{rel}} < 0.15$ is applied to suppress the contamination from muons contained in jets. The small differences in muon identification, isolation, and trigger efficiencies between data and simulated processes are compensated by applying corrections to the simulated events.

Hadronic jets are reconstructed from the PF candidates with a distance parameter of $R = 0.4$. Reconstructed jet energies are corrected by using p_{T}- and η-dependent corrections described in Sec. 3.5. The jets in simulated events are smeared by an η-dependent factor to account for the difference in energy resolution between data and simulation [60]. Jets are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ inside the acceptance of $|y| < 2.4$ and a spatial separation of $\Delta R > 0.4$ from muon candidates. Additional loose selection criteria are applied to each event to suppress nonphysical jets [112]. A number of vertexing-related and jet-shower-shape-related input variables are combined into a boosted decision tree yielding a single discriminator for the identification of pileup jets [112]. The contribution from pileup jets is reduced by applying a selection on the discriminator that has been optimized to minimize the dependency on the number of reconstructed vertices.

For the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$, the vector p_{T} sum of particles that were clustered as jets is replaced by the vector p_{T} sum of the jets including the jet energy corrections. Moreover, a

set of individual $E_{\mathrm{T}}^{\mathrm{miss}}$ filters are applied to veto events with anomalous $E_{\mathrm{T}}^{\mathrm{miss}}$ due to various subdetector malfunctions and algorithmic errors [113].

The W($\mu\nu$)+jets events are required to have exactly one muon and one or more jets. Events with additional muon PF candidates, which are not necessarily subject to the muon identification and isolation criteria, with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$ are removed. Events are further required to be in the transverse mass peak region for W bosons, defined by $M_{\mathrm{T}} > 50\,\mathrm{GeV}$. The M_{T} selection discriminates against non-W final states, such as QCD multijet background, that have a lepton candidate and nonzero $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, but a relatively low value of M_{T}. For the analysis of the $\Delta R(\mu$, closest jet) distribution, jets in the event are required to have $p_{\mathrm{T}} > 100\,\mathrm{GeV}$, with the leading jet $p_{\mathrm{T}} > 300\,\mathrm{GeV}$. This selection results in a boosted topology, where two jets recoil against each other and one of them can lose a significant amount of energy to the decay products of the emitted real W boson.

The contamination from $\mathrm{t\bar{t}}$ events is reduced by applying a b-quark tag veto to the events that contain one or more b-tagged jets. For this veto, the combined secondary vertex tagger (CSVv2) [65] is used as the b tagging algorithm with medium discrimination working point [64] corresponding to the misidentification probability of approximately 1% for light-flavor jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. After the implementation of the b tag veto, the expected contributions from the background processes and the observed data are given in Table 4.6 as a function of the jet multiplicity. For jet multiplicities of 1–6, the b tag veto rejects 71–88% of the predicted $\mathrm{t\bar{t}}$ background and 5–29% of the W+jets signal. Differences in the data and simulation b tagging efficiencies and mistagging rates are corrected by applying data-to-simulation scale factors [64].

TABLE 4.6: Numbers of events in simulation and data as a function of the exclusive jet multiplicity after the implementation of b tag veto. The processes included are: WW, WZ, and ZZ diboson (VV), QCD multijet, single top quark (Single t), Z/γ^*+jets Drell–Yan (DY+jets), $t\bar{t}$, and W($\mu\nu$)+jets signal processes. The QCD multijet background is estimated using control data samples. The $t\bar{t}$ background is scaled as discussed in Sec. 4.7.3.

N_{jets}	0	1	2	3	4	5	6
VV	4 302	1 986	774	205	45	10	2
QCD multijet	205 800	75 138	12 074	2 556	612	53	5
Single t	3 392	5 484	3 277	1 194	317	83	19
DY+jets	520 653	69 660	14 666	3 041	643	133	33
$t\bar{t}$	1 663	4 901	8 084	6 170	3 152	1 152	319
W($\mu\nu$)+jets	12 171 400	1 601 858	326 030	64 484	11 736	2 072	404
Total	12 907 210	1 759 027	364 905	77 650	16 505	3 503	782
Data	12 926 230	1 680 182	349 480	73 817	16 866	3 964	909

4.7.3 Background Estimation

Signal and background processes in these comparisons are simulated with the event generators described in Sec. 4.7.1. Similar to the analysis of W+jets at 8 TeV, the QCD multijet background is estimated using a data control region as described in 4.6.3.

At high jet multiplicities, where the W+jets signal is less dominant, the accuracy of the background modeling becomes more important, especially for the $t\bar{t}$ production process. We created a $t\bar{t}$-enriched control sample by requiring two or more b-tagged jets. The purity of this $t\bar{t}$ control sample increases towards higher jet multiplicities and ranges between 79–96% for jet multiplicities of 2–6. The differences between data and simulation observed for jet multiplicities of 2–6 in the $t\bar{t}$ control region are expressed in terms of $t\bar{t}$ data-to-simulation scaling factors that range between 0.75 and 1.15. The $t\bar{t}$ background events are scaled by these factors in all the reconstructed-level and unfolded distributions presented in this paper for events with jet multiplicities of 2–6.

The comparison of reconstructed distributions for data and simulated processes is shown in Fig. 4.16 for the jet multiplicity. The p_{T} distribution of the leading jet and the azimuthal correlation between the muon and the leading jet in events with at least one jet are shown in

Fig. 4.17. For each reconstructed distribution, the ratio of the sum of the simulated processes from signal and backgrounds to the data is presented to quantify possible disagreements (the corresponding error bars represent the statistical uncertainties stemming from both data and simulation). The data-to-simulation agreement is on the 5% level in almost all regions.

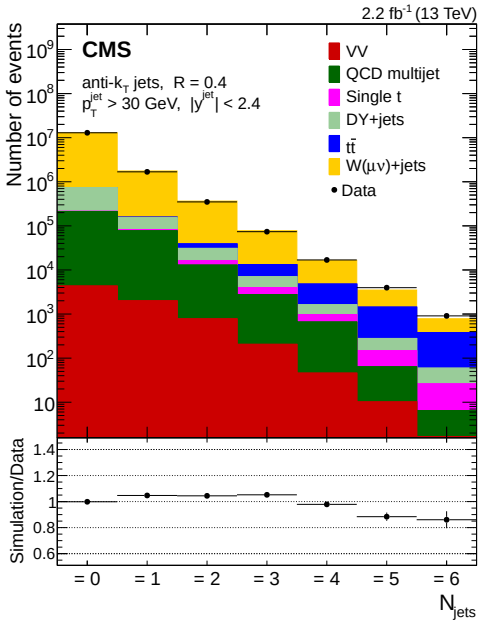


FIGURE 4.16: Data-to-simulation comparison as a function of the jet multiplicity. The processes included are listed in Table 4.6. The QCD multijet background is estimated using control samples in data. The $t\bar{t}$ background is scaled as discussed in Sec. 4.7.3. The error bars in the ratio panel represent the combined statistical uncertainty of the data and simulation.

4.7.4 $\Delta R(\mu$, closest jet)

In addition to the differential cross section measurement as functions of jet p_{T}, the jet $|y|$, H_{T}, and $\Delta\phi(\mu, j_i)$ up to multiplicities of four inclusive jets, the differential cross section as a function of the angular distance between the muon and the closest jet $\Delta R(\mu$, closest jet) for events with one or more jets is measured. In this case, the leading jet is required to have $p_{\mathrm{T}} > 300\,\mathrm{GeV}$, and all jets must have $p_{\mathrm{T}} > 100\,\mathrm{GeV}$.

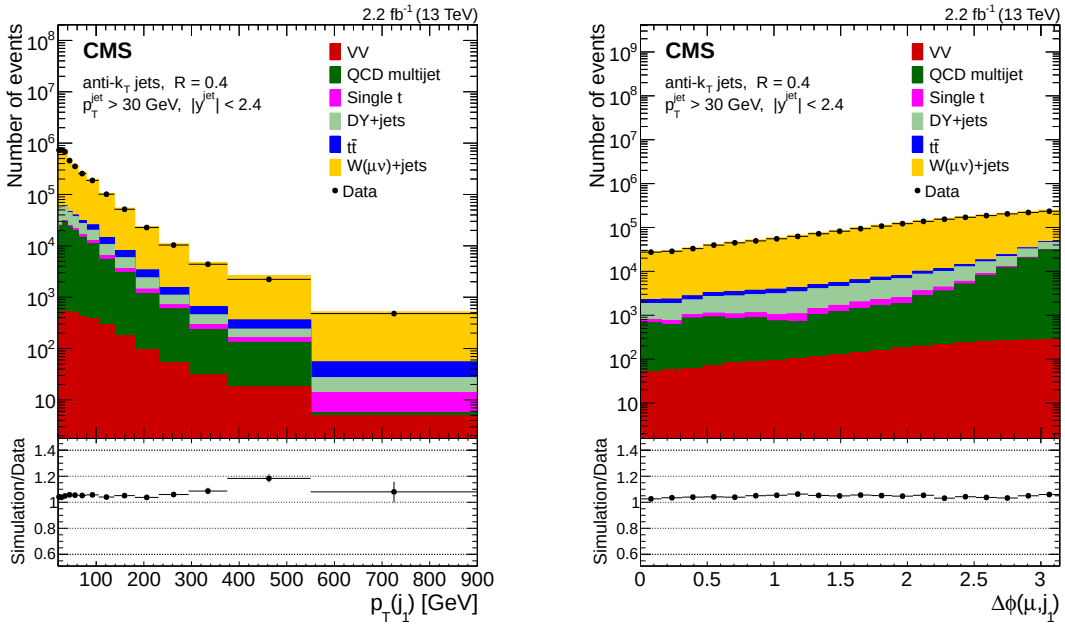


FIGURE 4.17: Data-to-simulation comparison as functions of the leading jet p_{T} (left) and $\Delta\phi(\mu, j_1)$ between the muon and the leading jet (right) for one jet inclusive production. The processes included are listed in Table 4.6. The QCD multijet background is estimated using control samples in data. The $\mathrm{t\bar{t}}$ background is scaled as discussed in Sec. 4.7.3. The error bars in the ratio panel represent the combined statistical uncertainty of the data and simulation.

The $\Delta R(\mu$, closest jet) variable separates the process of the electroweak emission of real W bosons from an initial- or final-state quark, which was recently studied by the ATLAS collaboration with 8 TeV data [23]. The contribution of electroweak radiative processes to the measurement of W+jets becomes significant with the increasing center-of-mass energy of collisions, leading to an enhancement in the collinear region of the distribution of the angular distance between the W boson and the closest jet [78, 114, 115, 116, 117, 118].

4.7.5 Unfolding

This unfolding procedure is applied separately to each measured differential cross section and includes corrections for the trigger and the muon selection efficiencies. This procedure is performed using the method of D'Agostini iteration with early stopping [68, 119, 120] that

is implemented in the statistical analysis toolkit RooUnfold [69]. A response matrix is constructed using generator and reconstruction levels of the NLO MG_aMC FxFx W+jets simulated sample. The details of the method and how the response matrix is constructed are discussed in Sec. 4.4 and Sec.4.6.5.

4.7.6 Systematic Uncertainties

The sources of systematic uncertainties considered in the analysis are:

- The dominant source of systematic uncertainty is the jet energy scale uncertainty. This uncertainty amount to 1.3 % for a jet multiplicity of 1 and increase with the number of reconstructed jets. Uncertainties in the jet energy scale are propagated to the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$

- The background cross sections are varied within their uncertainties. The cross section of the largest background contribution, coming from $\mathrm{t\bar{t}}$, is varied by 10%. The other backgrounds are also simultaneously varied up (down) by 7% for ZZ and WZ, 6% for WW, and 4% for Z+jets. For single-top processes, they are varied by 4% for s- and t-channel and 6% for the tW-channel. The QCD multijet background is estimated using data control regions and has an uncertainty based on the number of events in an inverted isolation sample where the shape is extracted, and in low M_{T} control regions, with nominal and inverted isolation requirements, which are used to calculate the normalization. This method is described in Ref. [22]. The uncertainty associated with the estimated multijet background is 0.6–24% for jet multiplicities 1–5.

- A systematic uncertainty associated with the generator used to build the unfolding response matrix is assigned by weighting the simulation to agree with the data in each distribution and building a reweighted response to unfold the data. The reweighting is done using a finer binning. The difference between the nominal results, and the results

unfolded using the reweighted response matrix is taken as the systematic uncertainty associated with the unfolding response matrix.

- The scale factors correcting for data-MC difference in jet energy resolution (JER) are applied. The uncertainty on these factors need to be considered. The effect of this is assessed by scaling the jets in W+jets sample with two additional sets of scale factors that correspond to varying the factors up and down by one sigma and evaluating the impact of these new sets. The resulting uncertainty is of the oder of 1%.

- The uncertainty of the pileup model (PU) is assessed by varying the minimum bias cross section by ±5%. The resulting uncertainty is of the order of 1%.

- The uncertainties on the data-to-simulation correction factors of the b-tagging efficiencies are considered (BtaggSF). This systematic uncertainty is assessed by adjusting the scale factors up and down according to their uncertainties found in the official payload. The entire analysis is performed with these variations and the final unfolded results are compared to the results of the standard analysis. The difference is taken to be the systematic uncertainty.

- The uncertainty on the data-to-simulation scale factors on the efficiency of muon selection is found by summing systematic uncertainties of trigger, identification, and isolation efficiencies in quadrature and is set as a global factor of 1.23%.

- The uncertainty on the integrated luminosity is taken to be 4.6%.

The systematic uncertainties assigned for each source of the systematics discussed above and the total systematic uncertainties are documented in this note. The systematic uncertainty values for the exclusive jet multiplicity is given in Table 4.7

TABLE 4.7: Differential cross section in exclusive jet multiplicity and break down of the systematic uncertainties for the muon decay channel.

Exclusive jet multiplicity

N_{jets}	$\frac{d\sigma}{dN_{\text{jets}}}$ [pb]	Tot. Unc [%]	stat [%]	MC stat [%]	JES [%]	JER [%]	PU [%]	XSEC [%]	Lumi [%]	BtagSF [%]	LER [%]	Unf [%]	Eff [%]
= 0	7.29e+03	5.3	0.038	0.055	0.040	0.088	1.9	0.16	4.9	0.40	0.0	0.0	0.53
= 1	904.	5.5	0.16	0.25	0.91	0.40	1.2	0.23	5.2	0.90	0.0	0.0	0.56
= 2	214.	8.5	0.36	0.63	6.3	0.37	1.4	0.53	5.2	1.5	0.0	0.0	0.57
= 3	45.4	12.	0.92	1.5	9.0	0.35	1.5	1.9	5.9	3.0	0.0	0.0	0.64
= 4	10.8	17.	2.1	3.3	12.	0.85	1.7	4.5	7.3	6.5	0.0	0.0	0.79
= 5	2.42	26.	5.2	8.0	18.	0.43	2.1	8.1	9.0	11.	0.0	0.0	0.98

4.7.7 Results

The theoretical predictions considered are listed below.

- **LO MG_aMC** : the LO prediction generated with MADGRAPH5_aMC@NLO interfaced with PYTHIA8 for parton showering and hadronization. The ME calculation includes the five processes pp $\rightarrow$ W+N-jets, $N = 0 \ldots 4$ and it is matched to the parton showering using the k_{T}-MLM [50, 121] scheme with the merging scale set at 19 GeV. The NNPDF 3.0 LO is used for the ME calculation while the NNPDF 2.3 LO PDF is used in the parton showering and hadronization with PYTHIA8 using the CUETP8M1 tune.

- **NLO MG_aMC FxFx** : the NLO prediction generated with MADGRAPH5_aMC@NLO interfaced with PYTHIA8 for parton showering and hadronization. The FXFX merging scheme is used with a merging scale parameter set to 30 GeV. This prediction has an NLO accuracy for pp $\rightarrow$ W+N-jets, $N = 0$, 1, 2, and LO accuracy for $N = 3$, 4. The NNPDF 3.0 NLO is used for the ME calculation , while the NNPDF 2.3 LO PDF is used in the parton showering and hadronization with PYTHIA8 using the CUETP8M1 tune.

- N_{jetti} **NNLO** : the fixed-order calculation based on the N-jettiness subtraction scheme (N_{jetti}) at NNLO for W+1-jet production [77, 78]. The comparison is made for the measured distributions of the leading jet p_T and $|y|$, H_T, $\Delta\phi(\mu, j_1)$, and $\Delta R(\mu$, closest jet) for events with one or more jets. The NNPDF 3.0 NNLO PDF is used in this calculation. To account for nonperturbative effects in the NNLO, the predictions with and without multiple parton interactions and hadronization are computed with LO MG_aMC interfaced with PYTHIA8. The value of this multiplicative correction applied to the NNLO calculation is mostly within the range of 0.93–1.10. The effect of final-state radiation (FSR) from the muon on the NNLO prediction is estimated to be less than 1%.

In the figure legends, LO and NLO predictions by MADGRAPH5_aMC@NLO are referred to as MG_aMC + PY8 ($\leq$ 4j LO + PS) and MG_aMC FxFx + PY8 ($\leq$ 2j NLO + PS) respectively.

The size of the data sample used in this analysis allows the measurements of cross sections of the W($\mu\nu$)+jets process for jet multiplicities up to six and fiducial cross sections as functions of several kinematic observables for up to four inclusive jets.

The measured W+jets differential cross section distributions are shown here in comparison with the predictions of the multileg NLO MG_aMC FxFx and multileg LO MG_aMC tree level k_T-MLM event generators, as described in above. Furthermore, the measured cross sections are compared to the fixed-order N_{jetti} NNLO calculation for W+1-jet production on the leading jet p_T and $|y|$, H_T, $\Delta\phi(\mu, j_1)$, and $\Delta R(\mu$, closest jet) distributions. The ratios of the predictions to the measurements are provided to make easier comparisons.

Total experimental uncertainties are quoted for the data in the differential cross section distributions. The multileg LO MG_aMC prediction is given only with its statistical uncertainty. The NLO MG_aMC FxFx prediction is given with both the statistical and systematic

uncertainties. Systematic uncertainties in the NLO MG_aMC FxFx prediction are obtained by varying the NNPDF 3.0 NLO PDFs and the value of α_s, and by varying independently the renormalization and factorization scales by a factor of 0.5 and 2. All possible combinations are used in variations of scales excluding only the cases where one scale is varied by a factor of 0.5 and the other one by a factor of 2. The total systematic uncertainty is the squared sum of these uncertainties. The systematic uncertainty due to variation of scale factors for the exclusive jet multiplicity distribution is computed using the method described in Refs. [122, 123]. For the NNLO prediction, the theoretical uncertainty includes both statistical and systematic components, where the systematic uncertainty is calculated by varying independently the central renormalization and factorization scales by a factor of 2 up and down, disallowing the combinations where one scale is varied by a factor of 0.5 and the other one by a factor of 2.0.

The measured differential cross sections as functions of the exclusive and inclusive jet multiplicities up to 6 jets are compared with the predictions of LO MG_aMC and NLO MG_aMC FxFx in Fig. 4.18. The measured cross sections and the predictions are in good agreement within uncertainties.

The measured cross sections for inclusive jet multiplicities of 1–4 are compared with the predictions as a function of the jet p_{T} ($|y|$) in Fig. 4.19 (Fig.4.20). The measured cross sections as functions of the jet p_{T} and $|y|$ are better described by the NLO MG_aMC FxFx prediction for all inclusive jet multiplicities and by the NNLO calculation for at least one jet. The LO MG_aMC prediction exhibits a slightly lower trend in estimating data in contrast to NLO MG_aMC FxFx and NNLO on jet p_{T} and $|y|$ distributions, particularly at low p_{T} and for inclusive jet multiplicities of 1–3.

The measured cross sections as functions of the H_{T} variable of the jets, which is sensitive to the effects of higher order corrections, are compared with the predictions. The H_{T} distributions for inclusive jet multiplicities of 1–4 are shown in Fig. 4.21. The predictions are in

good agreement with data for the H_{T} spectra of the jets for all inclusive jet multiplicities, with the exception of LO MG_aMC, which slightly underestimates the data at low H_{T}.

The differential cross sections are also measured as functions of angular variables: the azimuthal separation $\Delta\phi(\mu, j_i)$ between the muon and the jet for inclusive jet multiplicities of 1–4, and the angular distance between the muon and the closest jet $\Delta R(\mu, \text{closest jet})$ in events with one or more jets. The measured $\Delta\phi(\mu, j_i)$ distributions are compared with the predictions in Fig. 4.22 and they are well described within uncertainties. This observable is sensitive to the implementation of particle emissions and other nonperturbative effects modeled by parton showering algorithms in MC generators.

The comparison of the measured $\Delta R(\mu, \text{closest jet})$ with the predictions is shown in Fig. 4.23. This observable probes the angular correlation between the muon emitted in the W boson decay and the direction of the closest jet. In the collinear region (small ΔR values), it is sensitive to the modeling of W boson radiative emission from initial- or final-state quarks. The predictions are observed to be in fairly good agreement with data within the uncertainties, but there are some differences. Around $\Delta R = 2.0$–2.5, in the transition between the region dominated by back-to-back W+$N\geq$1-jet processes (high ΔR) and the region where the radiative W boson emission should be enhanced (low ΔR), the NLO MG_aMC FxFx prediction overestimates the measured cross section. In the high-ΔR region, the LO MG_aMC prediction underestimates the data, which is consistent with the other observables.

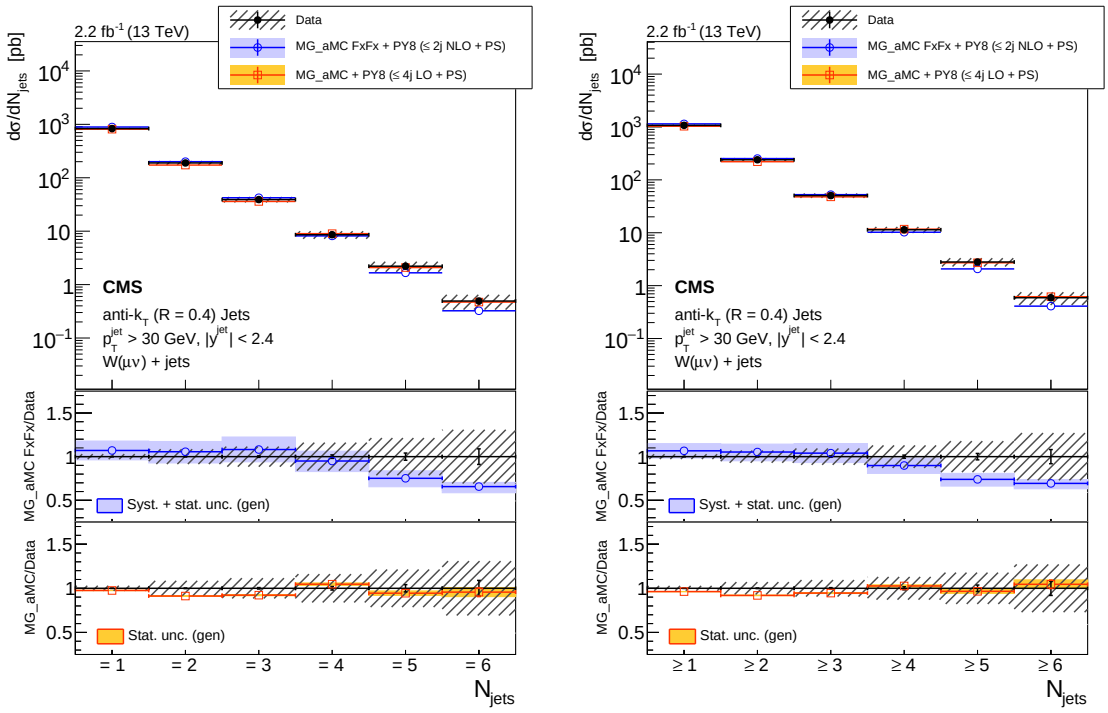


FIGURE 4.18: Differential cross section measurement for the exclusive (left) and inclusive jet multiplicities (right), compared to the predictions of NLO MG_aMC FxFx and LO MG_aMC. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The band around the NLO MG_aMC FxFx prediction represents its theoretical uncertainty including both statistical and systematic components. The lower panels show the ratios of the prediction to the unfolded data.

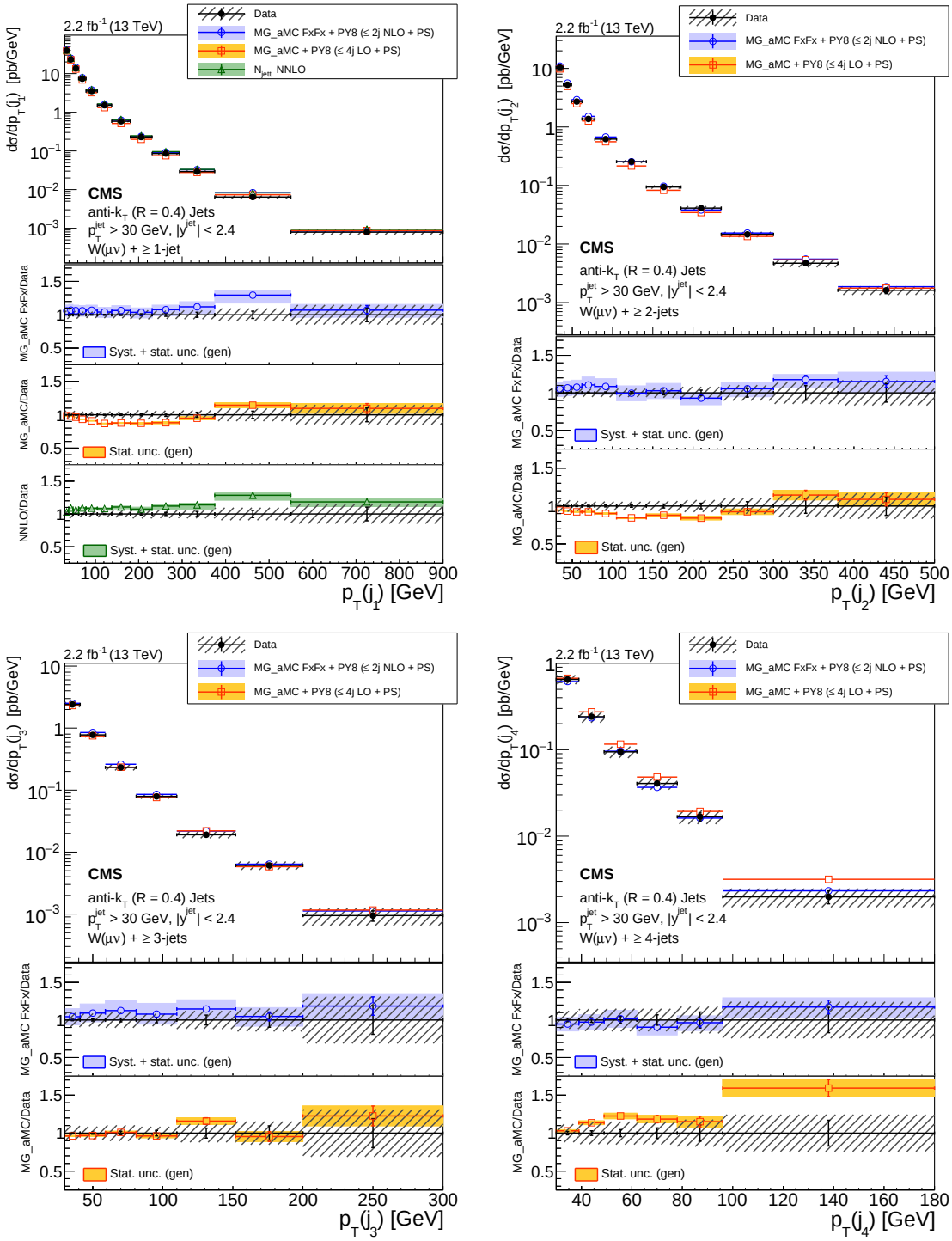


FIGURE 4.19: Differential cross section measurement for the transverse momenta of the four leading jets, shown from left to right for at least 1 and 2 jets (upper) and for at least 3 and 4 jets (lower) on the figures, compared to the predictions of NLO MG_aMC FxFx and LO MG_aMC. The NNLO prediction for W+1-jet is included in the first leading jet p_{T}. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The bands around the NLO MG_aMC FxFx and NNLO predictions represent their theoretical uncertainties including both statistical and systematic components. The lower panels show the ratios of the prediction to the unfolded data.

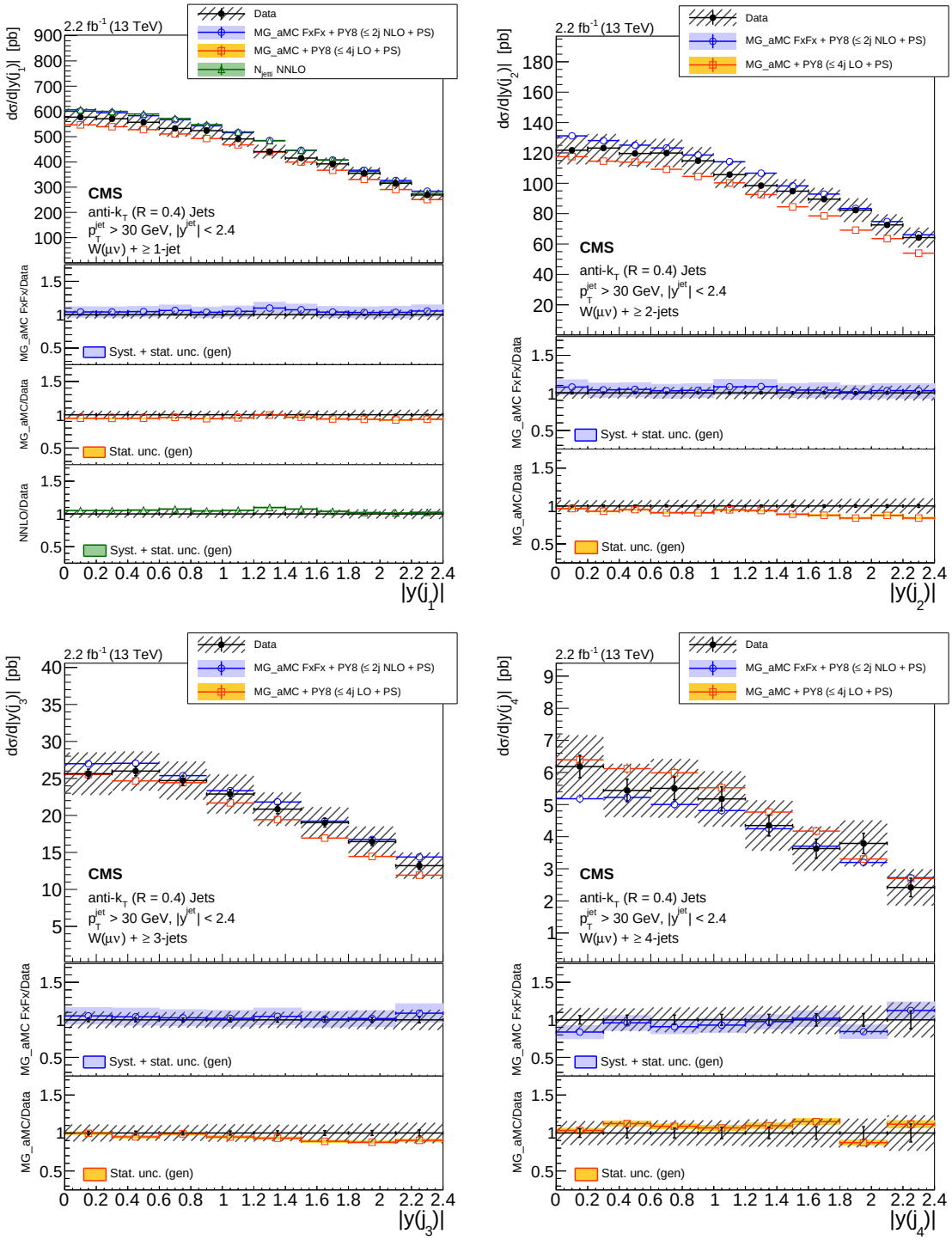


FIGURE 4.20: Differential cross section measurement for the absolute rapidities of the four leading jets, shown from left to right for at least 1 and 2 jets (upper) and for at least 3 and 4 jets (lower) on the figures, compared to the predictions of NLO MG_aMC FxFx and LO MG_aMC. The NNLO prediction for W+1-jet is included in the first leading jet $|y|$. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The bands around the NLO MG_aMC FxFx and NNLO predictions represent their theoretical uncertainties including both statistical and systematic components. The lower panels show the ratios of the prediction to the unfolded data.

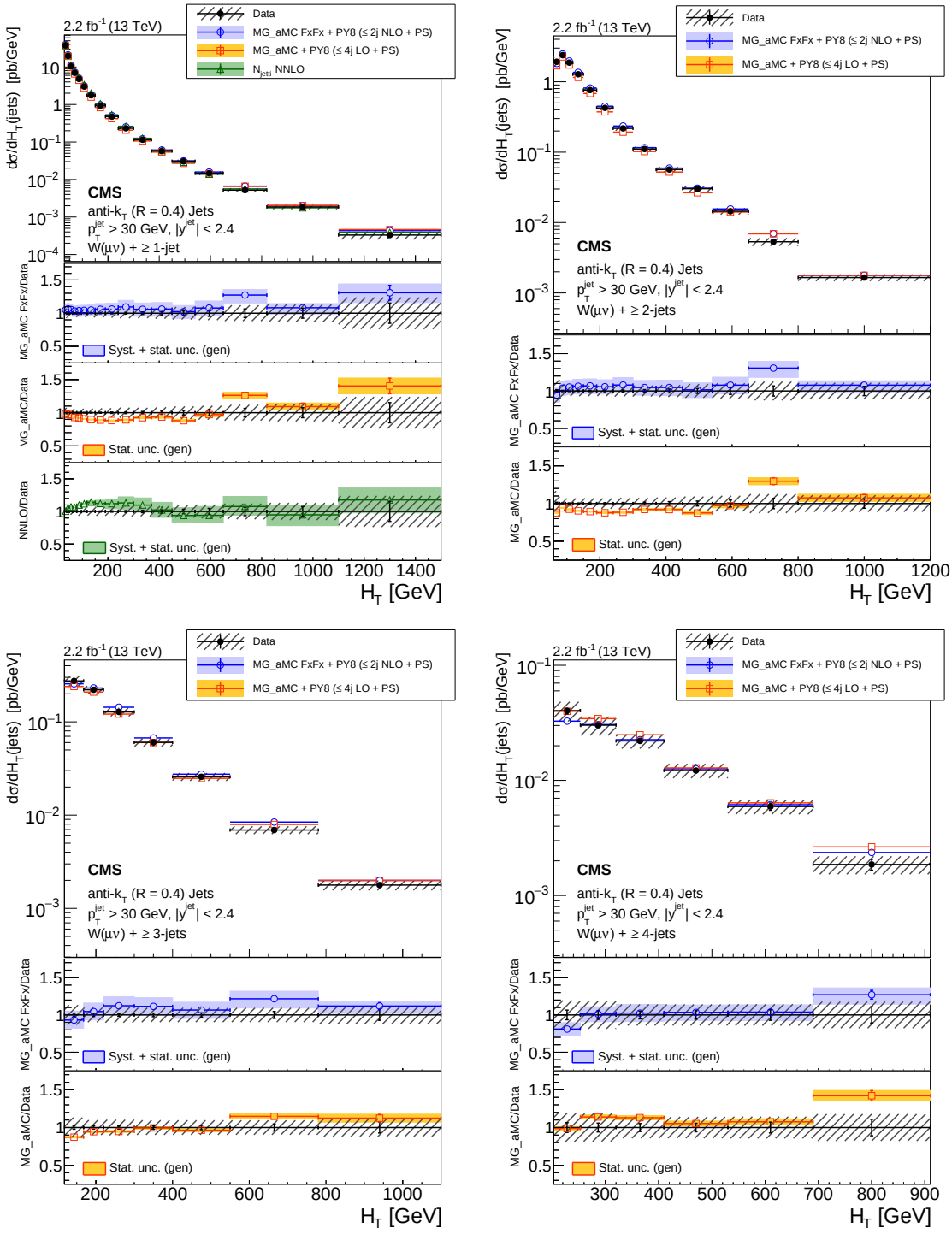


FIGURE 4.21: Differential cross section measurement for the jets H_{T}, shown from left to right for at least 1 and 2 jets (upper) and for at least 3 and 4 jets (lower) on the figures, compared to the predictions of NLO MG_aMC FxFx and LO MG_aMC. The NNLO prediction for W+1-jet is included in the jets H_{T} for one jet inclusive production. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The bands around the NLO MG_aMC FxFx and NNLO predictions represent their theoretical uncertainties including both statistical and systematic components. The lower panels show the ratio of the prediction to the unfolded data.

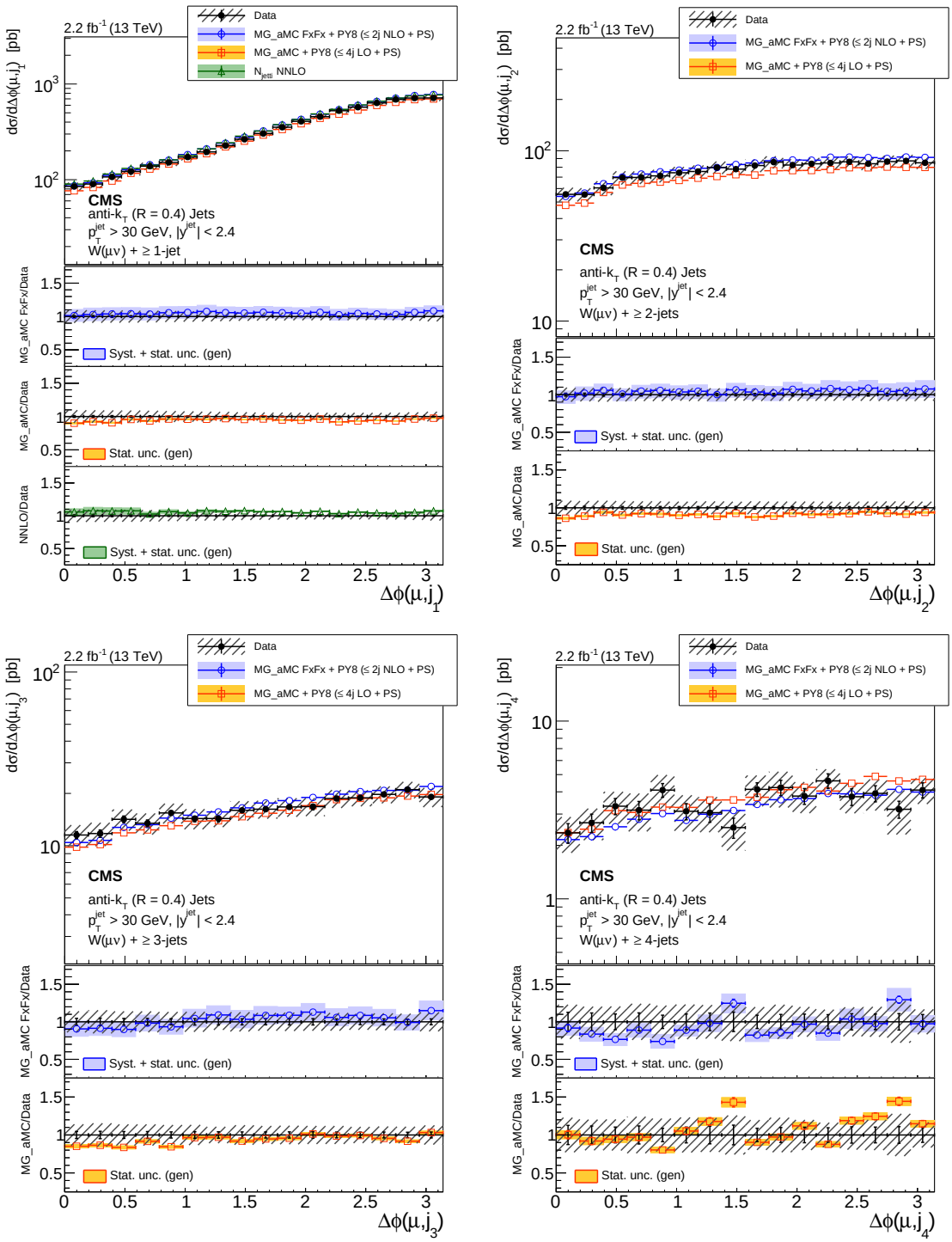


FIGURE 4.22: Differential cross section measurement for $\Delta\phi(\mu, j_i)$, shown from left to right for at least 1 and 2 jets (upper) and for at least 3 and 4 jets (lower) on the figures, compared to the predictions of NLO MG_aMC FxFx and LO MG_aMC. The NNLO prediction for W+1-jet is included in $\Delta\phi(\mu, j_1)$ for one jet inclusive production. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The bands around the NLO MG_aMC FxFx and NNLO predictions represent their theoretical uncertainties including both statistical and systematic components. The lower panels show the ratio of the prediction to the unfolded data.

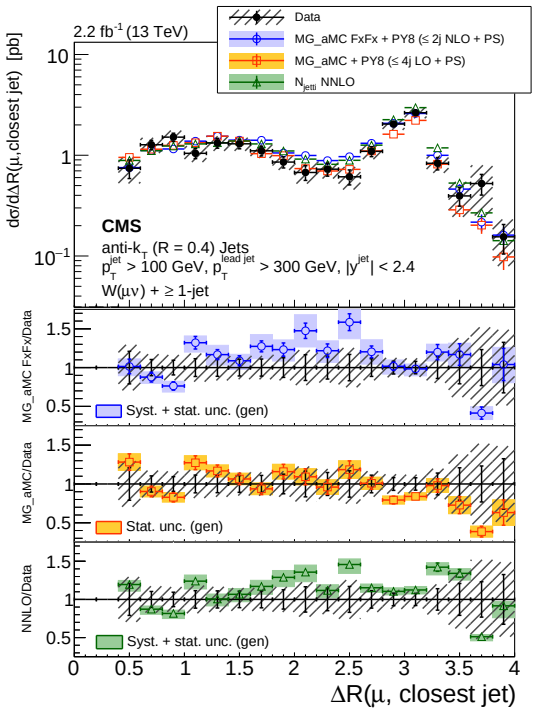


FIGURE 4.23: Differential cross section measurement for $\Delta R(\mu$, closest jet) for one jet inclusive production, compared to the predictions of NLO MG_aMC FxFx, LO MG_aMC, and the NNLO calculation. All jets in the events are required to have $p_{\mathrm{T}} > 100\,\mathrm{GeV}$, with the leading jet $p_{\mathrm{T}} > 300\,\mathrm{GeV}$. The black circular markers with the gray hatched band represent the unfolded data measurement and the total experimental uncertainty. The LO MG_aMC prediction is given only with its statistical uncertainty. The bands around the NLO MG_aMC FxFx and NNLO predictions represent their theoretical uncertainties including both statistical and systematic components. The lower panels show the ratio of the prediction to the unfolded data.

Chapter 5

Analysis II: HH Searches

5.1 Overview

In this analysis, a search for resonant X→HH production in the $b\bar{b}Z(\rightarrow \ell\ell)Z(\rightarrow jj)$ final state (the $b\bar{b}\ell\ell jj$ channel) is conducted using 35.9 fb^{-1} of data generated from pp collisions at $\sqrt{s}$ = 13 TeV and recorded in 2016 by the CMS detector at the LHC. The final state signature consists of two b jets from a Higgs boson decay, two charged leptons from a Z boson decay and two jets of any flavor from a second Z boson decay. The charged leptons are either electrons or muons. A Feynman diagram showing such final state decay is depicted in Fig. 5.1. The challenge of the search in the $b\bar{b}\ell\ell jj$ channel is to discriminate the signal of two b jets and two jets from multi-jet background events. In addition, one of the Z bosons from Higgs decay will be off-shell, disallowing a shape peak in the invariant mass distribution. These challenges are addressed by making use of a boosted decision tree (BDT) classification for optimal signal and background discrimination. In particular, the BDT distributions of data and estimated backgrounds are used to set the upper limits on the HH resonance production cross section using the asymptotic CL_S modified frequentist approach [124, 125].

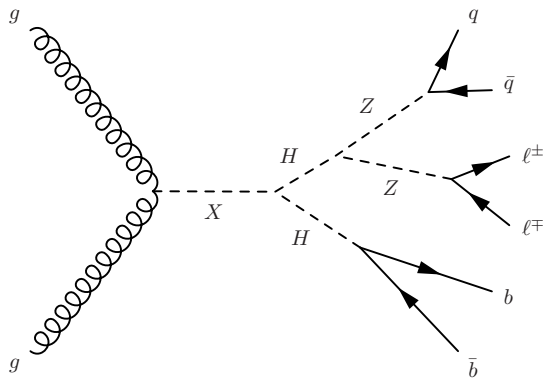


FIGURE 5.1: Feynman diagram showing a resonance produced in gluon fusion mode and decays into a pair of Higgs bosons which further decay in the $b\bar{b}\ell\ell jj$ channel. The figure was produced by G. Madigan.

The narrow width resonance models considered in this search arise from the Randall-Sundrum (RS) model [31] in the Warped Extra Dimension formulation (WED) [126, 127, 128, 129]. This scenario introduces one small spacial extra dimension with a non-factorizable geometry, where the SM particles are not allowed to propagate along the extra dimension, and is referred to, in this search, as RS1. The resonant particle produced can be a radion (spin-0) or the first Kaluza-Klein (KK) excitation of a graviton (spin-2). The production cross section of the radion is proportional to $1/\lambda_R^2$ where λ_R is the interaction scale parameter of the theory. In this analysis, we consider the cases where $\lambda_R = 1\,\text{TeV}$ with $kL = 35$, where k is the constant in the warp factor (e^{-kL}) appearing in the space-time metric of the theory and L is the size of the extra dimension. The free parameter of the model for the graviton case is $\tilde{k} = k/\overline{M_{Pl}}$ and we consider $\tilde{k} = 0.1$ in this analysis [130]. The production cross section at hadron colliders is expected to be mainly via gluon-gluon fusion, and we assume that the radion or graviton is produced exclusively via this process. The search covers a resonance mass range of 260–1000 GeV.

At the end, the results are combined with the search for resonant production of HH in the $b\bar{b}Z(\to \ell\ell)Z(\to \nu\nu)$ channel ($b\bar{b}\ell\ell\nu\nu$ channel) [131].

5.2 Samples and Dataset

5.2.1 Data Samples

The data used in this analysis were collected during the 2016 proton-proton run. Quality requirements were applied to exclude data taken during periods with known detector problems and only the validated data in the JSON file (Golden JSON) reported in Table 5.1 are utilized. The data sample corresponds to 35.9 fb^{-1} of integrated luminosity. Table 5.1 lists the datasets and corresponding luminosities. This analysis has been done using the `CMSSW_8_0_X` software release.

Events in the electron and muon channels come from the `DoubleEG` and `DoubleMu` primary dataset, respectively. Events are selected for the electron channel if they pass the double electron trigger listed in Table 5.2, and for the muon channel if they pass an OR of a suite of double muon triggers also listed in Table 5.2.

TABLE 5.1: Datasets with the corresponding range of run number and integrated luminosity, and JSON file used in the analysis.

Data sample	Run range
Run2016B-03Feb2017_ver2-v2	273150 - 275376
Run2016C-03Feb2017-v1	275656 - 276283
Run2016D-03Feb2017-v1	276315 - 276811
Run2016E-03Feb2017-v1	276831 - 277420
Run2016F-03Feb2017-v1	277932 - 278808
Run2016G-03Feb2017-v1	278820 - 280385
Run2016H-03Feb2017_ver2-v1	282035 - 281588
Run2016H-03Feb2017_ver3-v1	284036 - 284068
Total (Certified) [with brilcalc]	$\mathcal{L} = 35.9$ fb^{-1}

JSON
Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt

TABLE 5.2: Double Electron and Double Muon triggers used in the analysis.

Double Electron Triggers
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*

Muon Triggers
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_v *
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_v*
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_v*
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_v*

5.2.2 Simulated Samples

The main background processes to the production of two Higgs bosons in the $b\bar{b}\ell\ell jj$ channel are Z/γ^*+jets and $t\bar{t}$. Smaller contributions arise from single-top, W+jets, diboson + jets, standard model Higgs production, and QCD multijet production. Signal and background processes are modeled with Monte Carlo simulations, with the exception of the QCD multijet background which is estimated entirely using data in control regions.

Z/γ^*+jets and the W+jets processes are generated with MADGRAPH5_aMC@NLO [72], with a next-to-leading order (NLO) calculation. In this case, the MADGRAPH5_aMC@NLO generator uses the FXFX jet merging scheme [51].

The $t\bar{t}$ process is generated with POWHEG [79, 80, 81]. Single top processes and standard model Higgs productions are simulated either with POWHEG or MADGRAPH5-_aMC@NLO, depending on the particular channel. Diboson background processes are simulated with MADGRAPH5_aMC@NLO. The cross sections of these samples are normalized to their best known highest order QCD predictions, either evaluated at next-to-leading order (NLO) with MCFM [132] or at next-to-next-to-leading order (NNLO) with FEWZ [133], with the exception of $t\bar{t}$ and Z/γ^*+jets, which are normalized using data.

The signal processes of a radion spin-0 and a graviton spin-2 narrow-width resonance decaying to two Higgs bosons ($X \rightarrow HH$) are generated at leading order (LO) using MadGraph5_aMC@NLO. The mass of the Higgs boson is set to 125 GeV. The signal samples are produced considering resonance masses of 260, 270, 300, 350, 400, 450, 500, 550, 600, 650, 750, 800, 900, and 1000 GeV.

The simulated samples are interfaced with Pythia 8 for parton showering and hadronization. The Pythia 8 generator uses the CUETP8M1 tune scenario [99]. The NNPDF 3.0 NLO and LO PDF sets [96] are used for the different processes.

The simulated samples used in this analysis are tabulated in Table 5.3. All simulated samples were produced during the `RunIISummer2016` official `MiniAODv2` production campaign with a simulation of CMS detector response based on GEANT4 [48] applied. Digitization and reconstruction is performed using `CMSSW_8_0_X` software release. The presence of additional interactions in the same bunch crossing (PU), both in-time and out-of-time with respect to the primary interaction, is simulated with a multiplicity corresponding to the distribution measured in data.

Similar to the requirement on data, background and signal samples are required to pass the conditions imposed by the set of triggers described in Sec. 5.2.1. The difference between the trigger efficiency in data and in MC is corrected by applying scale factors which are a function of lepton p_T and η. For the simulation in the muon channel, a trigger scale factor of 1.0 is used for the combination of the muon triggers. This is consistent with what was measured by various analyses, for example in [134]. For the simulation in the electron channel, scale factors are taken from [135, 136] where the trigger scale factors have been measured for the electron identification used in this analysis.

TABLE 5.3: MC samples used in the analysis with their corresponding cross section. Abbreviations are used when listing the samples with **Sm16** and **Sp16** stands for *RunIISummer16MiniAODv2-PUMoriond17* and *RunIISpring16MiniAODv2-PUSpring16* respectively, **80X16** stands for *80X_mcRun2_asymptotic_2016*, **TIV** stands for *TrancheIV*, and **MIN** stands for *MINIAODSIM*.

DataSet	CrossSection (pb)
/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6_*/MIN	18610
/DYToLL_0J_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6*/MIN	4755.47
/DYToLL_1J_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6*/MIN	886.47
/DYToLL_2J_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6*/MIN	339.63
/WToLNu_0J_13TeV-amcatnloFXFX-pythia8/Sm16*_80X16_TIV_v6*/MIN	50279.20
/WToLNu_1J_13TeV-amcatnloFXFX-pythia8/Sm16*_80X16_TIV_v6*/MIN	8434.26
/WToLNu_2J_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6*/MIN	2813.24
/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/Sm16_*80X16_TIV_v6-v1/MIN	831.76
/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	49.997
/WWTo4Q_4f_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	51.723
/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v3/MIN	10.71
/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	3.033
/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	5.595
/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6-v1/MIN	4.42965
/WZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	6.324
/ZZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	4.04
/ZZTo4L_13TeV-amcatnloFXFX-pythia8/Sm16_80X16_TIV_v6_ext1-v1/MIN	1.212
/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	3.22
/ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	6.90
/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6*-v1/MIN	11.95
/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1/Sp16_80X16_miniAODv2_v2-v1/MIN	3.36
/ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/Sp16_80X16_miniAODv2_v2-v1/MIN	80.95
/ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1/Sp16_80X16_miniAODv2_v2-v1/MIN	136.02
/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/Sp16_80X16_miniAODv2_v2-v1/MIN	35.85
/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/Sp16_80X16_miniAODv2_v2-v2/MIN	35.85
/GluGluHToBB_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6*/MIN	25.34
/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.01212
/GluGluHToWWTo2L2Nu_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6-v1/MIN	43.92
/VBFHToBB_M125_13TeV_amcatnlo_pythia8/Sm16_80X16_TIV_v6-v1/MIN	2.1626
/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.001034
/VBFHToWWTo2L2Nu_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6*/MIN	3.748
/WH_HToBB_WToLNu_M125_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.173
/WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV6_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.0001471
/WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV6_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.0002339
/ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6*/MIN	0.173
/ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6*/MIN	0.006954
/GluGluZH_HToWW_M125_13TeV_powheg_pythia8/Sm16_80X16_TIV_v6-v1/MIN	0.8696
/ZH_HToZZ_2LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV709_pythia8/Sm16_HIG083_80X16_TIV_v6-v2/MIN	0.7534
/ttHJetToNonbb_M125_13TeV_amcatnloFXFX_madspin_pythia8_mWCutfix/Sm16_80X16_TIV_v6_ext1-v1/MIN	0.2151
/ttHJetTobb_M125_13TeV_amcatnloFXFX_madspin_pythia8/Sm16_80X16_TIV_v6_ext3-v1/MIN	0.2934

5.3 Event Selection

Events are selected with a set of double muon triggers with asymmetric thresholds in transverse momentum of $p_{\mathrm{T}} > 17\,\mathrm{GeV}$, and $p_{\mathrm{T}} > 8\,\mathrm{GeV}$ or a double electron trigger with the corresponding thresholds of $p_{\mathrm{T}} > 23\,\mathrm{GeV}$, and $p_{\mathrm{T}} > 12\,\mathrm{GeV}$ detailed in Sec. 5.2.1. All final state objects are reconstructed using particle flow (PF) algorithms described in Sec. 3.2.

5.3.1 Object identification

Muons are reconstructed as tracks in the muon system that are matched to the tracks reconstructed in the inner silicon tracking system [137]. The leading muon is required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, while the sub-leading muon must have $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. Muons are required to be reconstructed in the HLT fiducial volume, i.e. with $|\eta| < 2.4$, to ensure that the offline selection is at least as restrictive as the HLT requirements. The selected muons are required to satisfy a set of identification requirements based on the number of spatial measurements in the silicon tracker and in the muon system and the quality of the fit of the combined muon track [54].

Electrons are reconstructed by matching tracks in the silicon tracker to the clusters of energy deposited in the ECAL [138]. The leading (sub-leading) electron is required to have $p_{\mathrm{T}} > 25(15)\,\mathrm{GeV}$ and $|\eta| < 2.5$ to be within the geometrical acceptance, excluding candidates in the range $1.4442 < |\eta| < 1.566$, which is the transition region between the barrel and the forward ECAL detectors. Electrons are required to pass an identification requirement based on a multivariate analysis technique which combines information from various observables related to the shower shape in the ECAL and the quality of the matching between the tracks and the associated ECAL clusters [138]. They are further required to be consistent with originating from the primary vertex. Candidates that are identified as originating from photon conversions in the material of the detector are removed.

Both muons and electrons are required be isolated by requiring that the PF-based combined relative isolation with PU correction, be less than 0.25 and 0.15 respectively. The relative isolation is defined in Sec. 3.4.

Jets are reconstructed from the list of PF objects using the anti-k_{T} [58, 59] algorithm with a size parameter of $R = 0.4$. In order to reduce instrumental backgrounds and the contamination from pileup, selected jets are required to satisfy loose identification criteria [112]

based on the multiplicities and energy fractions carried by charged and neutral hadrons. The energy of reconstructed jets is calibrated using p_{T}- and η-dependent corrections described in Sec. 3.5. Corrections to the jet energies are propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$. The jets candidates that are assigned to be the ones from the decay of a Higgs and of a Z boson are required to have $p_T > 20\,\mathrm{GeV}$. Furthermore, jets are required to have a spatial separation from lepton candidates of $\Delta R > 0.3$.

Jets originating from b quarks are identified with the CMVA algorithm [64]. A jet is tagged as a b jet if the CMVA discriminant is above a certain threshold. The threshold is chosen to correspond to the medium working point of the algorithm, defined such that the misidentification rate for light-quark and gluon jets is about 1%. The b jet tagging efficiency for this working point is about 66% [63].

Simulated background and signal events are corrected for differences observed between data and simulation, in trigger efficiencies, in lepton p_{T}- and η-dependent identification and isolation efficiencies, and in b tagging efficiencies.

5.3.2 Jet Assignment and Event Selection

After object selection, an initial event selection is performed by requiring at least two muons or two electrons in each event. The two leptons (two muons or two electrons) are required to be oppositely charged. The invariant mass of the two leptons, $M_{\ell\ell}$, is required to be larger than 15 GeV. Four of the jets in an event are designated as the H and Z decay products. These jets are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and at least one of those must be b-tagged with a minimum requirement on the b tagging discriminant. We refer to this selection as the preselection.

Since the signal final state contains two b jets from the decay of a Higgs boson, and two jets of any flavor from the decay of a Z boson, it is important to carefully categorize the jets in

the event. Starting from a collection of jets identified as described above, the information from the b tagging discriminators as well as the kinematic properties of the jets are taken into account when assigning which jets belong to which particle's decay.

The following selection is applied to identify the b jets originating from the decay of the Higgs. The two jets with the highest b tagging scores above a certain minimum threshold (i.e. the jets with the highest probability of being b jets) are assigned to the decay of the Higgs. If only one jet is found that meets the minimum b tagging score value, a second jet that leads to an invariant mass closest to 125 GeV is selected. If no jets with b tagging scores above threshold are found, the two jets whose invariant mass is closest to 125 GeV are chosen.

After jets are assigned to the decay of $H \to \mathrm{b\bar{b}}$, the remaining jets which reconstruct to the four-object invariant mass $M(\ell\ell\mathrm{jj})$ closest to 125 GeV are assigned to the decay of the Z boson.

After preselection, additional requirements are imposed. At least one of the four jets assigned as the decay products of the H or Z boson must satisfy the medium b tagging working point, to increase the signal-to-background ratio. To impose orthogonality with the $\mathrm{b\bar{b}}\ell\ell\nu\nu$ decay channel [131], of which results are to be combined with the results of this analysis at the end, upper limits on the $p_{\mathrm{T}}^{\mathrm{miss}}$ are imposed as follows: $p_{\mathrm{T}}^{\mathrm{miss}} < 40\,\mathrm{GeV}$ for $260 \leq M_X < 350\,\mathrm{GeV}$, $p_{\mathrm{T}}^{\mathrm{miss}} < 75\,\mathrm{GeV}$ for $350 \leq M_X < 650\,\mathrm{GeV}$, and $p_{\mathrm{T}}^{\mathrm{miss}} < 100\,\mathrm{GeV}$ for $M_X \geq 650\,\mathrm{GeV}$. We refer to this selection as the final selection in the $\mathrm{b\bar{b}}\ell\ell\mathrm{jj}$ channel.

5.4 Background Estimation

The main processes that can mimic the signature of the signal in the $\mathrm{b\bar{b}}\ell\ell\mathrm{jj}$ channel are $Z/\gamma^* + \mathrm{jets}$, $\mathrm{t\bar{t}}$, QCD multijets, diboson +jets, $W + \mathrm{jets}$, and SM Higgs production. The

contributions from the latter three processes are smaller and are estimated from simulation.

The contribution from the principal background, $Z/\gamma^* + \text{jets}$, is estimated with simulated events normalized to the data at the preselection level in the Z-enriched region $80 < M_{\ell\ell} < 100\,\text{GeV}$. The contribution from $\mathrm{t\bar{t}}$ is estimated in a similar manner, with the $\mathrm{t\bar{t}}$-enriched region defined by $M_{\ell\ell} > 100\,\text{GeV}$, and $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\text{GeV}$. The data to simulation normalization factors derived from the two control regions are $R_{\mathrm{Z}} = 1.14 \pm 0.01$ (stat) and $R_{\mathrm{t\bar{t}}} = 0.91 \pm 0.01$ (stat) in the muon channel and $R_{\mathrm{Z}} = 1.24 \pm 0.01$ (stat) and $R_{\mathrm{t\bar{t}}} = 0.97 \pm 0.02$ (stat) in the electron channel.

The smaller contribution from QCD multijet processes is determined from data with a method that makes use of the fact that neither signal events nor events from other backgrounds produce final states with same-charge leptons at any significant level. Data events with same-sign isolated leptons are used to model the shape of the multijet background, after all non-QCD sources of background contributing to this selection are subtracted using simulation. The yield in this region is normalized with the ratio of the number of events with non-isolated opposite-sign leptons, to the number events with non-isolated same-sign leptons. Here, non-isolated leptons are those muons (electrons) that fail the relative isolation requirements described in Sec. 3.4. All non-QCD sources of background, estimated with simulated events, are subtracted from the numerator and the denominator before computing the ratio.

5.5 Signal Extraction

After final selection, twenty-two kinematic variables which have a high power of separation between signal and background processes are combined into a BDT discriminant. These variables are: $\Delta\Phi_{\ell 1,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}}$, the difference in azimuthal angle between the leading lepton and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$; $|cos(\theta^*_{CS})|$, $|cos(\theta^*_{b,Hbb})|$ and $|cos(\theta^*_{Z\ell\ell,HZZ})|$, the helicity angles, where θ^*_{CS} is defined

as the angle between the direction of the $H \to b\bar{b}$ candidate to the Collins-Soper reference frame (assuming each incoming particle in the scattering has an energy of 6.5 TeV); $\theta^*_{b,Hbb}$, the angle between the direction of the $H \to b\bar{b}$ candidate and the leading b jet; $\theta^*_{Z\ell\ell,HZZ}$, the angle between the direction of $Z \to \ell\ell$ candidate and the $H \to ZZ$ candidate; $|\phi_1|$, the angle between the $z' - z$ plane and the $H \to ZZ$ decay plane where z' is the direction of $H \to ZZ$ candidate; $|\phi_{1,Zjj}|$, the angle between the $z'_1 - z$ plane and the $Z \to jj$ decay plane where z'_1 is the direction of $Z \to jj$ candidate; $\Delta R_{\ell\ell}$, ΔR^H_{bb}, ΔR^Z_{jj}, $\Delta R_{\ell\ell,bb^H}$, $\Delta R_{\ell\ell,jj^Z}$, $\Delta R_{\ell 1 b 1}$, $\Delta R_{\ell 1 b 2}$, $\Delta R_{\ell 2 b 1}$, $\Delta R_{\ell 2 b 2}$, $\Delta R_{\ell 1 j 1}$, $\Delta R_{\ell 1 j 2}$, $\Delta R_{\ell 2 j 1}$, and $\Delta R_{\ell 2 j 2}$, the separation angles between objects; M^H_{bb}, M^Z_{jj}, and $M_{\ell\ell}$, the invariant masses.

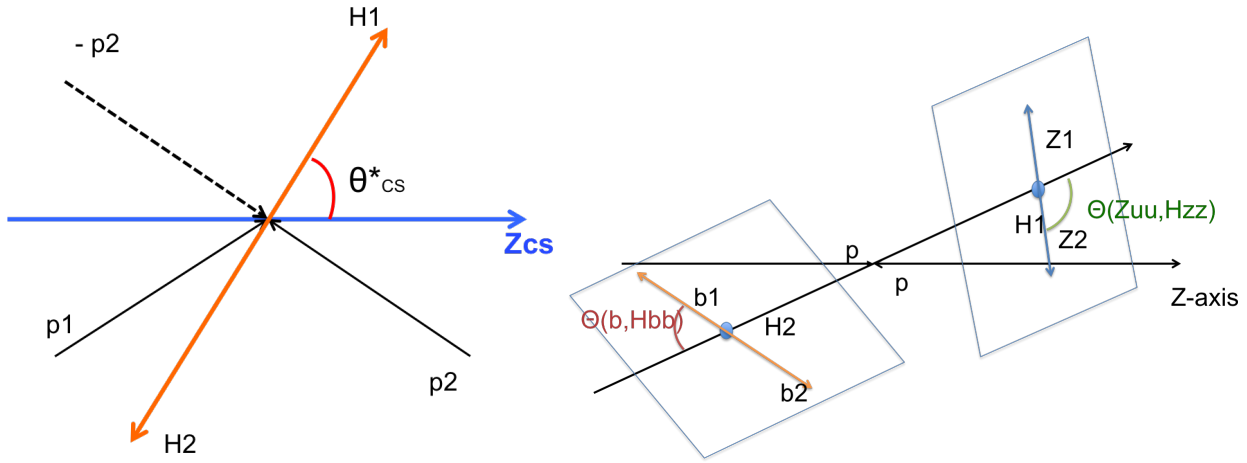


FIGURE 5.2: Diagrams illustrating the Collins-Soper (CS) reference frame (left) and the helicity angles of HH system, $|cos(\theta^*_{b,Hbb})|$ and $|cos(\theta^*_{Z\ell\ell,HZZ})|$ (right). The Collins-Soper (CS) reference frame is defined as the rest-frame of the Higgs-pair system. Since the Higgs-pair system has a transverse momentum, the directions of the initial protons are not collinear in this rest frame. The z-axis is defined in the Higgs-pair rest-frame such that it is bisecting the angle between the momentum of one of the protons and the inverse of the momentum of the second one. The sign of the z-axis is defined by the sign of the Higgs-pair momentum with respect to z-axis in the laboratory frame. To complete the coordinate system the y-axis is defined as the normal vector to the plane spanned by the two incoming proton momenta and the x-axis is chosen to set a right-handed Cartesian coordinate system with the other two axes.

In the resonance mass range of 260–300 GeV, the most important variables are $M_{\ell\ell}$, $\Delta R_{\ell 1 b 1}$, and M^H_{bb}. In the resonance mass range of 350–550 GeV, M^H_{bb} becomes the most important variable while $M_{\ell\ell}$ becomes less important and $\Delta R_{\ell\ell}$ gradually becomes more important

when the resonance mass increases. For masses higher than 550 GeV, $\Delta R_{\ell\ell}$ becomes the most important variable followed by M_{bb}^{H} and ΔR_{bb}^{H}.

The BDT is trained using all background processes described in Sec. 5.2, excluding the QCD multi-jet background, which is negligible. In each lepton channel and each spin hypothesis, one BDT is trained for each signal resonance mass simulated. In the training, signal events include samples from the two neighboring mass points, in addition to the targeted mass point. In total, 48 BDTs are trained. These BDT distributions for data and expected backgrounds are used as the final discriminating variable in the analysis.

5.6 Systematic Uncertainties

The dominant source of systematic uncertainty in this analysis is the jet energy scale uncertainty. The uncertainties in the jet energy scale are of the order of a few percent and are estimated as a function of jet p_T and η [60]. The η-dependent jet energy resolution correction factors are varied by ± 1 standard deviation in order to estimate the uncertainty of the jet energy resolution. Uncertainties in the jet energy are propagated to the calculation of $p_{\mathrm{T}}^{\mathrm{miss}}$.

An uncertainty of 2% per muon on the muon reconstruction, identification and isolation requirements, as well as a 1% per muon uncertainty on the muon HLT efficiency are assigned. A per-muon uncertainty due to to measured differences of tracking efficiency in data and simulation is estimated to be 0.5% for muon $p_T < 300$ and 1.0% for muon $p_T > 300$ [139]. Per-electron uncertainties on the electron HLT, identification and isolation requirements, estimated by varying the scale factors within their uncertainties, are applied. The uncertainties on the scale factors are generally less than 2% for HLT and less than 3% for identification and isolation. The effect of the variations on the yield of the total background is less than 1%. Uncertainties on the data-to-simulation correction factors of the b tagging and of light flavor mis-tagging efficiencies are included.

Normalization and shape uncertainties are assigned to the modeling of the backgrounds. Statistical uncertainties in the simulated samples for Z/γ^*+jets and $t\bar{t}$ background estimates result in uncertainties on the data-assisted normalization factors. A QCD scale shape uncertainty is determined by varying the factorization and the renormalization scales between its nominal value and 0.5 to 2.0 times nominal in the simulated signal and background samples. The non-physical variations where one scale increases and the other decreases are not considered. Each of the six possibilities of varying the renormalization and the factorization scales are considered, and the maximum variation among all the samples with respect to the central sample used in the analysis is taken as the systematic uncertainty, which is found to be 5–7% depending on the process. An uncertainty on the signal acceptance and background acceptance and cross section due to PDF uncertainties and to the value chosen for the strong coupling constant is estimated by varying the NNPDF set of eigenvectors within their uncertainties, following the PDF4LHC prescription [140].

An uncertainty of 2.5% is assigned to the determination of the integrated luminosity [141]. The uncertainty in the PU condition and modeling is assessed by varying the inelastic pp cross section from its central value by ±4.6%.

5.7 Results

Results are obtained by performing a binned maximum likelihood fit of the BDT distributions. The data and background predictions at final selection level are shown in Figure 5.3 and Figure 5.4, for the distributions of the BDT discriminant for signal masses of 500 and 1000 GeV, in the muon and electron final states, respectively.

In both channels the data distributions are well reproduced by the SM background processes. Upper limits on the resonance production cross section are set, using the asymptotic CL_s modified frequentist approach [124, 125].

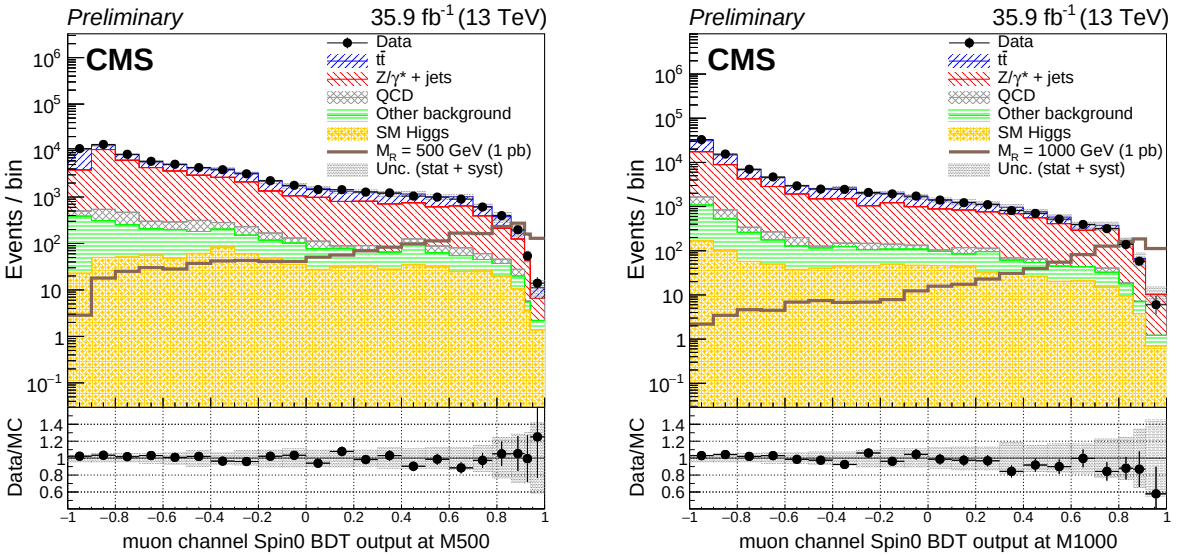


FIGURE 5.3: Comparison of the BDT discriminant for $m_X = 500\,\text{GeV}$ and $m_X = 1000\,\text{GeV}$ at final selection level in the muon channel of the $\text{b}\bar{\text{b}}\ell\ell\text{jj}$ channel. The signals of an RS1 radion with mass of 500 (left) and 1000 GeV (right) are normalized to 1 pb for the $\text{HH}\rightarrow \text{b}\bar{\text{b}}\text{ZZ} \rightarrow \text{b}\bar{\text{b}}\ell\ell\text{jj}$ process. The shaded area represents the combined statistical and systematic uncertainties.

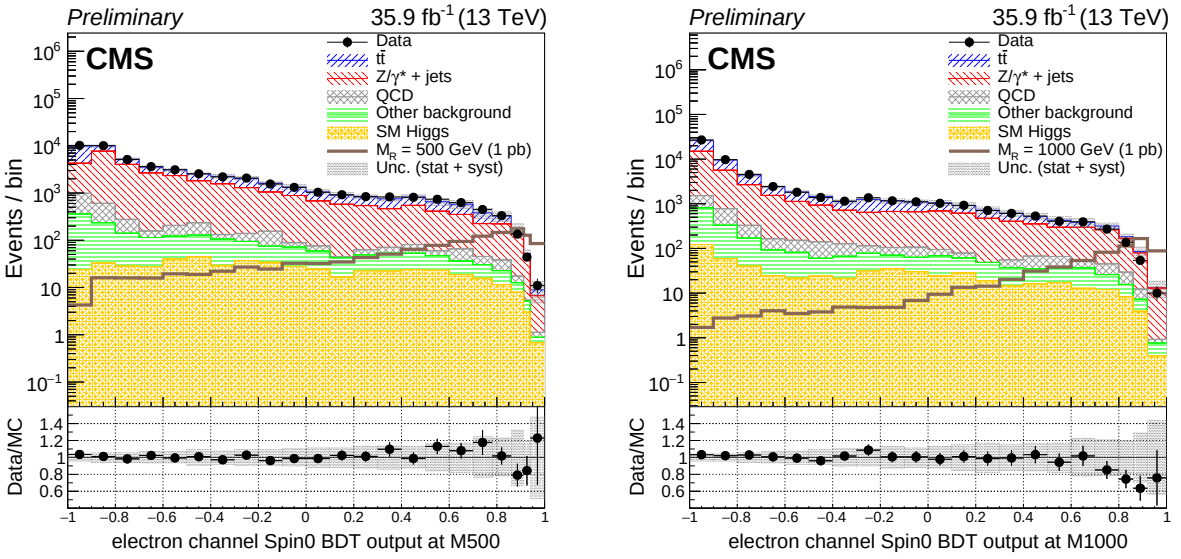


FIGURE 5.4: Comparison of the BDT discriminant for $m_X = 500\,\text{GeV}$ and $m_X = 1000\,\text{GeV}$ at final selection level in the electron channel of the $\text{b}\bar{\text{b}}\ell\ell\text{jj}$ channel. The signals of an RS1 radion with mass of 500 (left) and 1000 GeV (right) are normalized to 1 pb for the $\text{HH}\rightarrow \text{b}\bar{\text{b}}\text{ZZ} \rightarrow \text{b}\bar{\text{b}}\ell\ell\text{jj}$ process. The shaded area represents the combined statistical and systematic uncertainties.

The systematic uncertainties are represented by nuisance parameters that are varied in the fit according to their probability density functions, prescribed as follows. A log-normal probability density function is assumed for the nuisance parameters affecting the event yields of the various background contributions, whereas systematic uncertainties that affect the distributions are represented by nuisance parameters whose variation results in a continuous perturbation of the spectrum and which are assumed to have a Gaussian probability density function.

The statistical uncertainty from the limited number of events in the simulated samples is taken into account for each bin of the discriminant distributions by assigning a nuisance parameter to scale the sum of the process yields in that bin according to the statistical uncertainty using the Barlow-Beeston-lite prescription [142, 143].

The observed and expected 95% CL upper limits on $\sigma(\mathrm{pp}\to\mathrm{X}\to\mathrm{HH})$ in the $\mathrm{b\bar{b}}\ell\ell\mathrm{jj}$ channel as a function of RS1 radion and RS1 KK graviton mass are shown in Figure 5.5, together with the NLO predictions for the resonances production cross sections.

5.8 Combination with the $\mathrm{b\bar{b}}\ell\ell\nu\nu$ Channel

The search for resonant double Higgs production with bbZZ decays in the $\mathrm{b\bar{b}}\ell\ell\nu\nu$ final state was recently performed [131]. The results are combined with $\mathrm{b\bar{b}}\ell\ell\mathrm{jj}$ studied in this thesis to achieve more sensitivity in the search. The requirement on $p_{\mathrm{T}}^{\mathrm{miss}}$ mentioned in Sec. 5.3.2 ensures the orthogonality between both channels.

The 95% upper CL limits for the $\mathrm{b\bar{b}}\ell\ell\nu\nu$ channel are shown in Figure 5.6 [131]. As two different BDTs are defined for the search in the low and high mass ranges, the limit calculation is performed with both of the BDTs at the boundary of the two ranges, around 450 GeV,

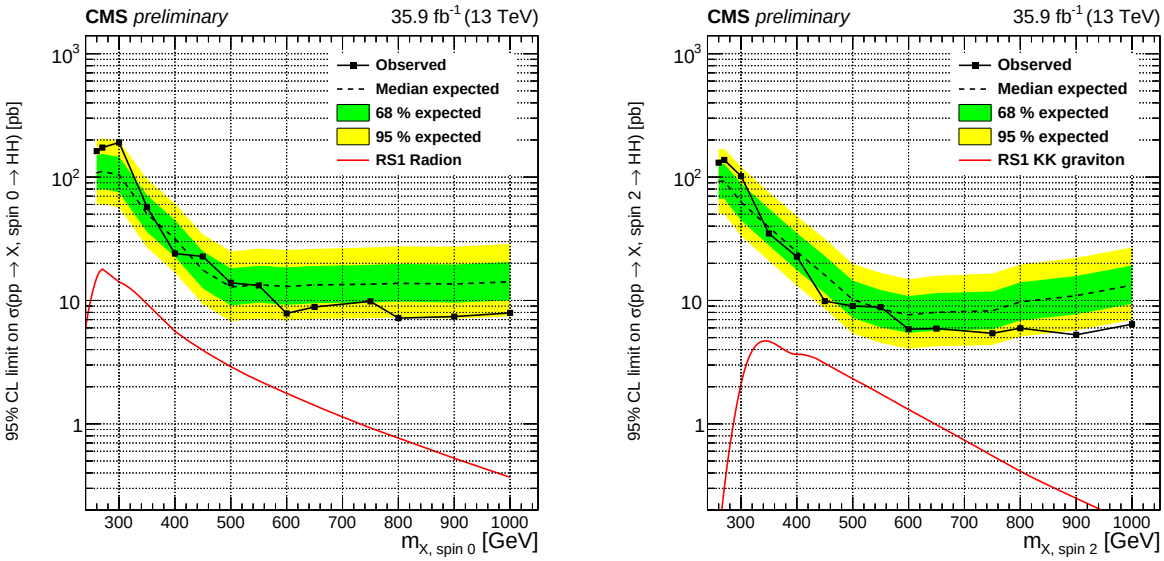


FIGURE 5.5: Expected (black dashed line) and observed (black solid line) limits on the cross section of resonant HH production as a function of the mass of the resonance for the $\mathrm{b\bar{b}}\ell\ell\mathrm{jj}$ channel. The RS1 radion case is shown on the left and the RS1 KK graviton case is shown on the right. The red solid lines show the theoretical prediction for the cross section of an RS1 radion with $\lambda_R = 1\,\mathrm{TeV}$ and $kL = 35$ (left) and an RS1 KK graviton with $\tilde{k} = 0.1$ (right).

where a discontinuity is seen. Also shown are the expected production cross sections as a function of resonance mass for an RS1 radion and an RS1 KK graviton.

Finally, the combined 95% CL upper limits from both channels on $\sigma(\mathrm{pp}\to\mathrm{X}\to\mathrm{HH})$ as a function of resonance mass are shown in Figure 5.7. Limits on the production cross section of RS1 radion and RS1 KK graviton range from 3.95 to 164 pb and 3.82 to 117 pb, respectively, in the range of resonance masses between 260 GeV and 1000 GeV.

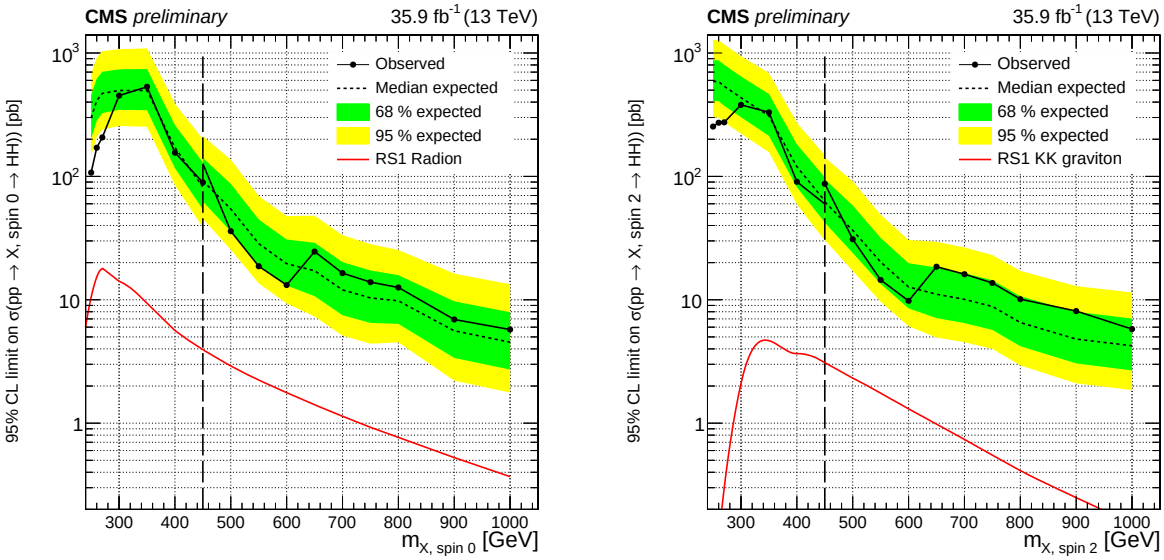


FIGURE 5.6: Expected (black dashed line) and observed (black solid line) limits on the cross section of resonant HH production as a function of the mass of the resonance for the $b\bar{b}\ell\ell\nu\nu$ channel. The RS1 radion case is shown on the left and the RS1 KK graviton case is shown on the right. The red solid lines show the theoretical prediction for the cross section of an RS1 radion with $\lambda_R = 1\,\text{TeV}$ and $kL = 35$ (left) and an RS1 KK graviton with $\tilde{k} = 0.1$ (right). The vertical black dashed line indicates the resonance mass of 450 GeV, a mass point where the BDT used in the analysis is switched from the one trained for low mass resonance to the one trained for high mass resonance.

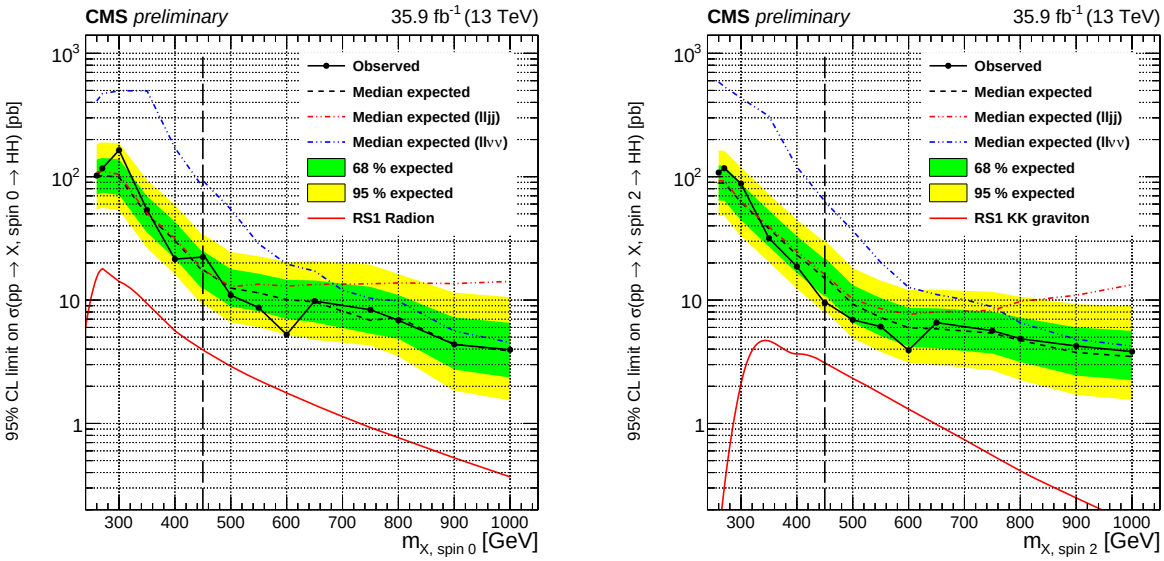


FIGURE 5.7: Expected (black dashed line) and observed (black solid line) limits on the cross section of resonant HH production as a function of the mass of the resonance for the combination of the $b\bar{b}\ell\ell jj$ and $b\bar{b}\ell\ell\nu\nu$ channels. The RS1 radion case is shown on the left and the RS1 KK graviton case is shown on the right. The red solid lines show the theoretical prediction for the cross section of an RS1 radion with $\lambda_R = 1\,\text{TeV}$ and $kL = 35$ (left) and an RS1 KK graviton with $\tilde{k} = 0.1$ (right). The expected limits for each individual channel are also shown with red dashed line for the $b\bar{b}\ell\ell jj$ channel and blue dashed line for the $b\bar{b}\ell\ell\nu\nu$ channel.

Chapter 6

Conclusion

This thesis presented the measurements of differential cross sections for a W boson produced in association with jets in pp collisions at a center-of-mass energy of 8 TeV [66] and 13 TeV [67]. The data correspond to an integrated luminosity of 19.6 fb^{-1} and 2.2 fb^{-1} respectively and were collected with the CMS detector at the LHC. Cross sections measured using the muon decay mode of the W boson were presented as functions of the jet multiplicity and the transverse momenta and pseudorapidities (rapidities in the 13 TeV analysis) of the four leading jets and H_{T} for jet multiplicities up to four. Cross sections were also measured as functions of the azimuthal separation between the muon direction from the W boson decay and the direction of the n^{th} leading jet for up to four inclusive jets.

For the W+jets 8 TeV analysis, cross sections were also presented as functions of the dijet p_{T}, invariant mass, and several angular correlation variables: rapidity difference, azimuthal angle difference, and ΔR between p_{T}-ordered and rapidity-ordered jets. The dependence of the average number of jets on H_{T} and on rapidity differences between jets was examined.

For the W+jets 13 TeV analysis, cross sections as a function of the angular distance between the muon and the closest jet in events with at least one jet with large p_{T} was also measured.

For the W+jets 8 TeV analysis, the results were corrected for detector effects by means of regularized unfolding and compared with particle-level simulated predictions using MadGraph5+pythia6; sherpa 2 and MadGraph5_aMC@NLO+pythia8 (multileg NLO); and with the fixed order calculations of BlackHat+sherpa (NLO); and N_{jetti} NNLO.

For the W+jets 13 TeV analysis, the results were compared with the predictions of MadGraph5_aMC@NLO at leading-order (LO) accuracy (LO MG_aMC) and at next-to-LO (NLO) accuracy (NLO MG_aMC FxFx) as well as N_{jetti} NNLO.

The cross section as a function of H_{T} is underestimated by BlackHat+sherpa for $N_{\text{jets}} \geq 1$ because the contribution from W + ≥3 jets is missing from an NLO prediction of W + ≥1 jet. The corresponding predictions from sherpa 2 overestimate the cross section, particularly at high H_{T}.

The NNLO predictions for W+1 jet production were compared with the measured cross sections differential in leading jet p_{T}, H_{T}, and leading jet $|\eta|$ ($|y|$ for 13 TeV analysis) and agree with data within uncertainties. The measured cross section as a function of the angular distance between the muon and the closest jet, which is sensitive to electroweak emission of W bosons, is best described by the NNLO calculation.

The search for the resonant production of two Higgs bosons decaying to two bottom quarks and two Z bosons, where one of the Z bosons decays to two leptons and the other decays to two quarks of any flavor or to two neutrinos was also presented [144]. The search used 13 TeV proton-proton collision data recorded by the CMS detector and corresponding to an integrated luminosity of 35.9 fb^{-1}. The results are in agreement with SM predictions and 95% CL upper limits are set on the resonant, narrow width, spin-0 radion and spin-2 Kaluza-Klein graviton production cross sections in the range of resonance masses between 260 GeV and 1000 GeV. These are the first limits to date in this decay channel.

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. http://www.slac.stanford.edu/~mpeskin/QFT.html.

[2] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. 1984.

[3] D. Griffiths, *Introduction to elementary particles*. 2008.

[4] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[5] https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[6] J. Schwinger, *Quantum Electrodynamics. I. A Covariant Formulation*, Phys. Rev. **74** (1948) 1439–1461.

[7] F. J. Dyson, *The S Matrix in Quantum Electrodynamics*, Phys. Rev. **75** (1949) 1736–1755.

[8] R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (1950) 440–457.

[9] S. L. Glashow, *The renormalizability of vector meson interactions*, Nucl. Phys. **10** (1959) 107–117.

[10] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367–377.

[11] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[12] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. **47B** (1973) 365–368.

[13] C.-N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191–195, [,150(1954)].

[14] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[15] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[16] UA1 Collaboration, G. Arnison et al., *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, Phys. Lett. **B122** (1983) 103–116, [,611(1983)].

[17] UA2 Collaboration, M. Banner et al., *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. **B122** (1983) 476–485, [,7.45(1983)].

[18] CDF Collaboration, T. Aaltonen et al., *Measurement of the cross section for* W^- *boson production in association with jets in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$*-TeV*, Phys. Rev. **D77** (2008) 011108, arXiv:0711.4044 [hep-ex].

[19] D0 Collaboration, V. M. Abazov et al., *Measurements of inclusive* W*+jets production rates as a function of jet transverse momentum in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Lett. **B705** (2011) 200–207, arXiv:1106.1457 [hep-ex].

[20] D0 Collaboration, V. M. Abazov et al., *Studies of W boson plus jets production in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D88** (2013) 092001, arXiv:1302.6508 [hep-ex].

[21] ATLAS Collaboration, G. Aad et al., *Measurements of the W production cross sections in association with jets with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 82, arXiv:1409.8639 [hep-ex].

[22] CMS Collaboration, V. Khachatryan et al., *Differential cross section measurements for the production of a W boson in association with jets in proton–proton collisions at* $\sqrt{s} = 7$ *TeV*, Phys. Lett. **B741** (2015) 12–37, arXiv:1406.7533 [hep-ex].

[23] ATLAS Collaboration, M. Aaboud et al., *Measurement of* W *boson angular distributions in events with high transverse momentum jets at* $\sqrt{s} = 8$ *TeV using the ATLAS detector*, Phys. Lett. B **765** (2017) 132, arXiv:1609.07045 [hep-ex].

[24] ATLAS Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[25] CMS Collaboration, S. Chatrchyan et al., *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[26] CMS Collaboration, A. M. Sirunyan et al., *Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at* $\sqrt{s}$ *=13TeV*, Phys. Lett. **B778** (2018) 101–127, arXiv:1707.02909 [hep-ex].

[27] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922 [hep-ph].

[28] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1–102, arXiv:1106.0034 [hep-ph].

[29] P. Fayet, *Supergauge Invariant Extension of the Higgs Mechanism and a Model for the electron and Its Neutrino*, Nucl. Phys. **B90** (1975) 104–124.

[30] P. Fayet, *Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions*, Phys. Lett. **69B** (1977) 489.

[31] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett. **83** (1999) 3370–3373.

[32] LHC Higgs Cross Section Working Group Collaboration, J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347 [hep-ph].

[33] CMS Collaboration, A. M. Sirunyan et al., *Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{(s)} = 8$ TeV*, Phys. Rev. **D96** (2017) 072004, arXiv:1707.00350 [hep-ex].

[34] ATLAS Collaboration, G. Aad et al., *Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector*, Phys. Rev. **D92** (2015) 092004, arXiv:1509.04670 [hep-ex].

[35] CMS Collaboration, A. M. Sirunyan et al., *Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **122** (2019) 121803, arXiv:1811.09689 [hep-ex].

[36] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions at 13 TeV with the ATLAS experiment*, ATLAS CONF Note ATLAS-CONF-2018-043, 2018. http://cds.cern.ch/record/2638212.

[37] CMS Collaboration, *Summary of run-II $\sigma(gg \to X) \times B(X \to HH)$ 95% CL upper limits assuming spin-0 and spin-2 hypotheses*, Higgs PAG Summary Plots. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryResultsHIG#Summary_of_run_II_sigma_gg_to_X.

[38] P. Mouche, *Overall view of the LHC. Vue d'ensemble du LHC*, https://cds.cern.ch/record/1708847, General Photo.

[39] https://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/images/lhc-schematic.jpg.

[40] https://project-hl-lhc-industry.web.cern.ch/content/project-schedule.

[41] S. Dailler, *Cross section of LHC dipole. Dipole LHC: coupe transversale.*, AC Collection. Legacy of AC. Pictures from 1992 to 2002., Apr, 1999.

[42] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018*, http://cds.cern.ch/record/2636343, General Photo.

[43] CMS Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[44] T. Sakuma and T. McCauley, *Detector and Event Visualization with SketchUp at the CMS Experiment*, Journal of Physics: Conference Series **513** (2014) 022032, https://doi.org/10.1088%2F1742-6596%2F513%2F2%2F022032.

[45] CMS Collaboration, A. M. Sirunyan et al., *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, JINST **13** (2018) P06015, arXiv:1804.04528 [physics.ins-det].

[46] CMS Collaboration, *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. https://cds.cern.ch/record/343814.

[47] CMS Collaboration, S. Chatrchyan et al., *The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC*, JINST **8** (2013) P11002, arXiv:1306.6905 [physics.ins-det].

[48] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4—a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[49] T. Gleisberg, S. Hoeche, F. Krauss, A. Schalicke, S. Schumann, and J.-C. Winter, *SHERPA 1. alpha: A Proof of concept version*, JHEP **02** (2004) 056, arXiv:hep-ph/0311263 [hep-ph].

[50] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. L. Mangano, M. Moretti, C. G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek, *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. C **53** (2008) 473, arXiv:0706.2569 [hep-ph].

[51] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012) 061, arXiv:1209.6215 [hep-ph].

[52] CMS Collaboration, S. Chatrchyan et al., *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, JINST **9** (2014) P10009, arXiv:1405.6569 [physics.ins-det].

[53] CMS Collaboration, A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, JINST **12** (2017) P10003, arXiv:1706.04965 [physics.ins-det].

[54] CMS Collaboration, S. Chatrchyan et al., *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, JINST **7** (2012) P10002, arXiv:1206.4071 [physics.ins-det].

[55] CMS Collaboration, V. Khachatryan et al., *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV*, JINST **10** (2015) P06005, arXiv:1502.02701 [physics.ins-det].

[56] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B **659** (2008) 119, arXiv:0707.1378.

[57] CMS Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, Tech. Rep. CMS-PAS-JME-13-002, CERN, Geneva, 2013. https://cds.cern.ch/record/1599732.

[58] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[59] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[60] CMS Collaboration, V. Khachatryan et al., *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, JINST **12** (2017) P02014, arXiv:1607.03663 [hep-ex].

[61] CMS Collaboration, S. Chatrchyan et al., *Determination of jet energy calibration and transverse momentum resolution in CMS*, JINST **6** (2011) 11002, arXiv:1107.4277.

[62] CMS Collaboration, V. Khachatryan et al., *Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV*, JINST **10** (2015) P02006, arXiv:1411.0511 [hep-ex].

[63] CMS Collaboration, A. M. Sirunyan et al., *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, JINST **13** (2018) P05011, arXiv:1712.07158 [physics.ins-det].

[64] CMS Collaboration, *Identification of b quark jets at the CMS experiment in the LHC Run 2*, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016. http://cds.cern.ch/record/2138504.

[65] CMS Collaboration, S. Chatrchyan et al., *Identification of b-quark jets with the CMS experiment*, JINST **8** (2013) P04013, arXiv:1211.4462 [hep-ex].

[66] CMS Collaboration, V. Khachatryan et al., *Measurements of differential cross sections for associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Rev. D **95** (2017) 052002, arXiv:1610.04222 [hep-ex].

[67] CMS Collaboration, A. M. Sirunyan et al., *Measurement of the differential cross sections for the associated production of a W boson and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. **D96** (2017) 072005, arXiv:1707.05979 [hep-ex].

[68] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. A **362** (1995) 487.

[69] T. Adye, *Unfolding algorithms and tests using RooUnfold*, p. , 313. arXiv:1105.1160 [physics.data-an].
https://inspirehep.net/record/898599/files/arXiv:1105.1160.pdf.

[70] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: going beyond*, JHEP **06** (2011) 128, arXiv:1106.0522 [hep-ph].

[71] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[72] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[73] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **12** (2008) 039, arXiv:0808.3674 [hep-ph].

[74] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv:0709.1027 [hep-ph].

[75] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[76] Z. Bern, L. J. Dixon, F. Febres Cordero, S. Höche, H. Ita, D. A. Kosower, and D. Maître, *Ntuples for NLO events at hadron colliders*, Comput. Phys. Commun. **185** (2014) 1443, arXiv:1310.7439 [hep-ph].

[77] R. Boughezal, C. Focke, X. Liu, and F. Petriello, *W-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD*, Phys. Rev. Lett. **115** (2015) 062002, arXiv:1504.02131 [hep-ph].

[78] R. Boughezal, X. Liu, and F. Petriello, *W-boson plus jet differential distributions at NNLO in QCD*, Phys. Rev. D **94** (2016) 113009, arXiv:1602.06965 [hep-ph].

[79] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[80] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[81] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[82] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, arXiv:0907.4076 [hep-ph].

[83] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through $O(\alpha_S^2)$*, Phys. Rev. D **74** (2006) 114017, arXiv:hep-ph/0609070 [hep-ph].

[84] J. Campbell, R. K. Ellis, and F. Tramontano, *Single top-quark production and decay at next-to-leading order*, Phys. Rev. D **70** (2004) 094012, arXiv:hep-ph/0408158 [hep-ph].

[85] J. Campbell and F. Tramontano, *Next-to-leading order corrections to Wt production and decay*, Nucl. Phys. B **726** (2005) 109, arXiv:hep-ph/0506289 [hep-ph].

[86] J. M. Campbell, R. Frederix, F. Maltoni, and F. Tramontano, *Next-to-Leading-Order Predictions for t-Channel Single-Top Production at Hadron Colliders*, Phys. Rev. Lett. **102** (2009) 182003, arXiv:0903.0005 [hep-ph].

[87] J. M. Campbell, R. K. Ellis, and C. Williams, *Vector boson pair production at the LHC*, JHEP **07** (2011) 018, arXiv:1105.0020 [hep-ph].

[88] M. Czakon, P. Fiedler, and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004, arXiv:1303.6254 [hep-ph].

[89] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky1, and W.-K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, arXiv:hep-ph/0201195 [hep-ph].

[90] CMS Collaboration, V. Khachatryan et al., *Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV*, JHEP **01** (2011) 079, arXiv:1011.5531.

[91] CMS Collaboration, *Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_{T}^{miss}*, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. http://cdsweb.cern.ch/record/1194487.

[92] CMS Collaboration, *Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector*, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010. http://cdsweb.cern.ch/record/1247373.

[93] J. Campbell and R. K. Ellis, *Next-to-leading order corrections to W+2 jet and Z+2 jet production at hadron colliders*, Phys. Rev. D **65** (2002) 113007, arXiv:hep-ph/0202176 [hep-ph].

[94] CMS Collaboration, S. Chatrchyan et al., *Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV*, Phys. Lett. B **722** (2013) 5, arXiv:1210.6718.

[95] CMS Collaboration, *Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update*, CMS Physics Analysis Summary CMS-PAS-SMP-12-008, 2012. http://cdsweb.cern.ch/record/1434360.

[96] NNPDF Collaboration, R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph].

[97] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph].

[98] NNPDF Collaboration, M. Ubiali, *Towards the NNPDF3.0 parton set for the next LHC run*, in *Proceedings of the XXII Int'l Workshop on Deep-Inelastic Scattering and Related Subjects*, p. , 041. Warsaw, Poland, May, 2014. arXiv:1407.3122. http://pos.sissa.it/archive/conferences/203/041/DIS2014_041.pdf. PoS(DIS2014)041.

[99] CMS Collaboration, V. Khachatryan et al., *Event generator tunes obtained from underlying event and multiparton scattering measurements*, Eur. Phys. J. C **76** (2016) 155, arXiv:1512.00815 [hep-ex].

[100] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, R. Pittau, and P. Torrielli, *Four-lepton production at hadron colliders: aMC@NLO predictions with theoretical uncertainties*, JHEP **02** (2012) 099, arXiv:1110.4738 [hep-ph].

[101] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, H. Ita, D. A. Kosower, and Maître, *One-Loop Calculations with BlackHat*, in *Proceedings of the 9th DESY Workshop on Elementary Particle Theory*. 2008. arXiv:0807.3705 [hep-ph]. Nucl. Phys. B, Proc. Suppl. 183 (2008) 313.

[102] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and Maître, *Vector Boson + Jets with BlackHat and SHERPA*, in *Loops and Legs in Quantum Field Theory — Proccedings of the 10th DESY Workshop on Elementary Particle Theory*. 2010. arXiv:1005.3728 [hep-ph]. Nucl. Phys. B, Proc. Suppl. 205-206 (2010) 92.

[103] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[104] S. Hoeche, F. Krauss, M. Schonherr, and F. Siegert, *A critical appraisal of NLO+PS matching methods*, JHEP **09** (2012) 049, arXiv:1111.1220 [hep-ph].

[105] C. F. Berger, Z. Bern, L. J. Dixon, F. F. Cordero, D. Forde, T. Gleisberg, H. Ita, D. A. Kosower, and D. Maïtre, *Next-to-leading order QCD predictions for W+3-jet*

distributions at hadron colliders, Phys. Rev. D **80** (2009) 074036, arXiv:0907.1984 [hep-ph].

[106] D. Maître and S. Sapeta, *Simulated NNLO for high-p_T observables in vector boson + jets production at the LHC*, Eur. Phys. J. C **73** (2013) 2663, arXiv:1307.2252.

[107] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[108] T. Melia, P. Nason, R. Rontsch, and G. Zanderighi, W^+W^-, WZ *and* ZZ *production in the POWHEG BOX*, JHEP **11** (2011) 078, arXiv:1107.5051 [hep-ph].

[109] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012 [hep-ph].

[110] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl. Phys. B **838** (2010) 136, arXiv:1002.4407 [hep-ph].

[111] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Impact of heavy quark masses on parton distributions and LHC phenomenology*, Nucl. Phys. B **849** (2011) 296, arXiv:1101.1300 [hep-ph].

[112] CMS Collaboration, *Jet algorithms performance in 13 TeV data*, CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017. http://cds.cern.ch/record/2256875.

[113] CMS Collaboration, *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*, CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016. http://cds.cern.ch/record/2205284.

[114] U. Baur, *Weak boson emission in hadron collider processes*, Phys. Rev. D **75** (2007) 013005, arXiv:hep-ph/0611241 [hep-ph].

[115] J. R. Christiansen and T. Sjöstrand, *Weak gauge boson radiation in parton showers*, JHEP **04** (2014) 115, arXiv:1401.5238 [hep-ph].

[116] F. Krauss, P. Petrov, M. Schönherr, and M. Spannowsky, *Measuring collinear W emissions inside jets*, Phys. Rev. D **89** (2014) 114006, arXiv:1403.4788 [hep-ph].

[117] R. Boughezal, C. Focke, and X. Liu, *Jet vetoes versus giant K-factors in the exclusive Z+1-jet cross section*, Phys. Rev. D **92** (2015) 094002, arXiv:1501.01059 [hep-ph].

[118] J. R. Christiansen and S. Prestel, *Merging weak and QCD showers with matrix elements*, Eur. Phys. J. C **76** (2016) 39, arXiv:1510.01517 [hep-ph].

[119] W. H. Richardson, *Bayesian-based iterative method of image restoration*, J. Opt. Soc. Am. **62** (1972) 55.

[120] L. B. Lucy, *An iterative technique for the rectification of observed distributions*, Astron. J. **79** (1974) 745.

[121] J. Alwall, S. de Visscher, and F. Maltoni, *QCD radiation in the production of heavy colored particles at the LHC*, JHEP **02** (2009) 017, arXiv:0810.5350 [hep-ph].

[122] S. Gangal and F. J. Tackmann, *Next-to-leading-order uncertainties in Higgs+2 jets from gluon fusion*, Phys. Rev. D **87** (2013) 093008, arXiv:1302.5437 [hep-ph].

[123] I. W. Stewart and F. J. Tackmann, *Theory uncertainties for Higgs and other searches using jet bins*, Phys. Rev. D **85** (2012) 034011, arXiv:1107.2117 [hep-ph].

[124] T. Junk, *Confidence Level Computation for Combining Searches with Small Statistics*, Nucl. Instrum. Meth. A **434** (1999) 435, hep-ex/9902006.

[125] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, CERN-OPEN 2000-205, 2000. http://cdsweb.cern.ch/record/451614. 1st Workshop on Confidence Limits, CERN, Jan. 2000.

[126] O. DeWolfe and M. B. Wise, *Modulus Stabilization with Bulk Fields*, Phys. Rev. Lett. **83** (1999) 4922–4925.

[127] W. D. Goldberg, D. Freedman, S. Gubser, and A. Karch, *Modeling the fifth dimension with scalars and gravity*, Phys. Rev. D **62** (2000) 046008.

[128] C. Csaki, M. Graesser, L. Randall, and J. Terning, *Cosmology of Brane Models with Radion Stabilization*, Phys. Rev. D **62** (1999) 045015.

[129] C. Csaki, J. Hubisz, and S. J. Lee, *Radion Phenomenology in Realistic Warped Space Models*, Phys. Rev. D **76** (2007) 125015.

[130] A. Oliveira, *Gravity particles from Warped Extra Dimensions, predictions for LHC*, arXiv:1404.0102 [hep-ph].

[131] CMS Collaboration, *Search for resonant double Higgs production with $bbZZ$ decays in the $b\bar{b}\ell\ell\nu\nu$ final state*, Tech. Rep. CMS-PAS-HIG-17-032, CERN, Geneva, 2018. http://cds.cern.ch/record/2648796.

[132] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC*, Nucl. Phys. Proc. Suppl. **205** (2010) 10–15, arXiv:1007.3492 [hep-ph].

[133] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, *W physics at the LHC with FEWZ 2.1*, Comput. Phys. Commun. **184** (2013) 209, arXiv:1201.5896 [hep-ph].

[134] CMS Collaboration, A. M. Sirunyan et al., *Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state of two muons and two τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **11** (2018) 018, arXiv:1805.04865 [hep-ex].

[135] CMS Collaboration, A. M. Sirunyan et al., *Evidence for the Higgs boson decay to a bottom quark–antiquark pair*, Phys. Lett. **B780** (2018) 501–532, arXiv:1709.07497 [hep-ex].

[136] CMS Collaboration, *Search for the Standard Model Higgs Boson Produced in Association with W and Z and Decaying to Bottom Quarks*, CMS Note 2015/168, 2015. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/168.

[137] CMS Collaboration, *Performance of muon identification in pp collisions at* $\sqrt{s} = 7$ *TeV*, CMS Physics Analysis Summary CMS-PAS-MUO-10-002, 2010. http://cdsweb.cern.ch/record/1279140.

[138] CMS Collaboration, *Electron Reconstruction and Identification at* $\sqrt{s} = 7$ *TeV*, CMS Physics Analysis Summary CMS-PAS-EGM-10-004, 2010. http://cdsweb.cern.ch/record/1299116.

[139] CMS Collaboration, *Search for new physics in l+MET channel with 2016 data*, CMS Note 2016/204, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2016/204.

[140] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, arXiv:1510.03865 [hep-ph].

[141] CMS Collaboration, *Current recommendations for luminosity estimations*, https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM (2016).

[142] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.

[143] J. S. Conway, *Incorporating nuisance parameters in likelihoods for multisource spectra*, p. , 115. 2011. arXiv:1103.0354 [physics.data-an].

[144] CMS Collaboration,, *Search for the resonant production of a pair of Higgs bosons decaying to the bb-barZZ final state*, Tech. Rep. CMS-PAS-HIG-18-013, CERN, Geneva, 2019. http://cds.cern.ch/record/2682621.

Appendix A

W+Jets Analysis: Additional Materials

A.1 Data-Simulation Comparisons in W+Jets 13 TeV

The goal of this analysis is to measure differential cross sections for jets produced in association with a W boson as a function of several kinematic observables with early 13 TeV data. We consider the jet multiplicity, both inclusive and exclusive, and the basic kinematic variables which are sensitive to higher order corrections, such as the transverse momentum p_{T} and absolute rapidity $|y|$ of the jets in the event and the total hadronic activity H_{T} in different jet multiplicity bins. We also examine the differential cross sections for angular variables of the azimuthal separation $\Delta\phi(\mu, \text{jet})$ and the angular distance $\Delta R(\mu, \text{closest jet})$ between the muon and the jets. We compare data with simulation at the detector level, where signal and background processes are simulated by MC based generators with the exception of QCD multijet background which is estimated using a data-driven method. We scale the $\mathrm{t\bar{t}}$ background events to take into account the data-simulation difference in $\mathrm{t\bar{t}}$-enriched control region for events with two or more (exclusive) inclusive jets as described in Sec. 4.7.3.

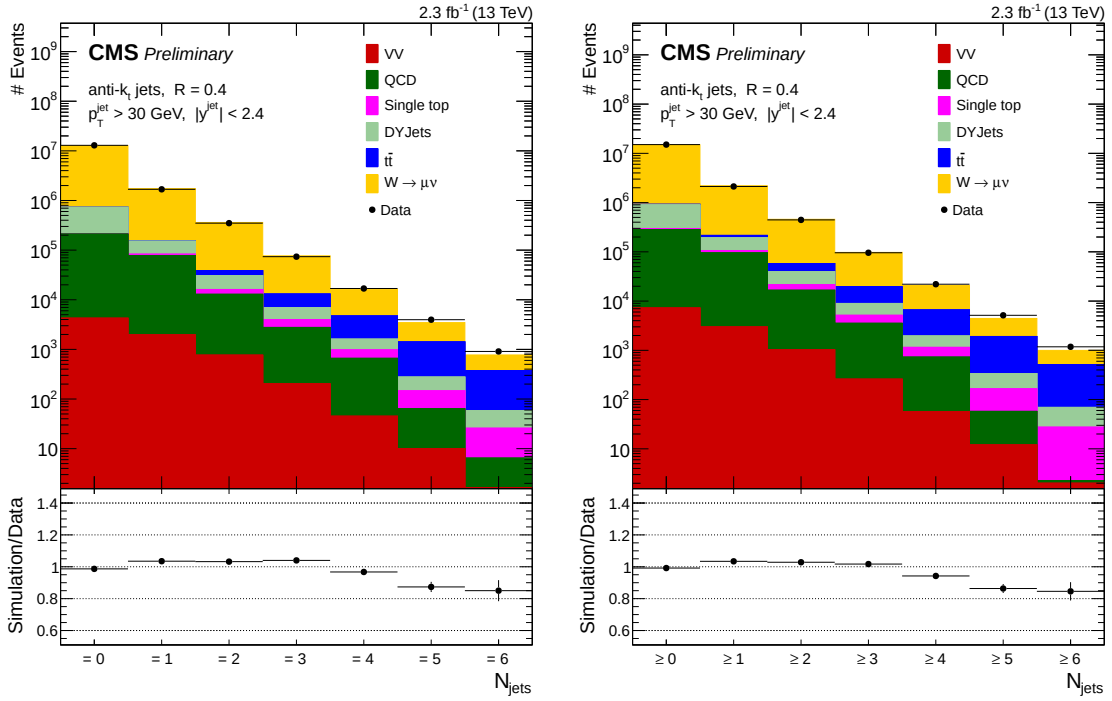


FIG. A.1: Data to simulation comparison of exclusive (left) and inclusive (right) jet multiplicity. The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

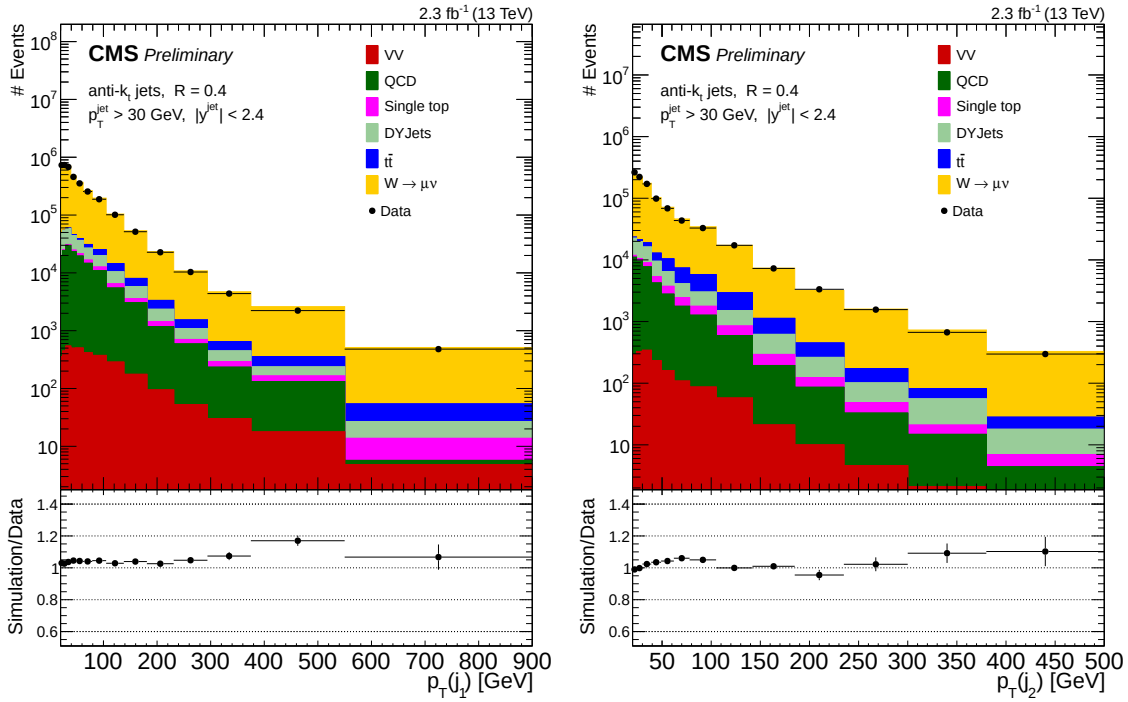


FIG. A.2: Data to simulation comparison of 1^{st} (left) and 2^{nd} (right) jet p_{T}. The QCD background is estimated using a data-driven method. The $\mathrm{t\bar{t}}$ background is scaled to compensate for the data-simulation difference in $\mathrm{t\bar{t}}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

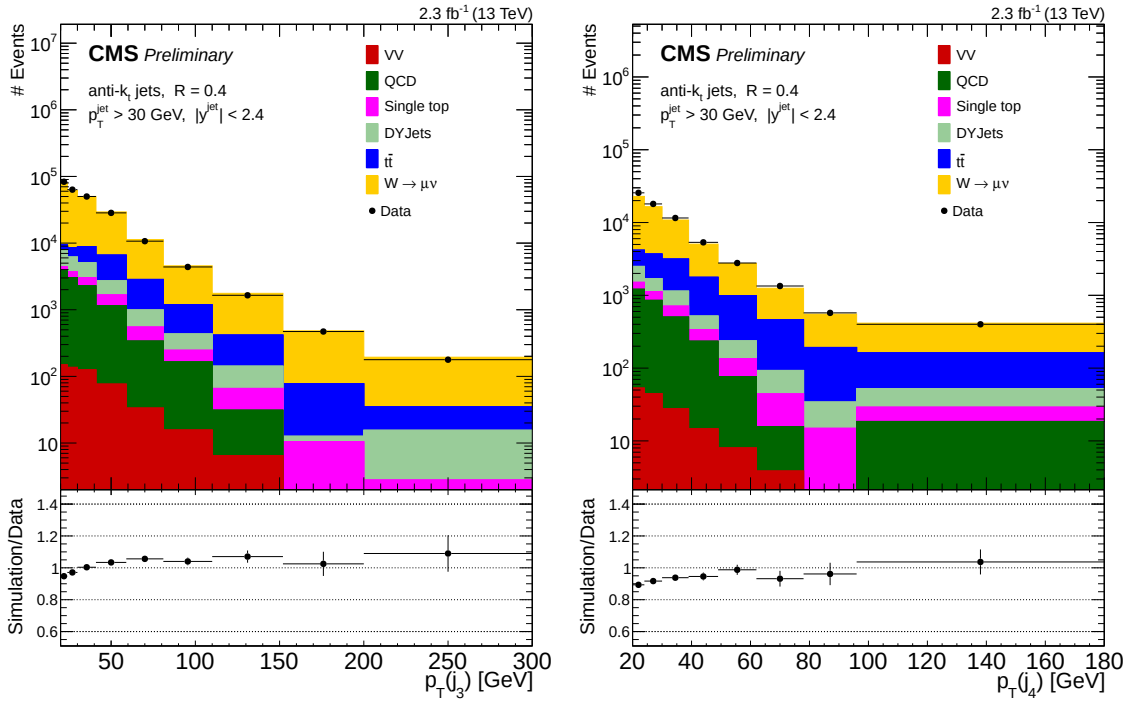


FIG. A.3: Data to simulation comparison of 3^{rd} (left) and 4^{th} (right) jet p_{T}. The QCD background is estimated using a data-driven method. The $\mathrm{t\bar{t}}$ background is scaled to compensate for the data-simulation difference in $\mathrm{t\bar{t}}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

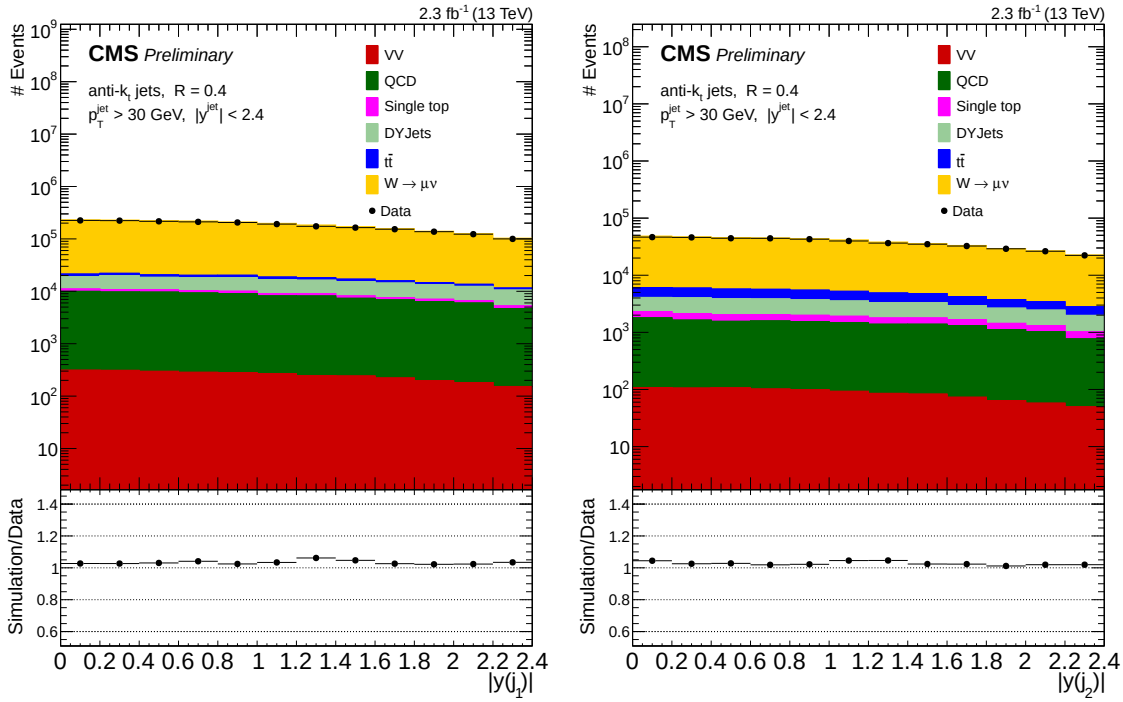


FIG. A.4: Data to simulation comparison of 1^{st} (left) and 2^{nd} jet y. The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

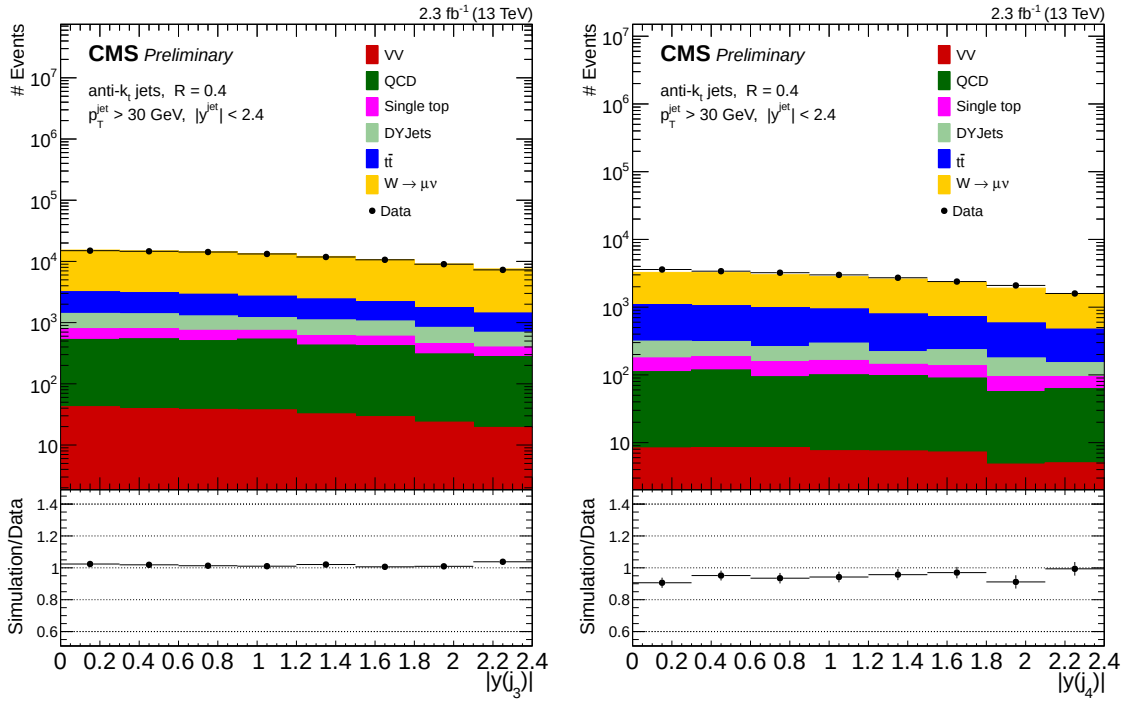


FIG. A.5: Data to simulation comparison of 3^{rd} (left) and 4^{th} (right) jet y. The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

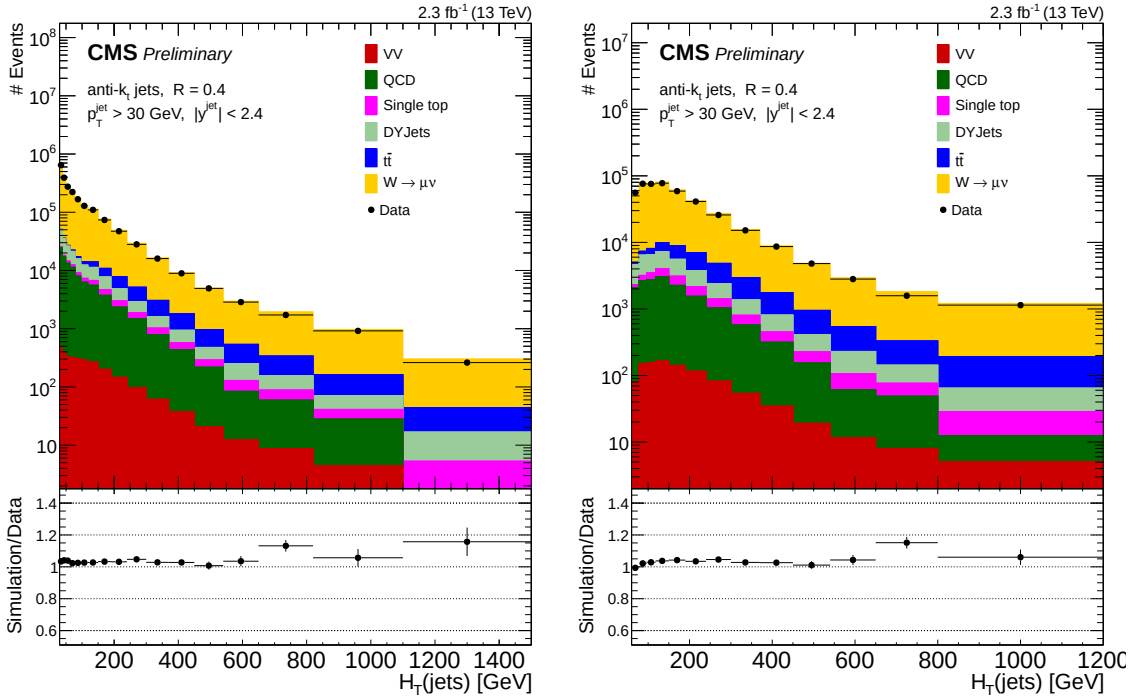


FIG. A.6: Data to simulation comparison of jet H_T for $N_{jets} \geq 1$ (left) and $N_{jets} \geq 2$ (right). The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

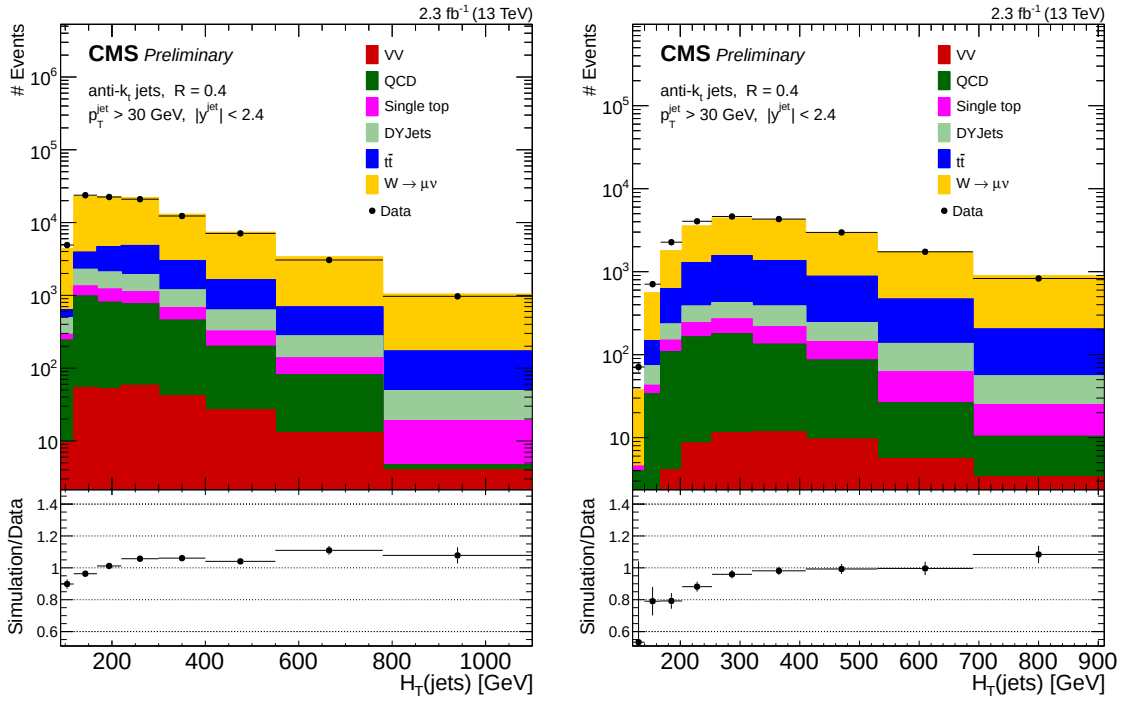


FIG. A.7: Data to simulation comparison of jet H_T for $N_{jets} \geq 3$ (left) and $N_{jets} \geq 4$ (right). The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

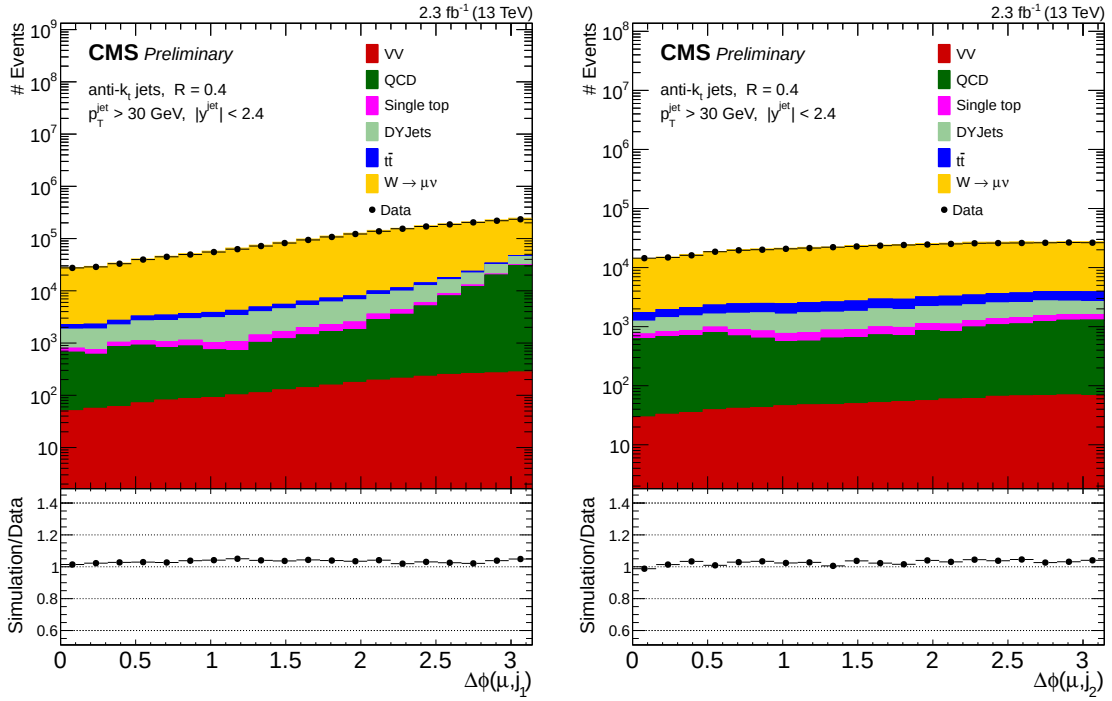


FIG. A.8: Data to simulation comparison of $\Delta\phi(\mu, \text{jet})$ for $N_{jets} \geq 1$ (left) and $N_{jets} \geq 2$ (right). The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

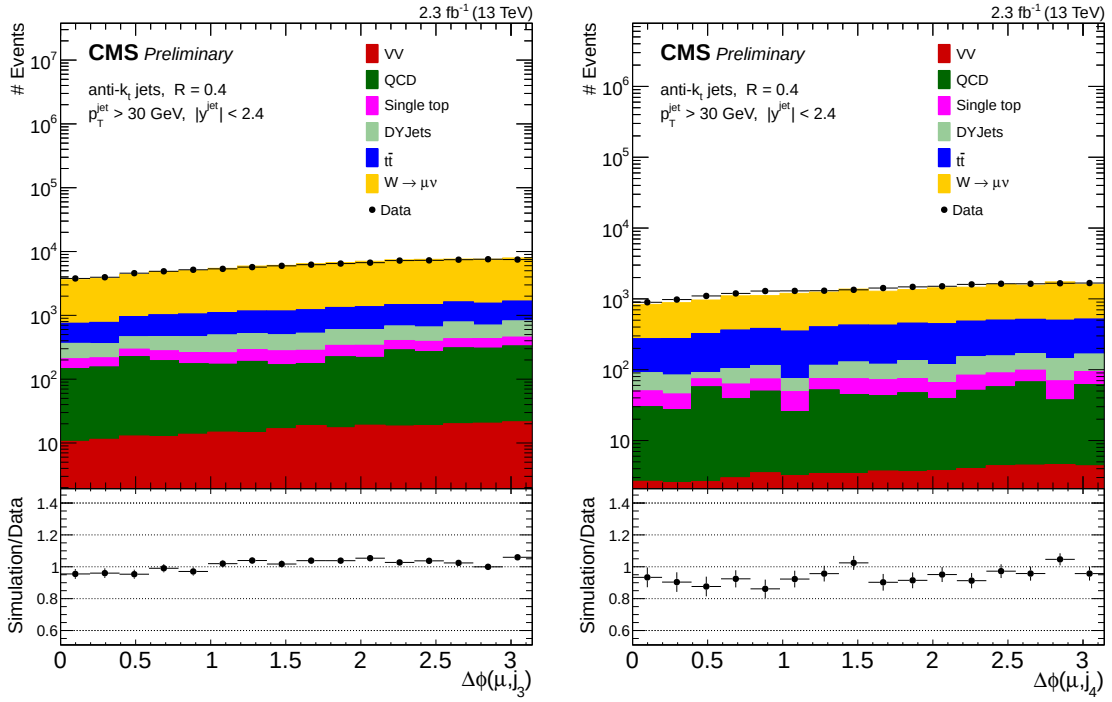


FIG. A.9: Data to simulation comparison of $\Delta\phi(\mu, \text{jet})$ for $N_{jets} \geq 3$ (left) and $N_{jets} \geq 4$ (right). The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

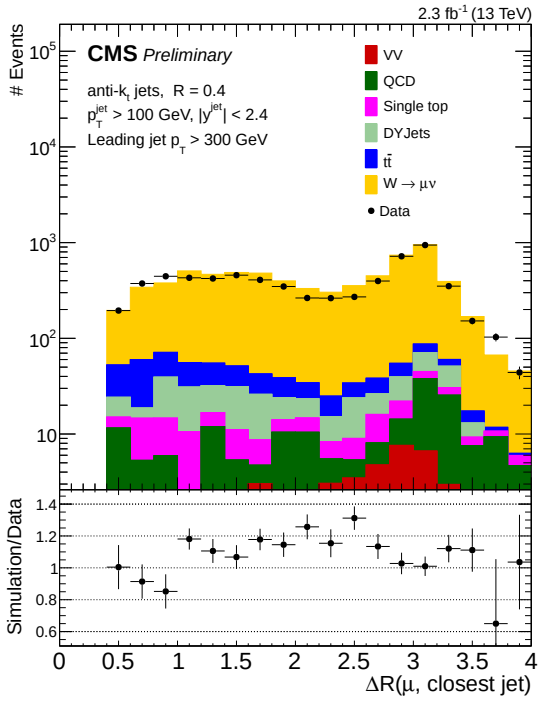


FIG. A.10: Data to simulation comparison of $\Delta R(\mu$, closest jet) for $N_{jets} \geq 1$. The QCD background is estimated using a data-driven method. The $t\bar{t}$ background is scaled to compensate for the data-simulation difference in $t\bar{t}$-enriched control region for events with at least two or more jets. The diboson samples (WW, WZ, and ZZ) are represented by VV.

Appendix B

HH Analysis: Additional Materials

B.1 Estimation of Z/γ^*+jets and $t\bar{t}$ backgrounds

As described in Sec. 5.4, the contributions from Z+jets and $t\bar{t}$, are estimated with MC simulation normalized to data at the preselection level in the Z-enriched region $80 < M_{\ell\ell} < 100$ GeV and in the $t\bar{t}$-enriched region $M_{\ell\ell} > 100$ GeV and $E_{\mathrm{T}}^{\mathrm{miss}} > 100$ GeV respectively. The distributions of the $M_{\ell\ell}$ in these control regions are shown in Fig. B.1. On the ratio subplots the statistical uncertainty of data and background is shown as bars on the black points and the full systematic uncertainty is shown as a shaded band.

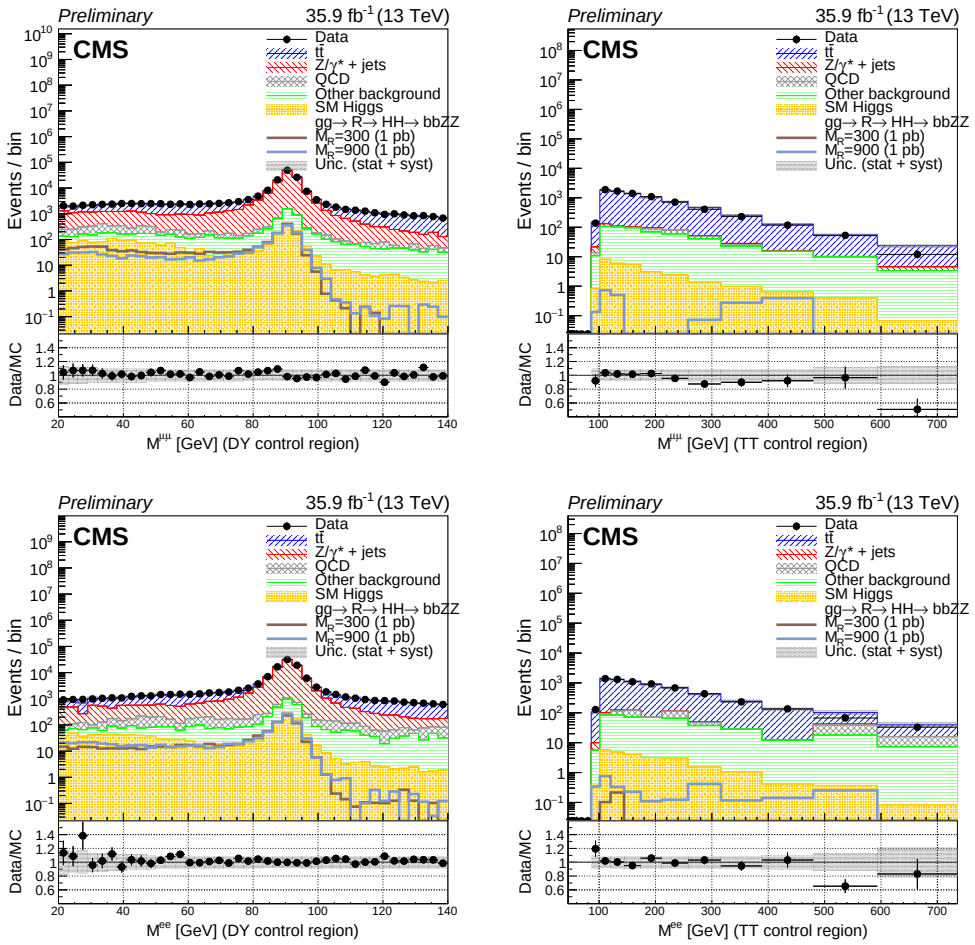


FIGURE B.1: Comparison of the $M_{\ell\ell}$ at preselection level in the Z-enriched region (left), and in the $t\bar{t}$-enriched region (right) in the muon (top) and electron (bottom) channels, after the Data/MC normalization scale factors have been applied.

B.2 Data-MC Comparisons of Discriminating Variables

This section compiles the data-simulation comparisons of discriminating variables used in the training of BDTs. The $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is not used in the training of BDTs but is included in this section because it is used as a requirement to ensure orthogonality with $\mathrm{b\bar{b}}\ell\ell\nu\nu$ analysis channel.

B.2.1 Muon channel preselection

Good agreement is found for all final state distributions at preselection level in the muon channel (Figures B.2– B.8). For all plots, the Z/γ^*+jets and $\mathrm{t\bar{t}}$ MC predictions are normalized to data as described in Sec. 5.4. On the ratio subplots, the statistical uncertainty of data and background is shown as a bar on the black points and the full systematic uncertainty is shown as a shaded band.

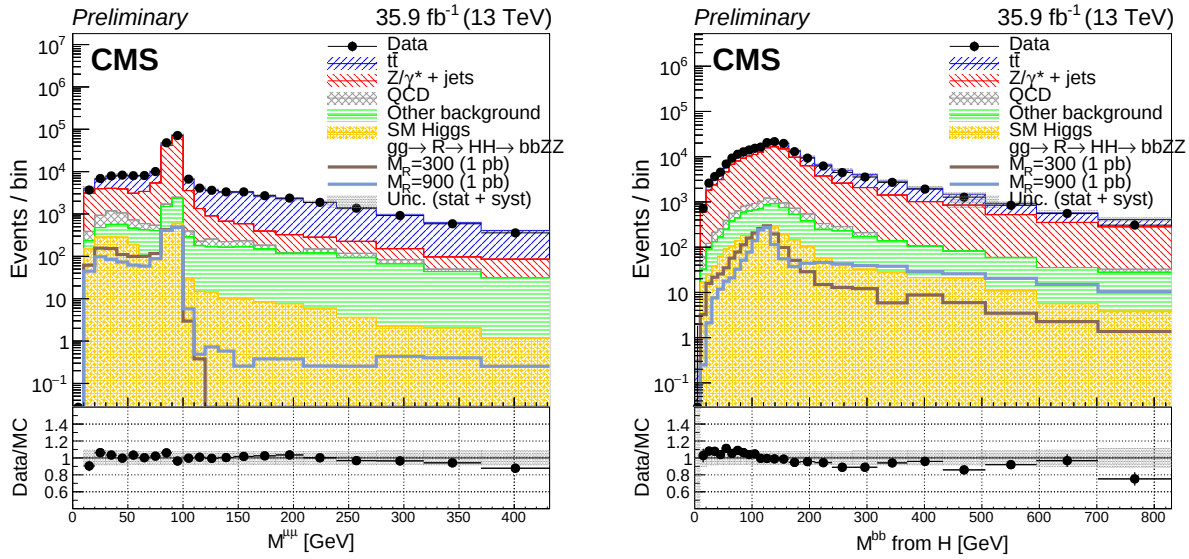


FIGURE B.2: Comparison of the $M_{\mu,\mu}$ and M_{bb}^{H} at preselection level in the muon channel.

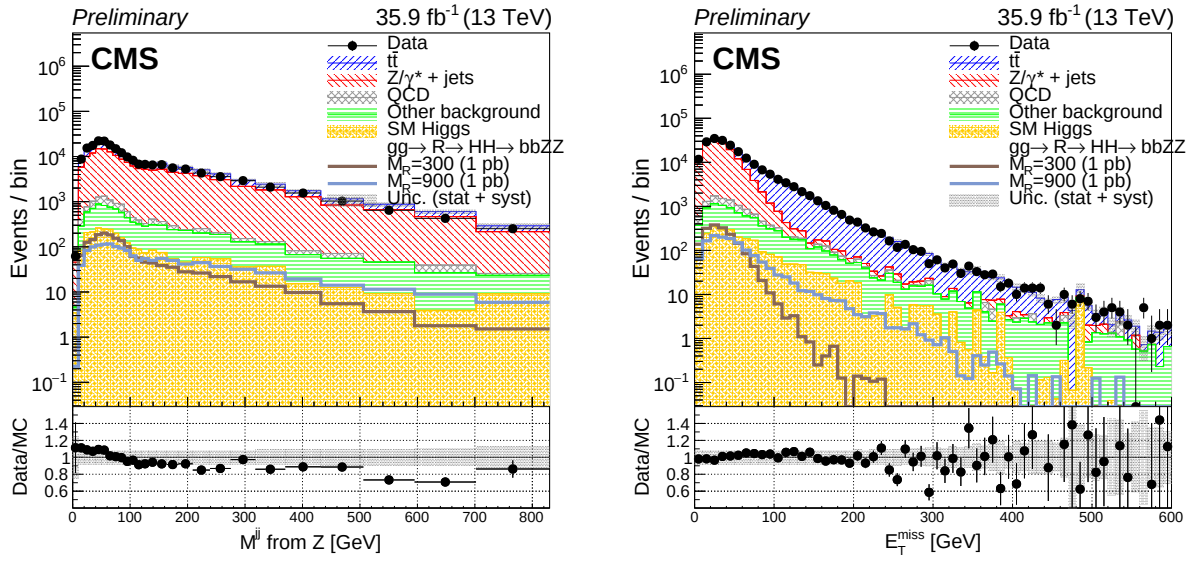


FIGURE B.3: Comparison of the M_{jj}^{Z} and $E_{\mathrm{T}}^{\mathrm{miss}}$ at preselection level in the muon channel.

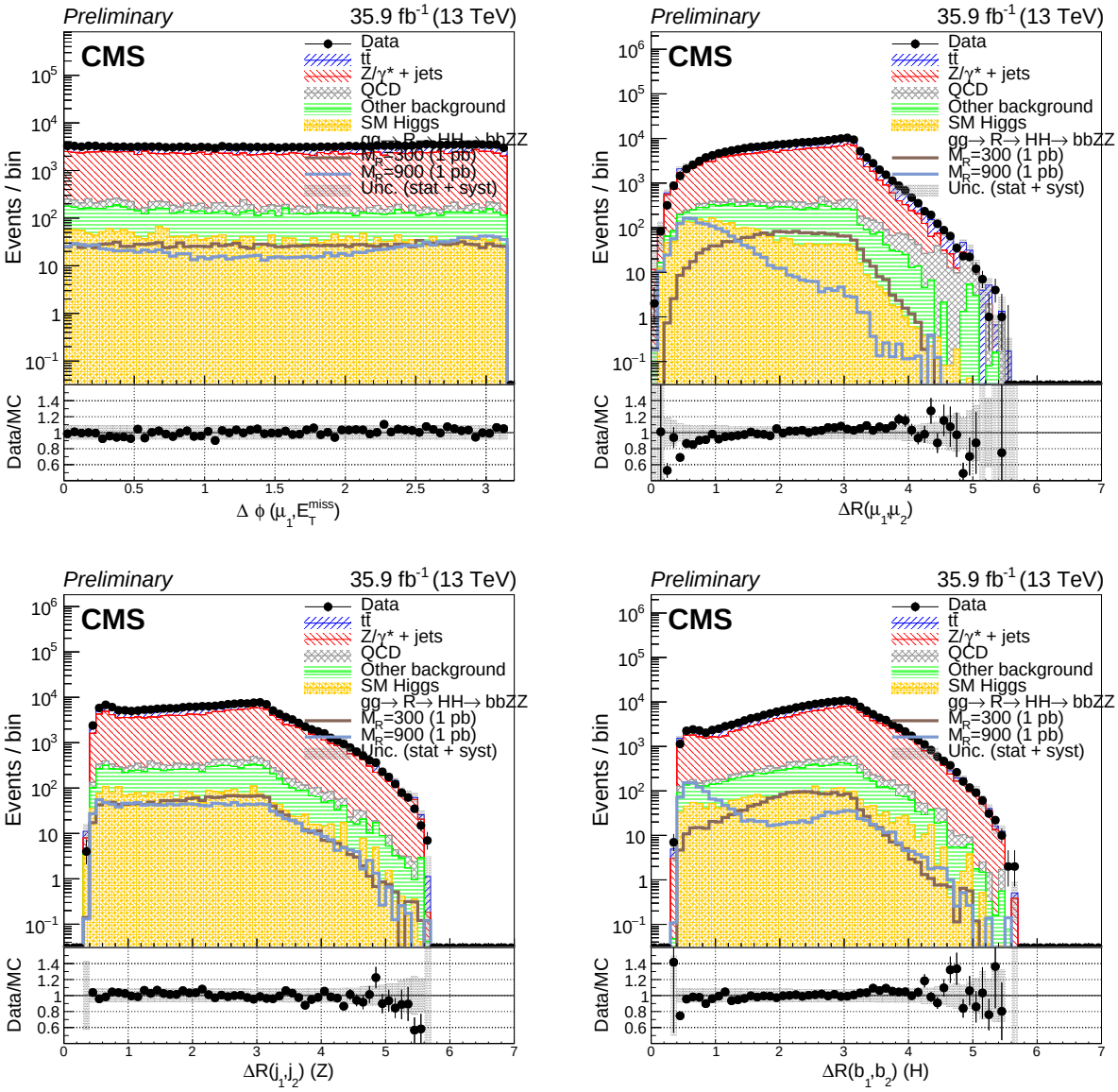


FIGURE B.4: Comparison of $\Delta\Phi_{\mu 1,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}}$, $\Delta R_{\mu\mu}$, ΔR^{Z}_{jj}, and ΔR^{H}_{bb} at preselection level in the muon channel.

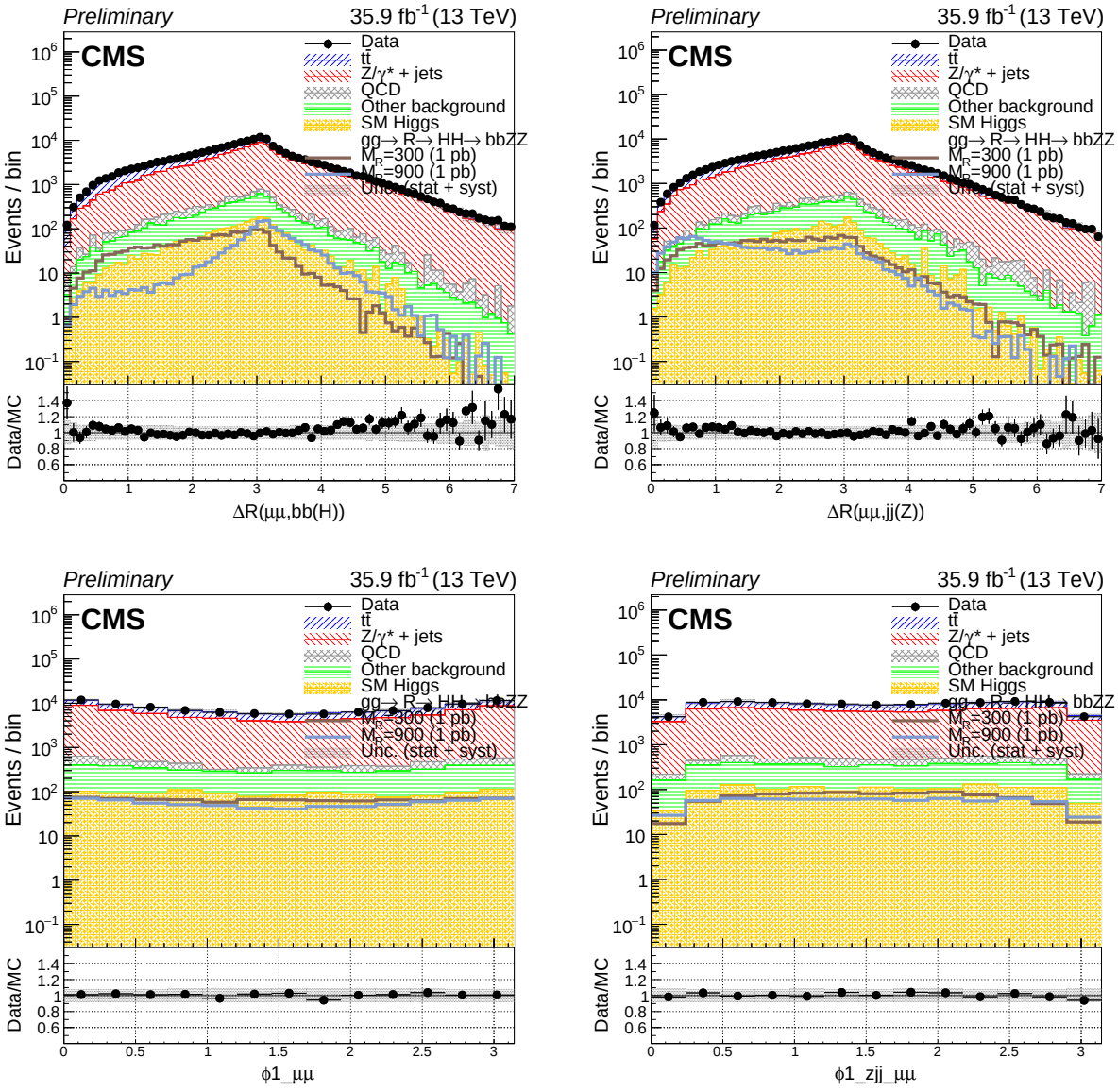


FIGURE B.5: Comparison of $\Delta R_{\mu\mu,bb^H}$, $\Delta R_{\mu\mu,jj^Z}$, $|\phi_1|$, and $|\phi_{1,Zjj}|$ at preselection level in the muon channel.

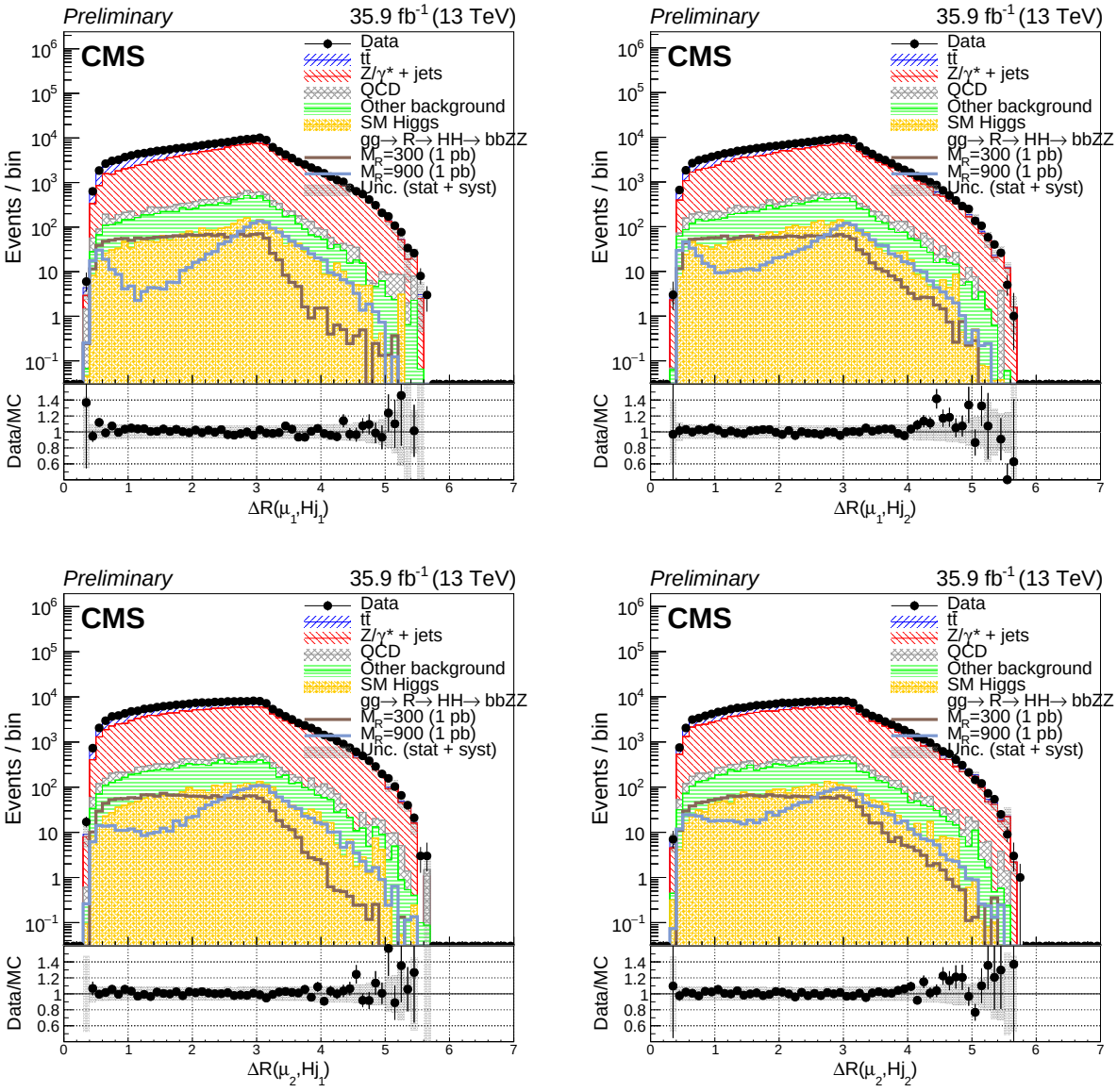


FIGURE B.6: Comparison of $\Delta R_{\mu 1 b 1}$, $\Delta R_{\mu 1 b 2}$, $\Delta R_{\mu 2 b 1}$, and $\Delta R_{\mu 2 b 2}$ at preselection level in the muon channel.

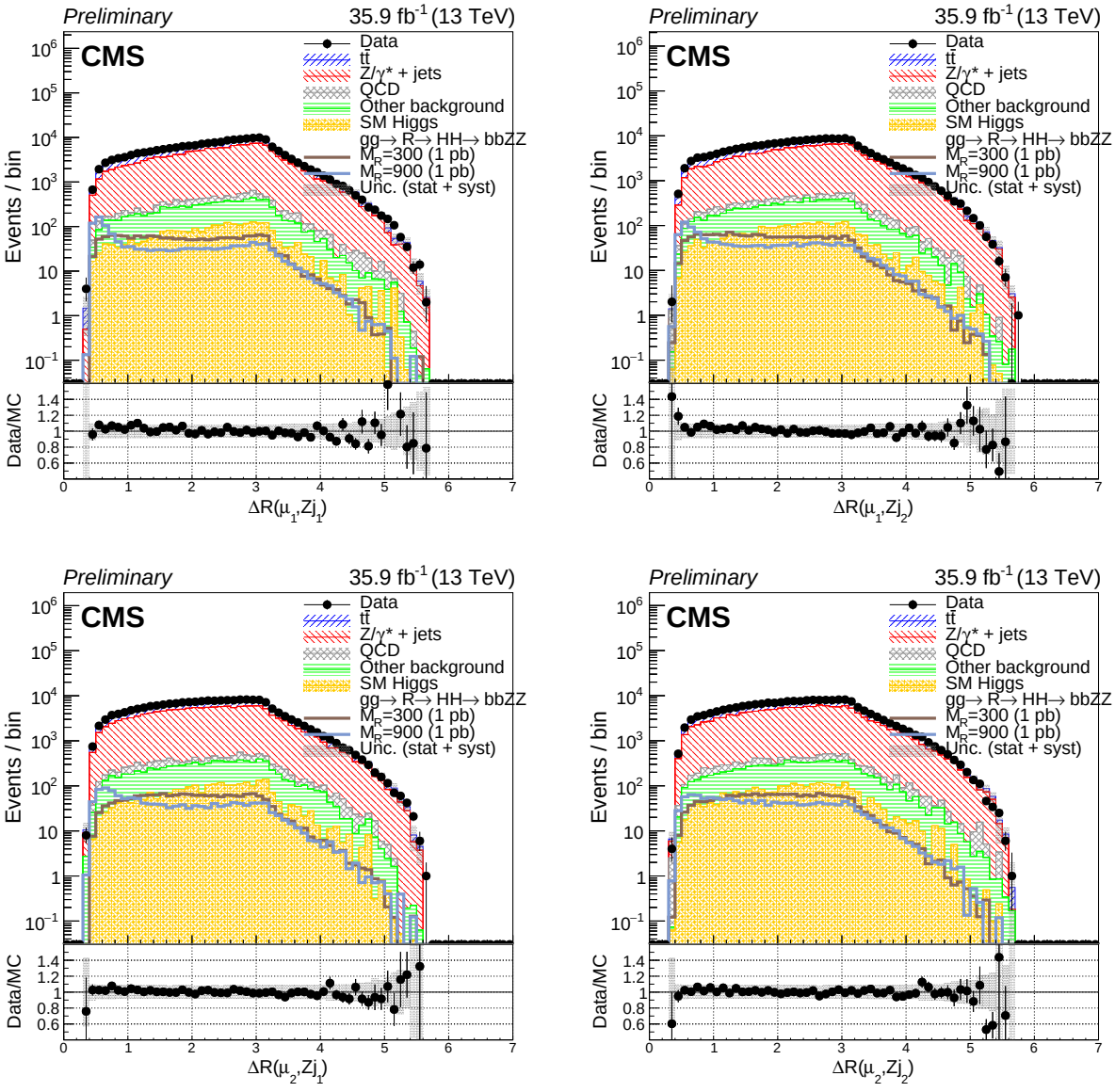


FIGURE B.7: Comparison of $\Delta R_{\mu 1 j 1}$, $\Delta R_{\mu 1 j 2}$, $\Delta R_{\mu 2 j 1}$, and $\Delta R_{\mu 2 j 1}$ at preselection level in the muon channel.

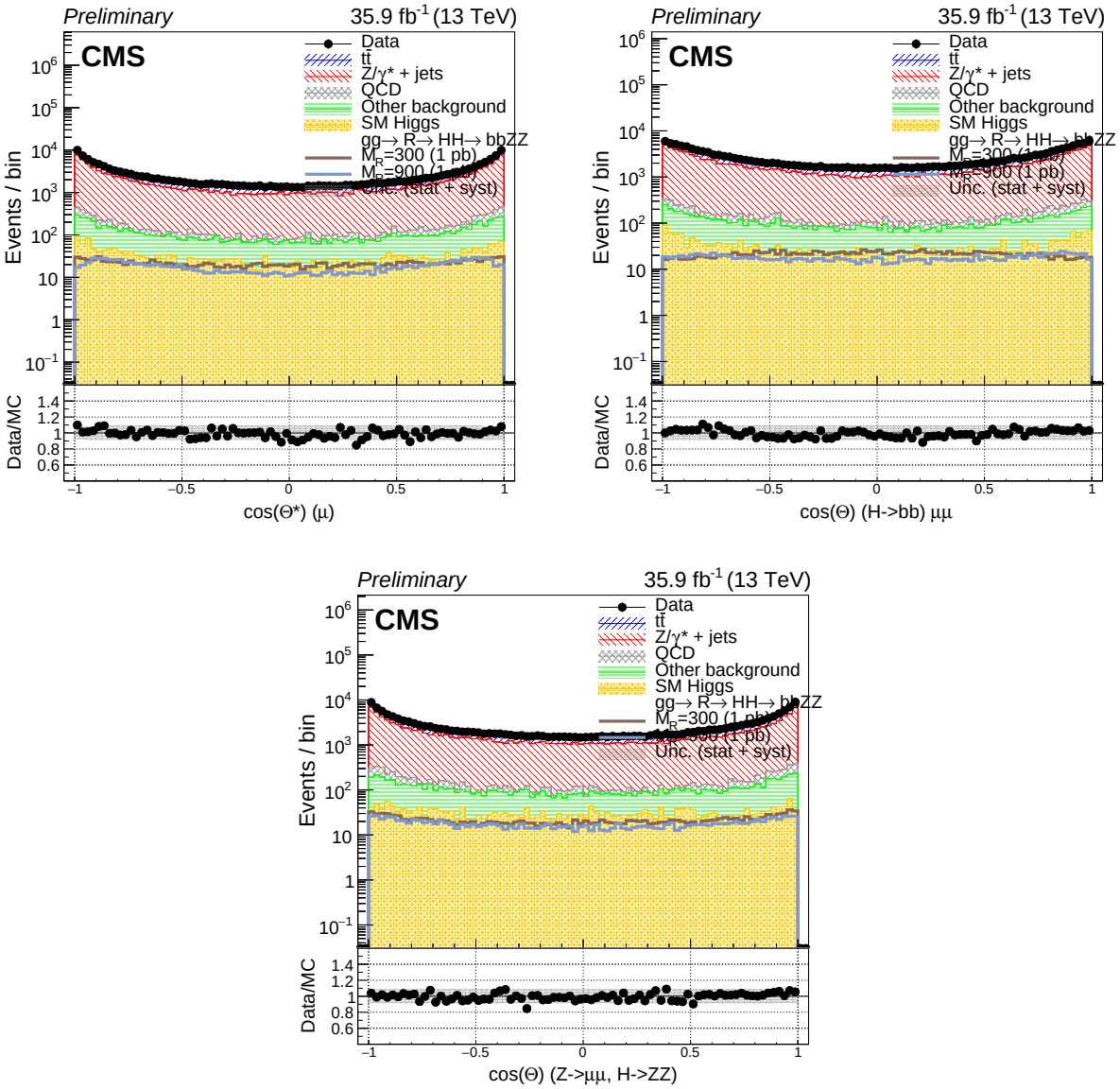


FIGURE B.8: Comparison of $|cos(\theta^*_{CS})|$, $|cos(\theta^*_{b,Hbb})|$ and $|cos(\theta^*_{Zuu,Hzz})|$ at preselection level in the muon channel.

B.2.2 Electron channel preselection

Good agreement is found for all final state distributions at preselection level in the electron channel (Figures B.9– B.15). For all plots, the Z/γ^*+jets and $t\bar{t}$ MC predictions are normalized to data as described in Sec. 5.4.

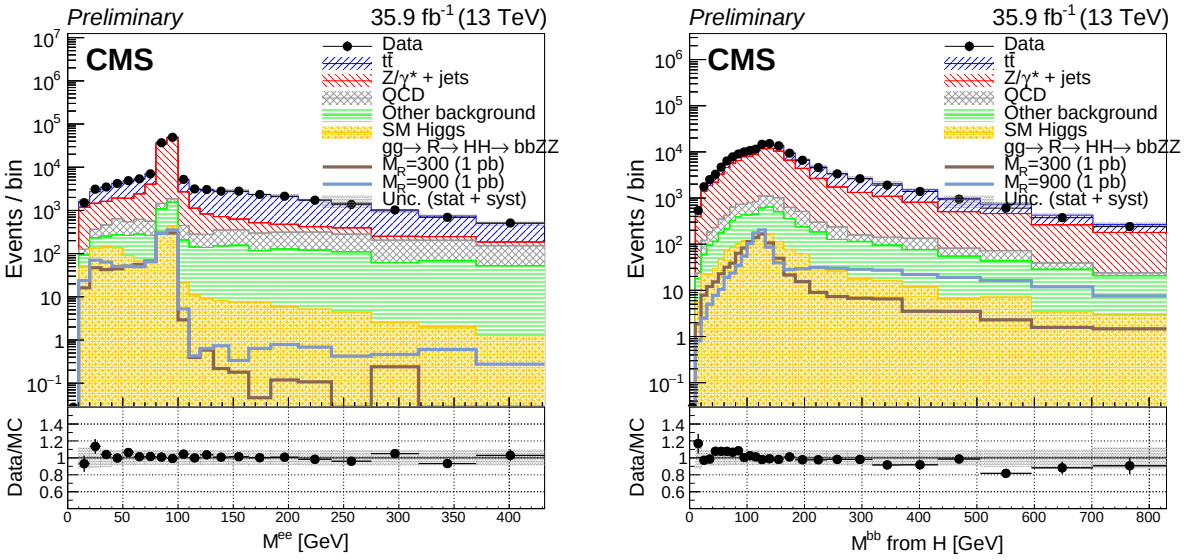


FIGURE B.9: Comparison of the $M_{e,e}$ and M_{bb}^{H} at preselection level in the electron channel.

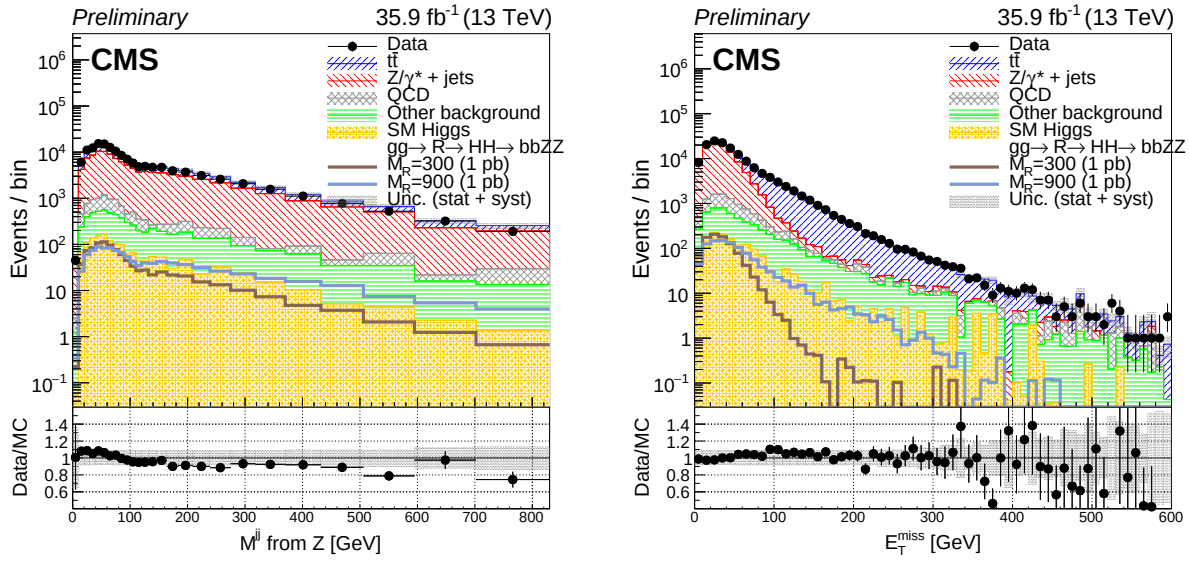


FIGURE B.10: Comparison of the M_{jj}^{Z} and $E_{\mathrm{T}}^{\mathrm{miss}}$ at preselection level in the electron channel.

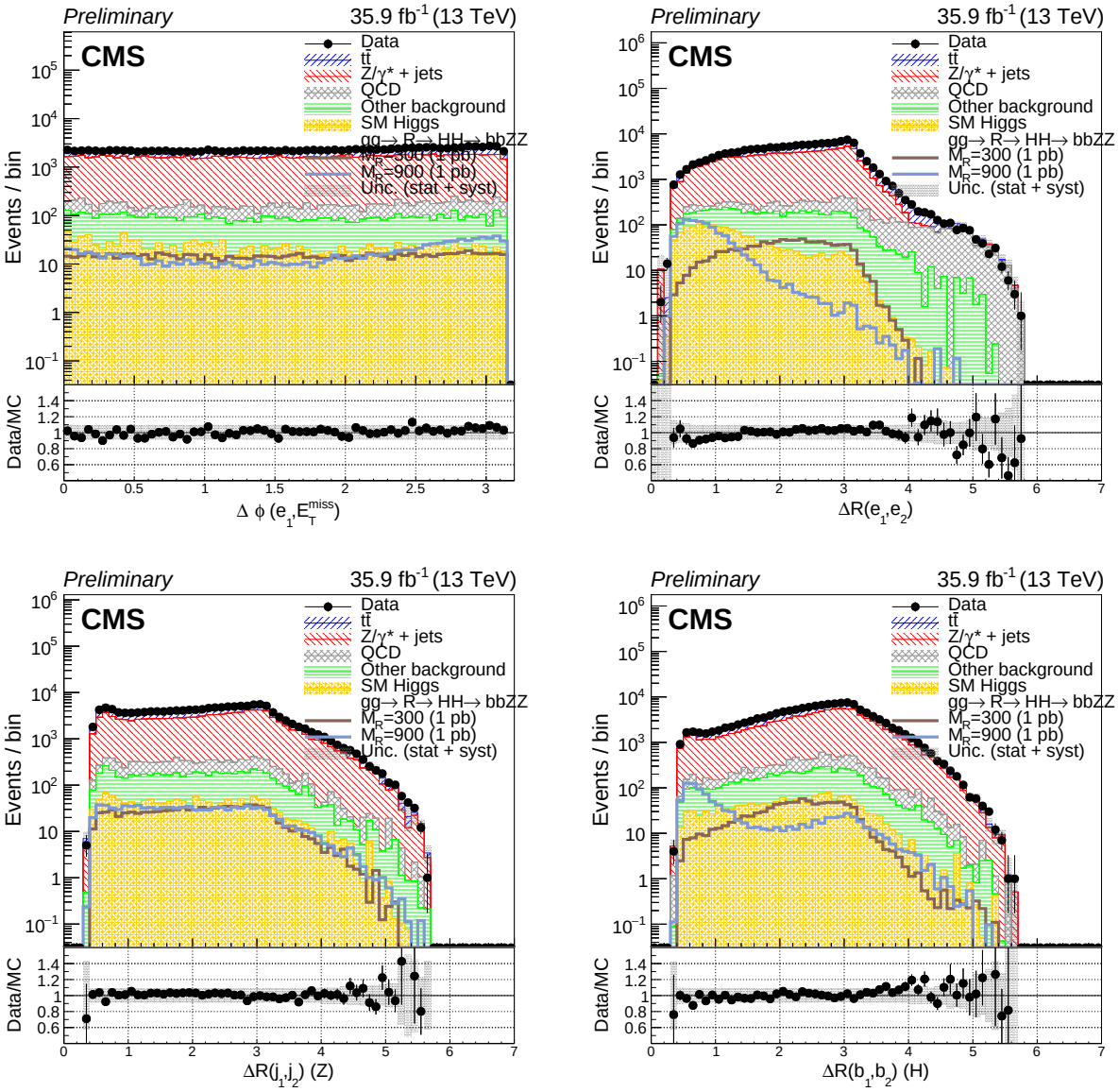


FIGURE B.11: Comparison of $\Delta\Phi_{e1,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}}$, ΔR_{ee}, ΔR_{jj}^{Z}, and ΔR_{bb}^{H} at preselection level in the electron channel.

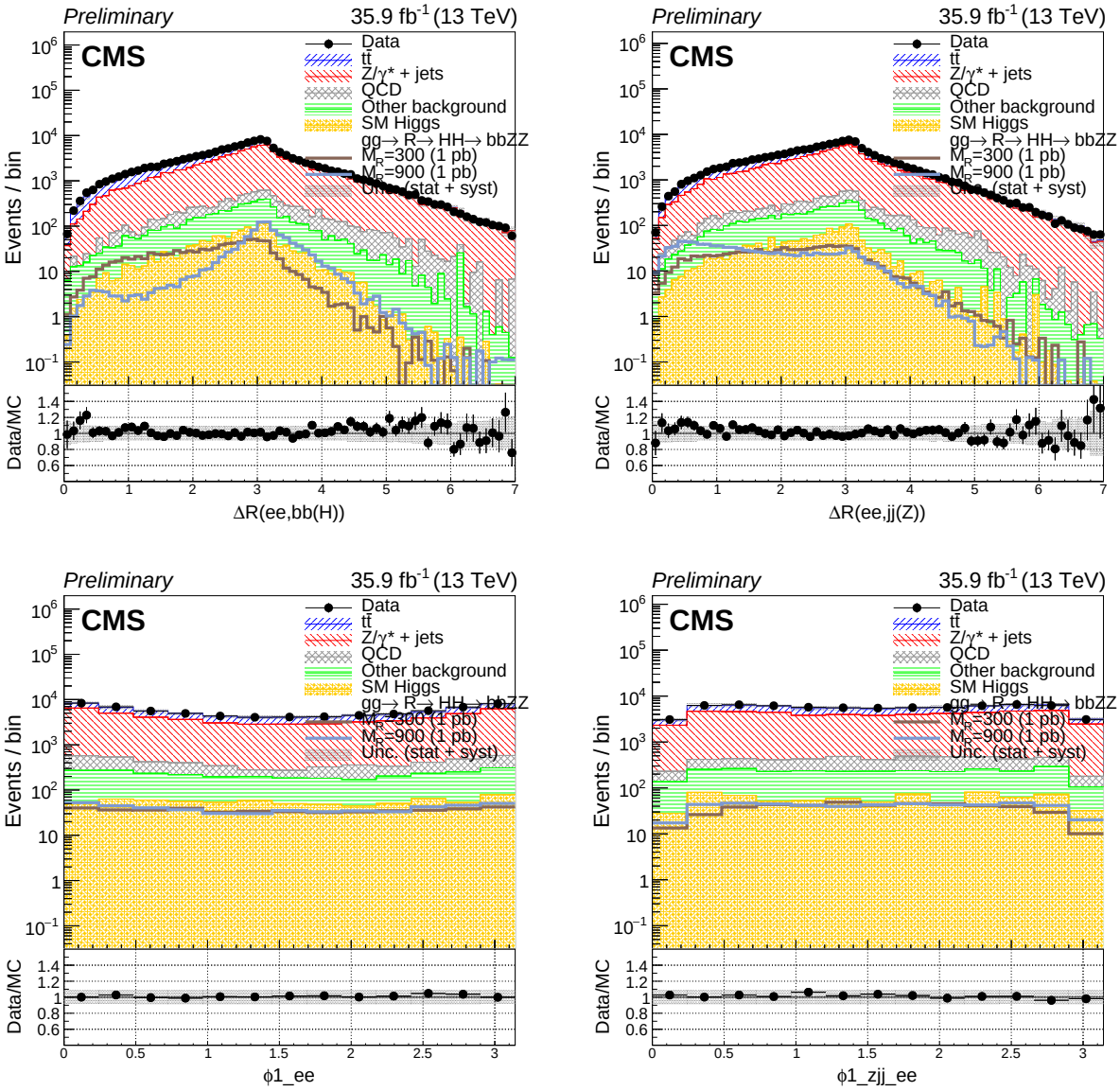


FIGURE B.12: Comparison of $\Delta R_{ee,bb^H}$, $\Delta R_{ee,jj^Z}$, $|\phi_1|$, and $|\phi_{1,Zjj}|$ at preselection level in the electron channel.

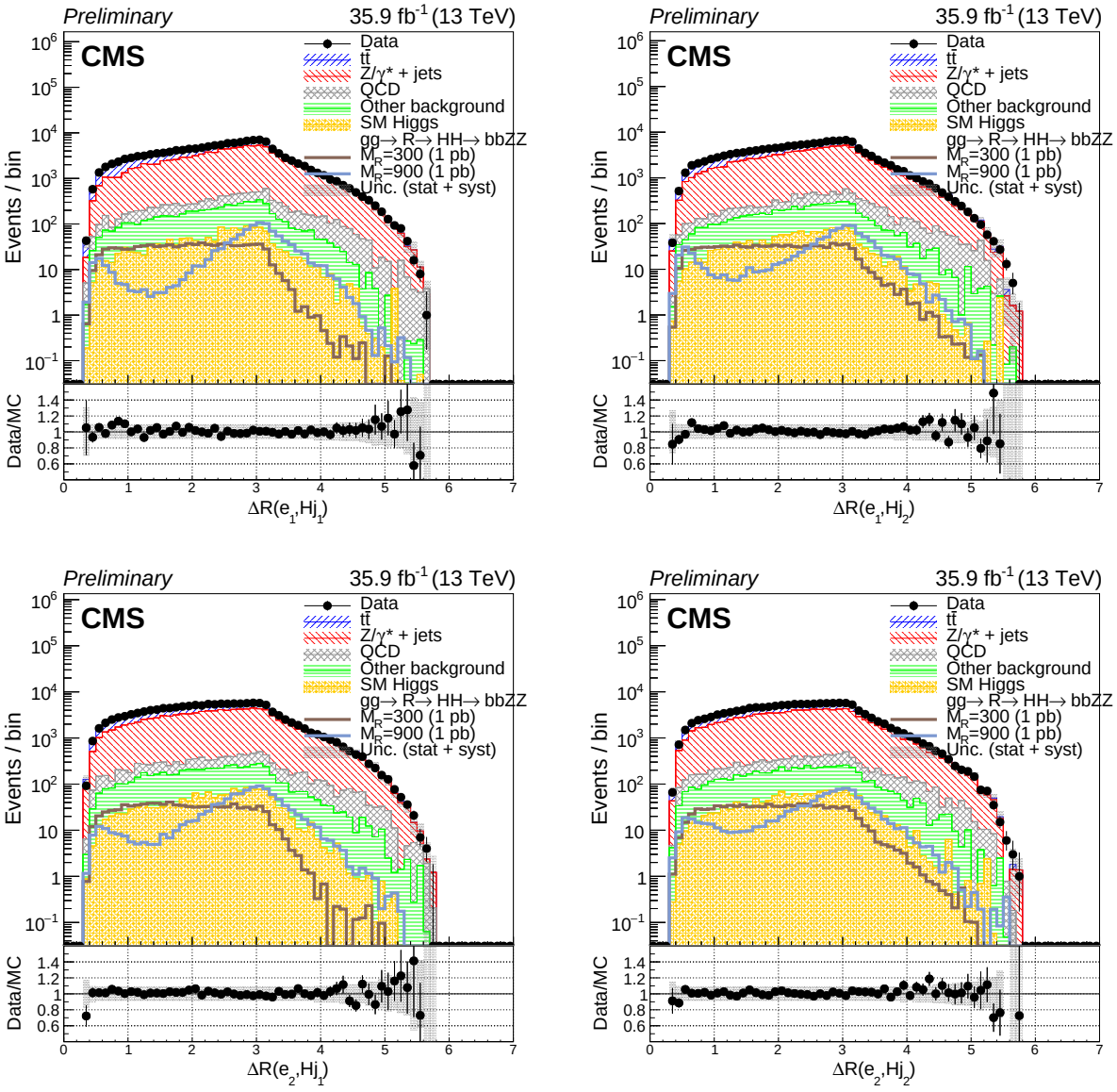


FIGURE B.13: Comparison of ΔR_{e1b1}, ΔR_{e1b2}, ΔR_{e2b1}, and ΔR_{e2b2} at preselection level in the electron channel.

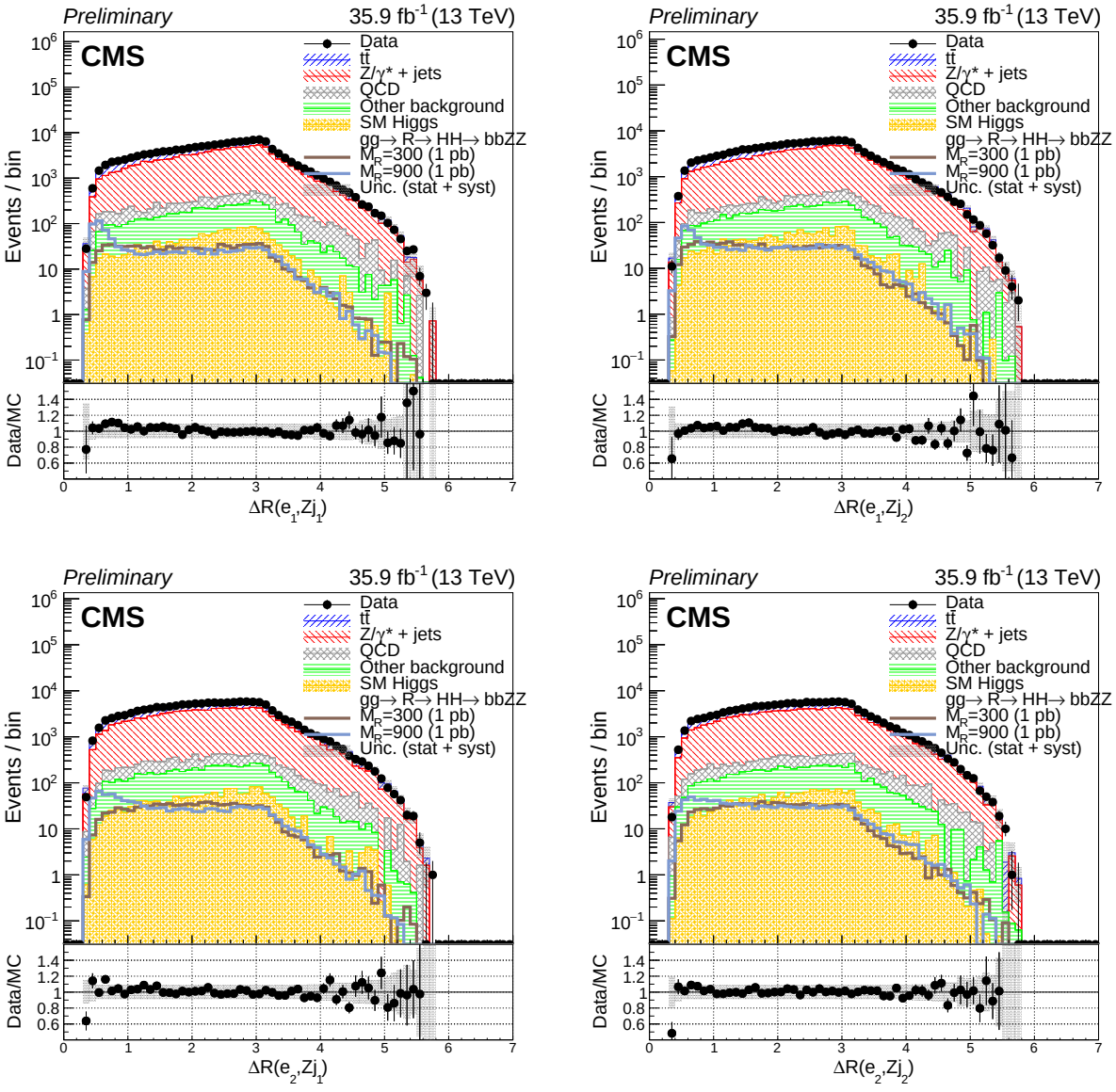


FIGURE B.14: Comparison of ΔR_{e1j1}, ΔR_{e1j2}, ΔR_{e2j1}, and ΔR_{e2j1} at preselection level in the electron channel.

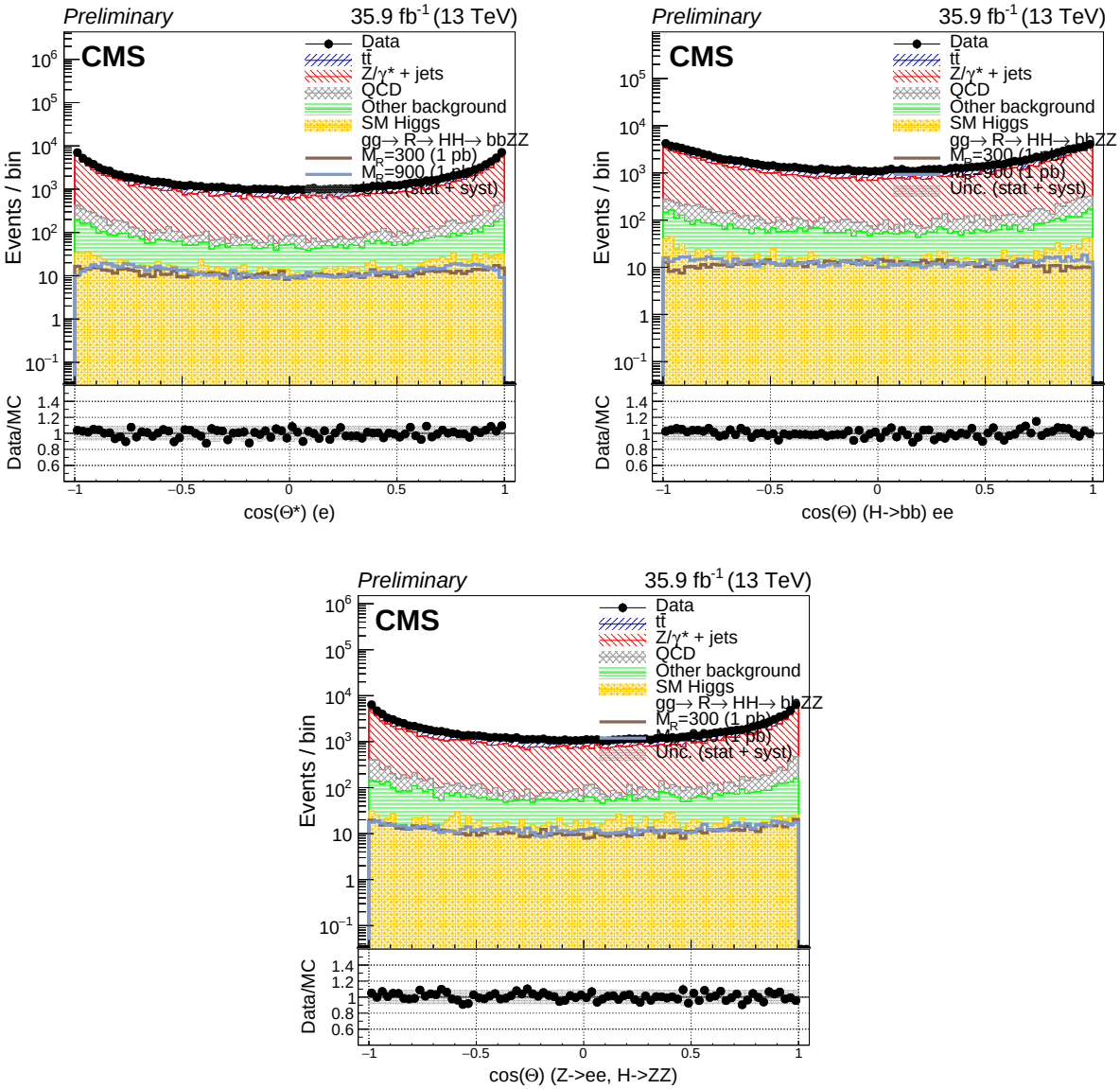


FIGURE B.15: Comparison of $|cos(\theta^*_{CS})|$, $|cos(\theta^*_{b,Hbb})|$ and $|cos(\theta^*_{Zee,Hzz})|$ at preselection level in the electron channel.

B.3 BDT Training and Optimization

The BDTs are trained with the multivariate analysis (MVA) tool, TMVA, implemented in ROOT. The algorithm used is the "Gradient Boosted" BDT. Events in each sample are split equally into the training and test samples by randomly picking events.

A set of 50 kinematic variables were initially selected based on their possibility to have high separation power between signal and background. Generally, the higher performance of the BDTs based on the integral of the ROC (Receiver Operating Characteristic) curve is obtained when using more input variables in the training. However, only the 22 most effective variables listed in Sec. 5.5 are selected to be implemented, while including the others were proved to have no significant benefit. The variables that are not well modeled by the MC, considering the agreement with data, are not included in the training.

Fig. B.16 and B.18 shows the linear correlation matrix between these variables for spin-0 signal ($m_X = 1000\,\text{GeV}$) and background.

In order to get the highest sensitivity, a BDT for each signal mass point is trained with the corresponding signal sample merged with its two neighboring mass point samples. For the lowest mass point (260 GeV), since there is no lower neighboring mass point available, the BDT that was trained for $m_X = 270\,\text{GeV}$ is used. The same reasoning applies to the highest mass point (1000 GeV), where the BDT that was trained for $m_X = 900\,\text{GeV}$ is used.

Overfitting in the model is investigated by plotting the BDT output distributions of the training sample and test sample overlaid, as shown in Fig. B.17 and B.19. The BDT output distributions at final selection level, comparing data and simulation are shown in Sec. B.4. No overfitting is observed.

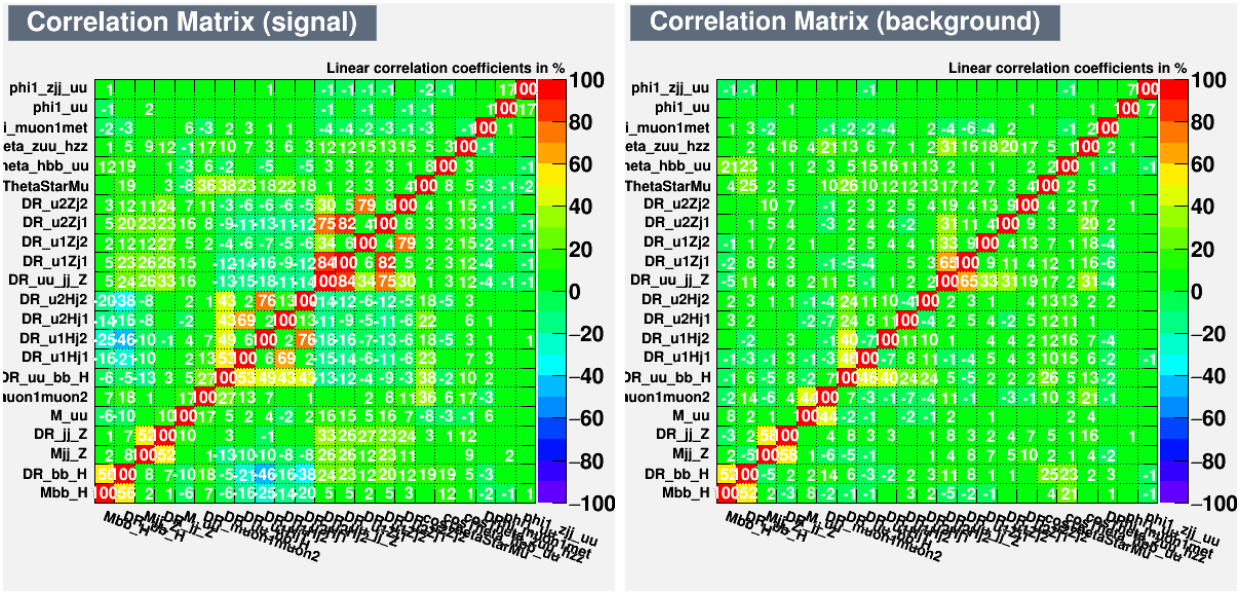


FIGURE B.16: Correlation matrix for spin-0 signal with $m_X = 1000\,\text{GeV}$ (left) and background (right) in the muon channel.

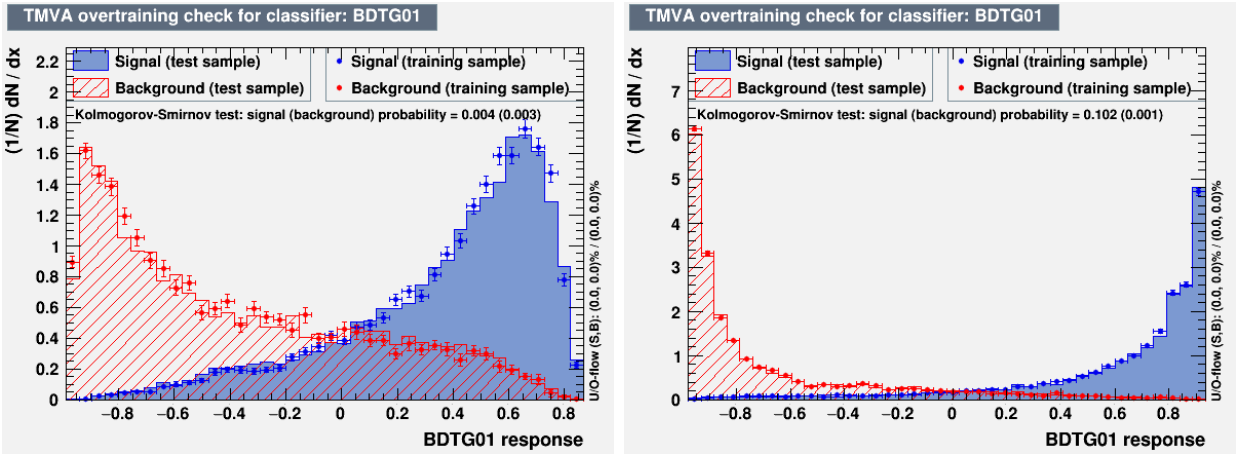


FIGURE B.17: The distribution of the BDT discriminator for spin-0 case with $m_X = 260\,\text{GeV}$ (left) and $m_X = 1000\,\text{GeV}$ (right) in the muon channel. The histograms shown in blue are signal and the ones in red are background. The filled histograms are from test sample, the dotted plots are from training sample.

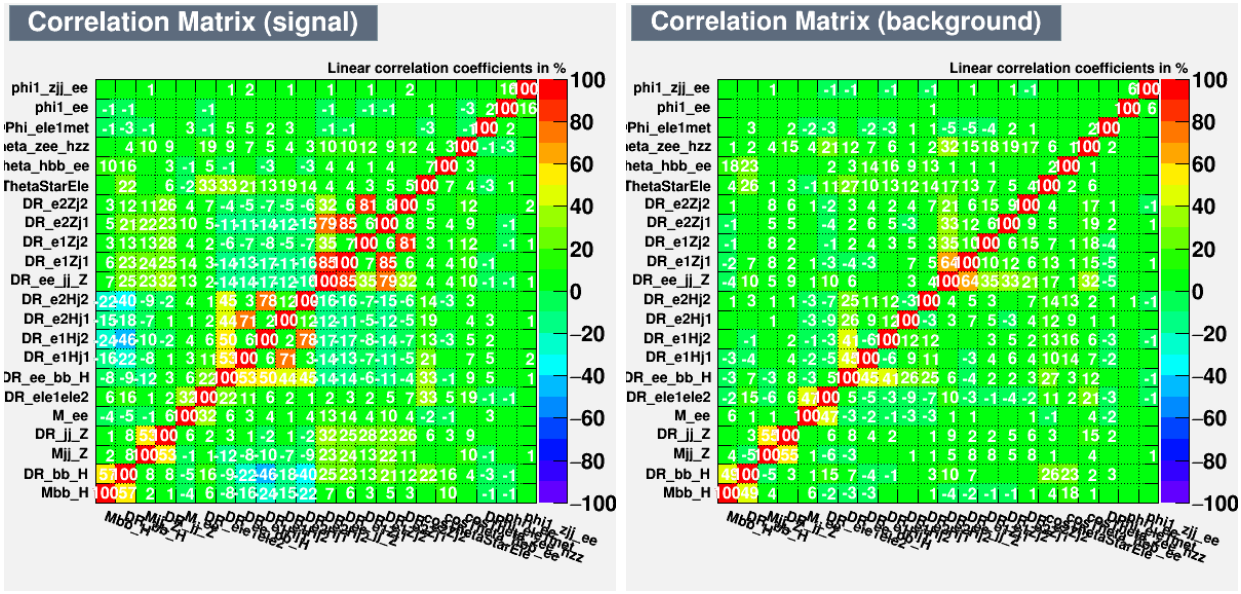


FIGURE B.18: Correlation matrix for spin-0 signal with $m_X = 1000\,\text{GeV}$ (left) and background (right) in the electron channel.

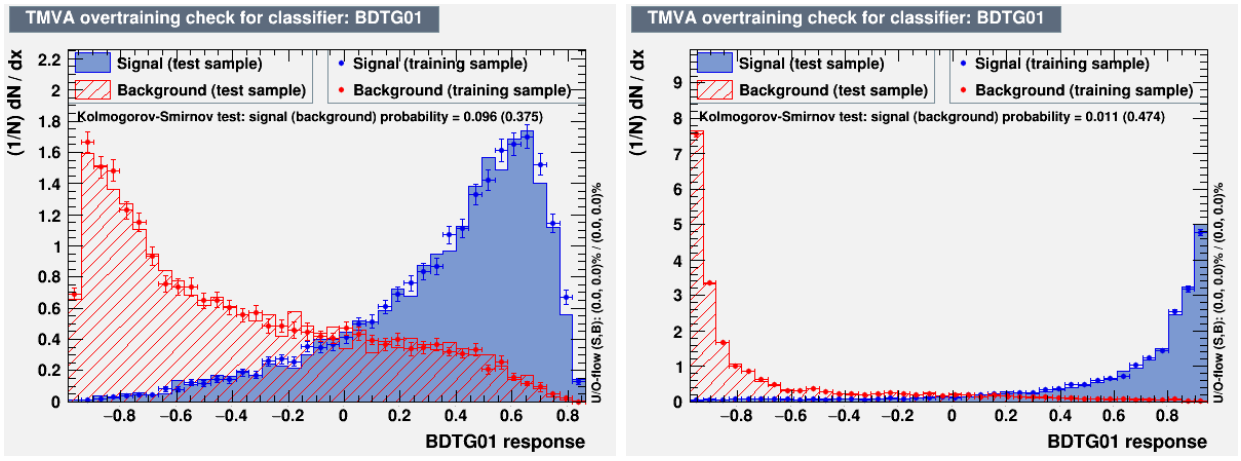


FIGURE B.19: The distribution of the BDT discriminator for spin-0 case with $m_X = 260\,\text{GeV}$ (left) and $m_X = 1000\,\text{GeV}$ (right) in the electron channel. The histograms shown in blue are signal and the ones in red are background. The filled histograms are from test sample, the dotted plots are from training sample.

B.4 BDT Distributions

B.4.1 BDT Distributions for Spin-0 Case

B.4.1.1 Muon Channel: BDT Distributions for Spin-0 Case

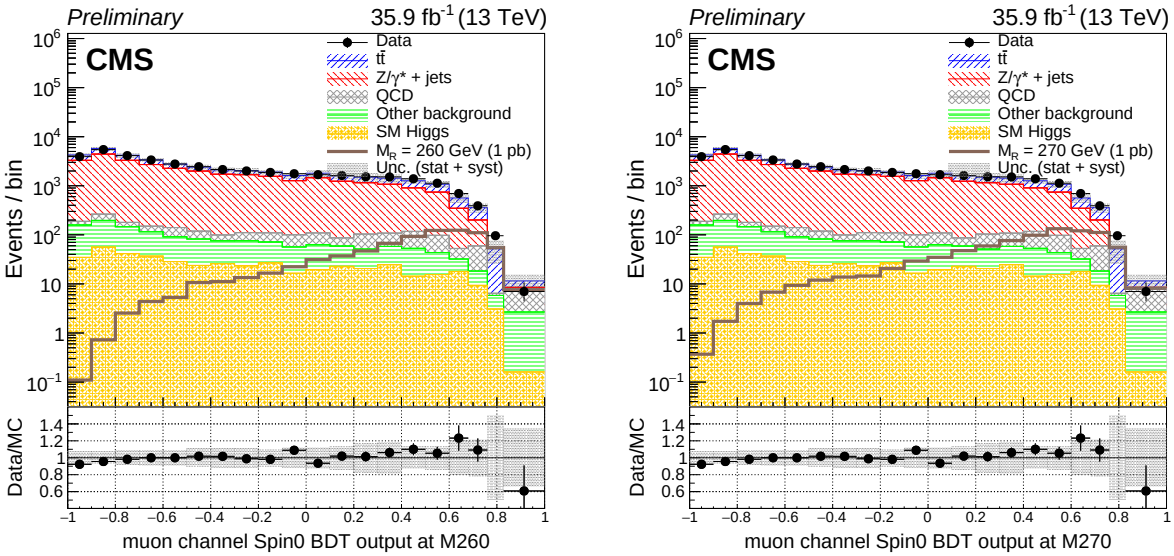


FIGURE B.20: Comparison of BDT discriminant for m_X=260 GeV and m_X=270 GeV at final selection level in the muon channel. The signals of radion with mass of 260 (left) and 270 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

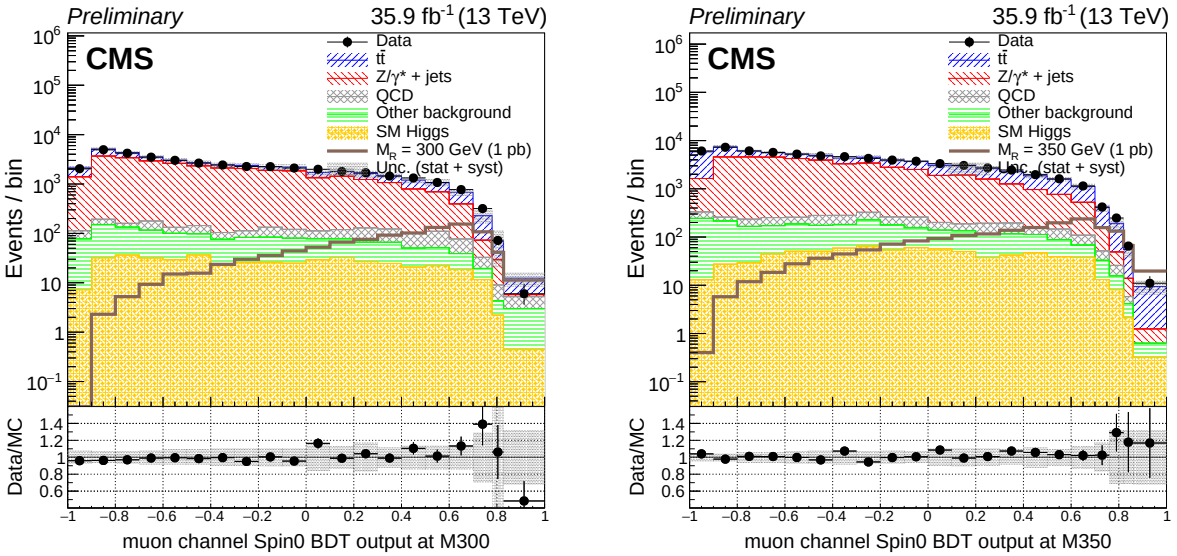


FIGURE B.21: Comparison of BDT discriminant for m_X=300 GeV and m_X=350 GeV at final selection level in the muon channel. The signals of radion with mass of 300 (left) and 350 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

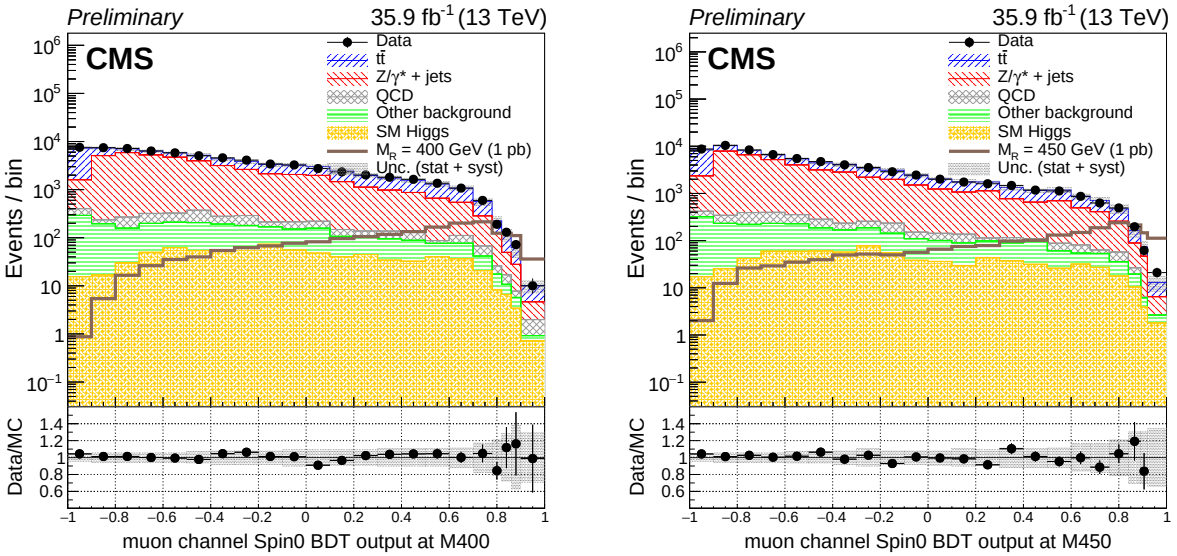


FIGURE B.22: Comparison of BDT discriminant for m_X=400 GeV and m_X=450 GeV at final selection level in the muon channel. The signals of radion with mass of 400 (left) and 450 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

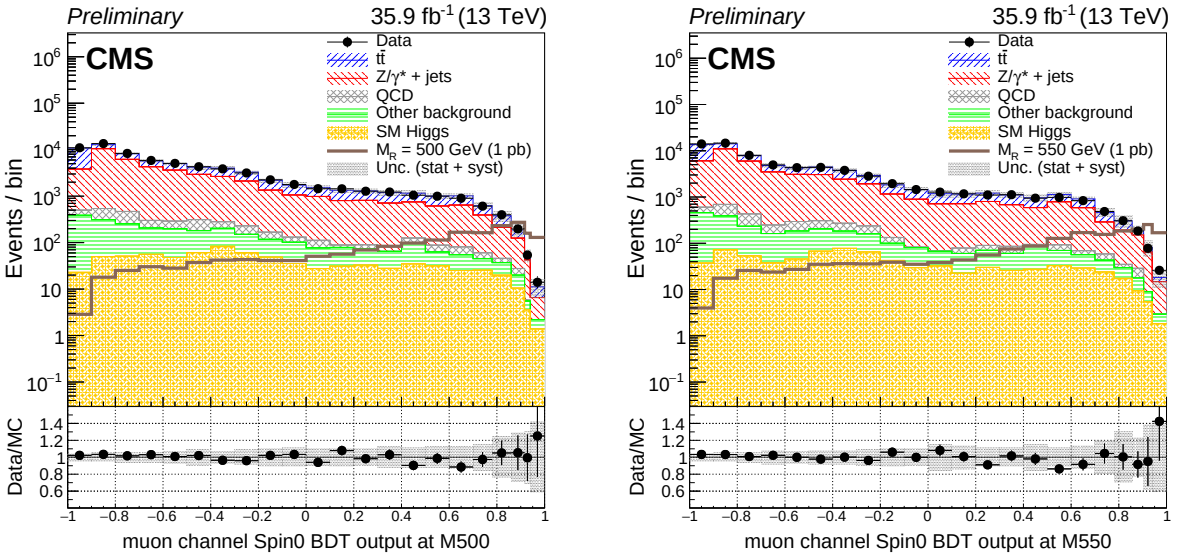


FIGURE B.23: Comparison of BDT discriminant for m_X=500 GeV and m_X=550 GeV at final selection level in the muon channel. The signals of radion with mass of 500 (left) and 550 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

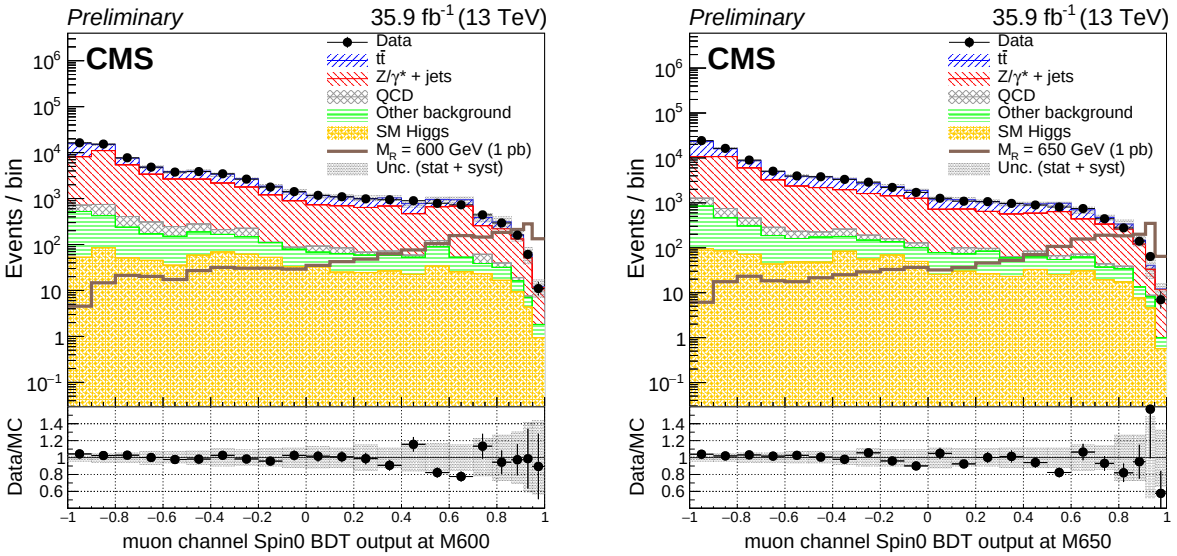


FIGURE B.24: Comparison of BDT discriminant for m_X=600 GeV and m_X=650 GeV at final selection level in the muon channel. The signals of radion with mass of 600 (left) and 650 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

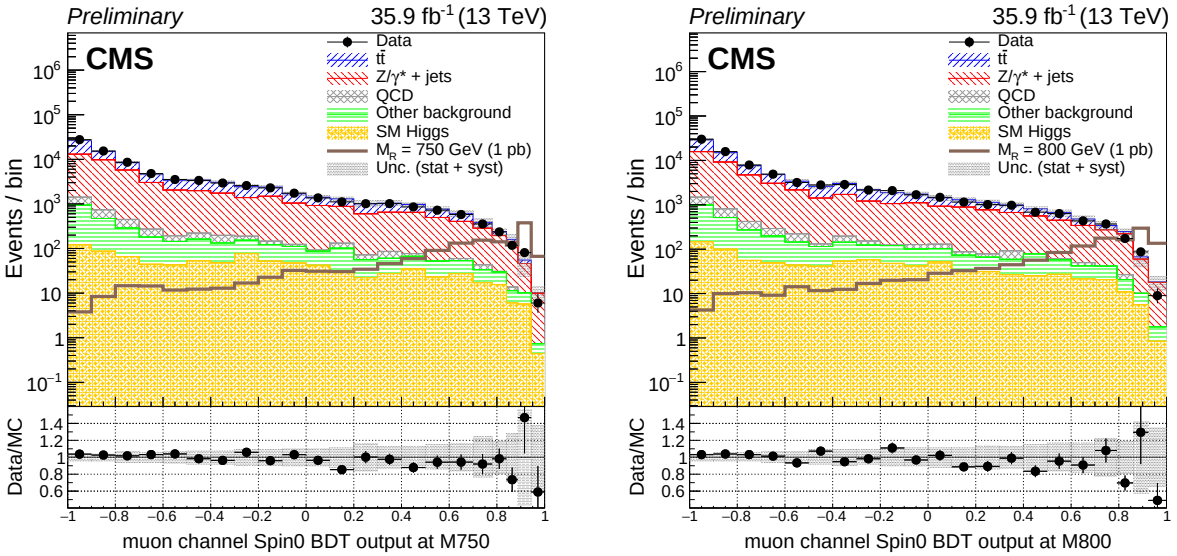


FIGURE B.25: Comparison of BDT discriminant for m_X=750 GeV and m_X=800 GeV at final selection level in the muon channel. The signals of radion with mass of 750 (left) and 800 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

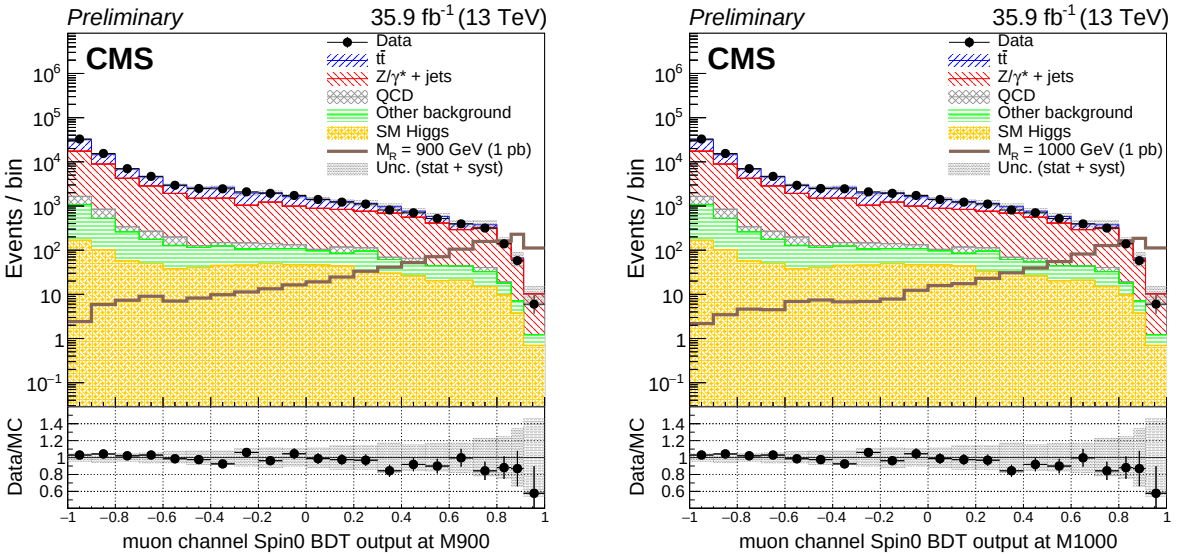


FIGURE B.26: Comparison of BDT discriminant for m_X=900 GeV and m_X=10000 GeV at final selection level in the muon channel. The signals of radion with mass of 900 (left) and 1000 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

B.4.1.2 Electron Channel: BDT Distributions for Spin-0 Case

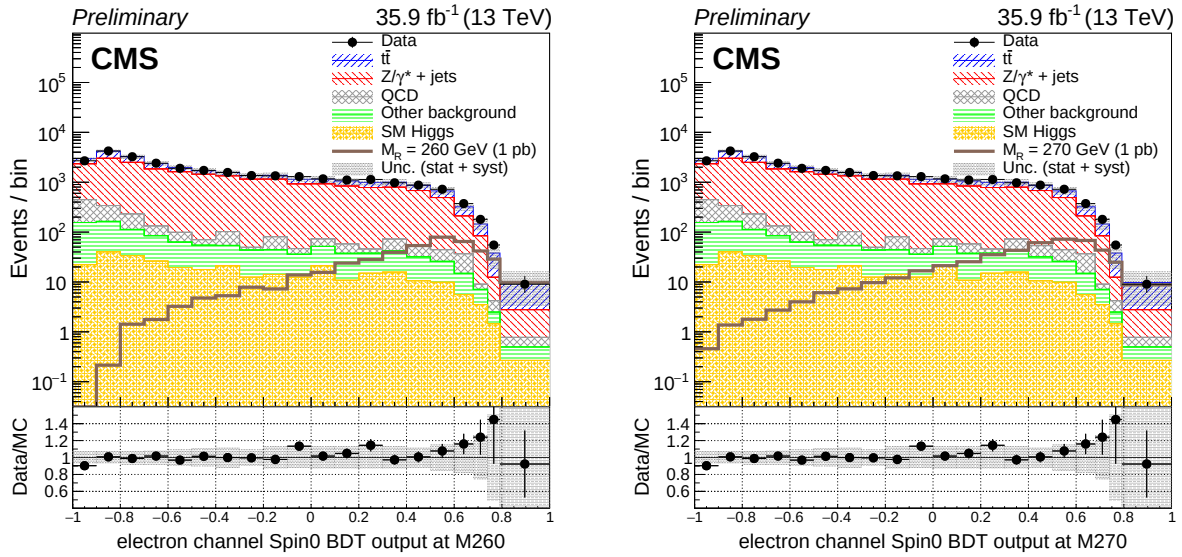


FIGURE B.27: Comparison of BDT discriminant for m_X=260 GeV and m_X=270 GeV at final selection level in the electron channel. The signals of radion with mass of 260 (left) and 270 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

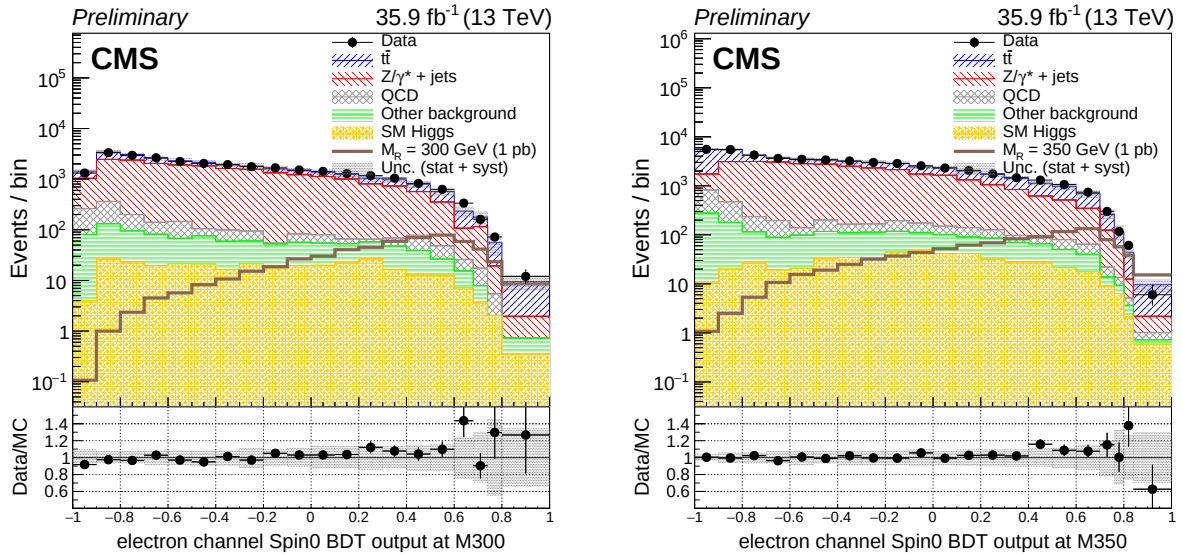


FIGURE B.28: Comparison of BDT discriminant for m_X=300 GeV and m_X=350 GeV at final selection level in the electron channel. The signals of radion with mass of 300 (left) and 350 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

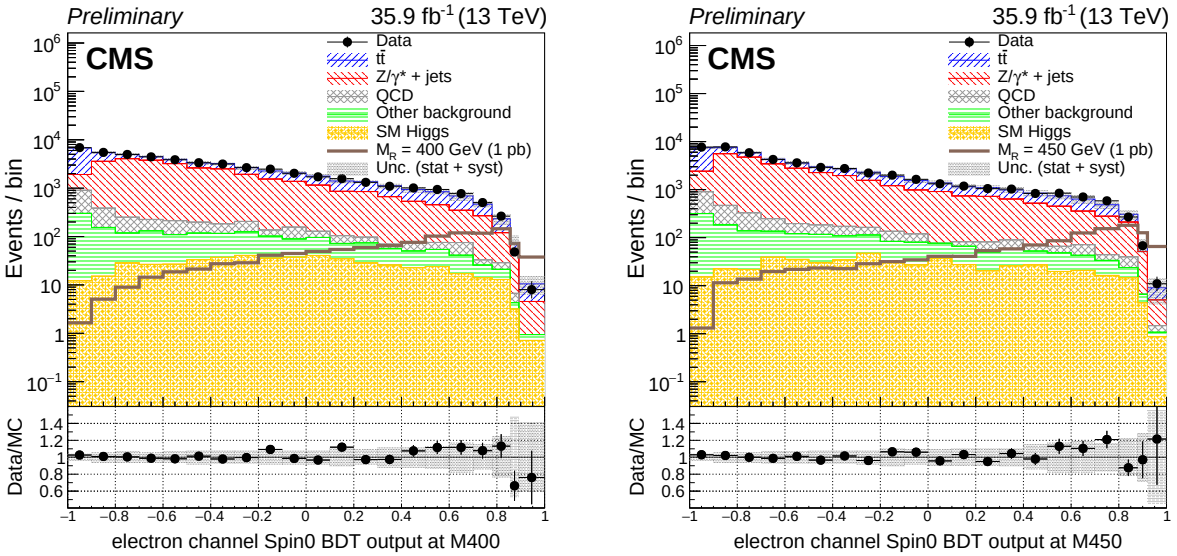


FIGURE B.29: Comparison of BDT discriminant for m_X=400 GeV and m_X=450 GeV at final selection level in the electron channel. The signals of radion with mass of 400 (left) and 450 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

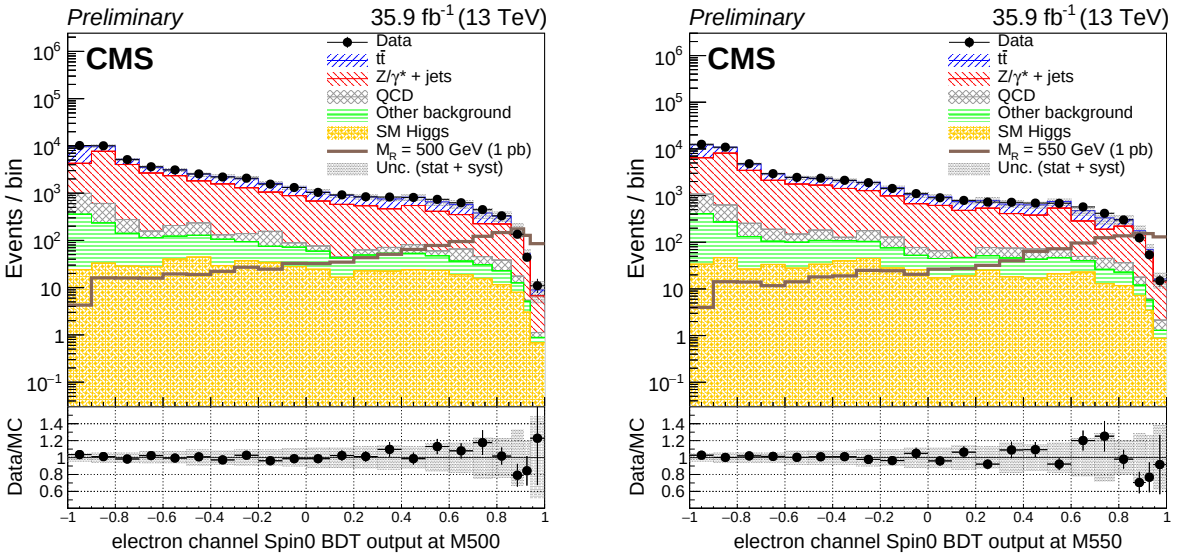


FIGURE B.30: Comparison of BDT discriminant for m_X=500 GeV and m_X=550 GeV at final selection level in the electron channel. The signals of radion with mass of 500 (left) and 550 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

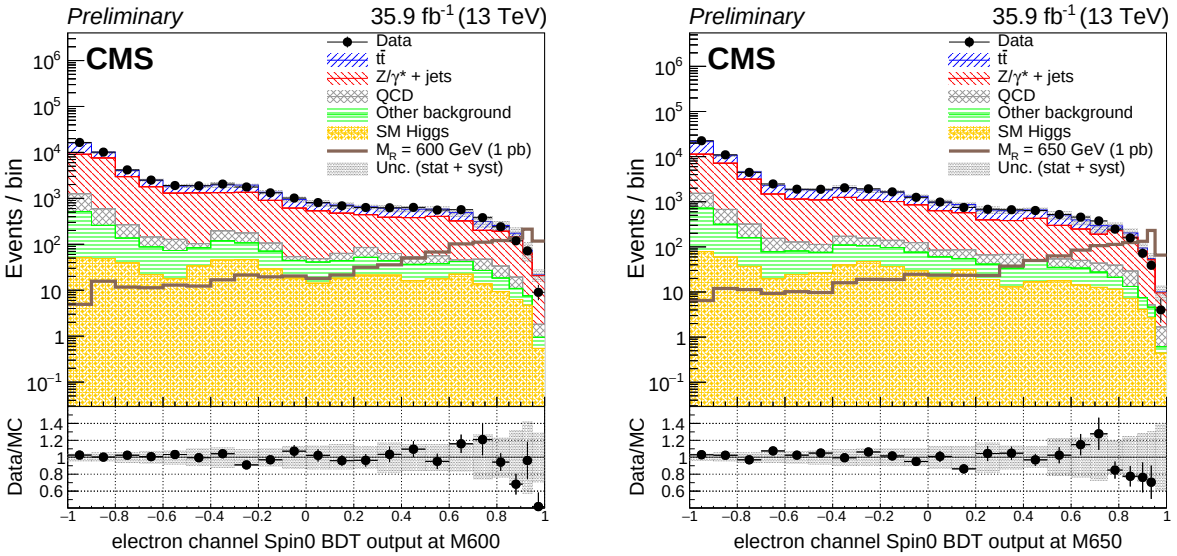


FIGURE B.31: Comparison of BDT discriminant for m_X=600 GeV and m_X=650 GeV at final selection level in the electron channel. The signals of radion with mass of 600 (left) and 650 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

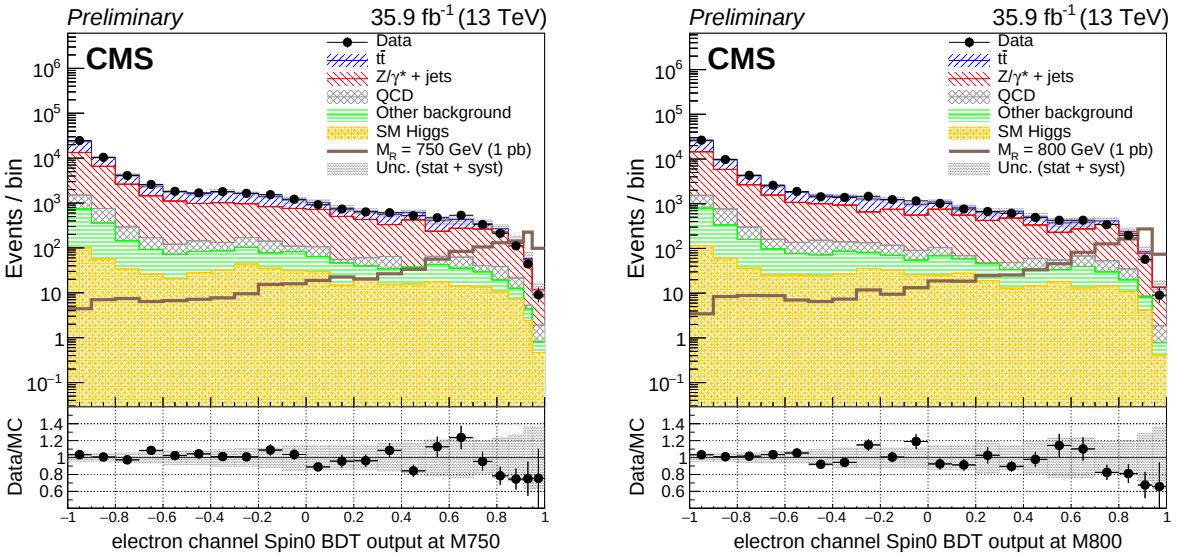


FIGURE B.32: Comparison of BDT discriminant for m_X=750 GeV and m_X=800 GeV at final selection level in the electron channel. The signals of radion with mass of 750 (left) and 800 GeV (right) are normalized to 1 pb for the HH→ $b\bar{b}ZZ \rightarrow b\bar{b}\ell\ell jj$ process.

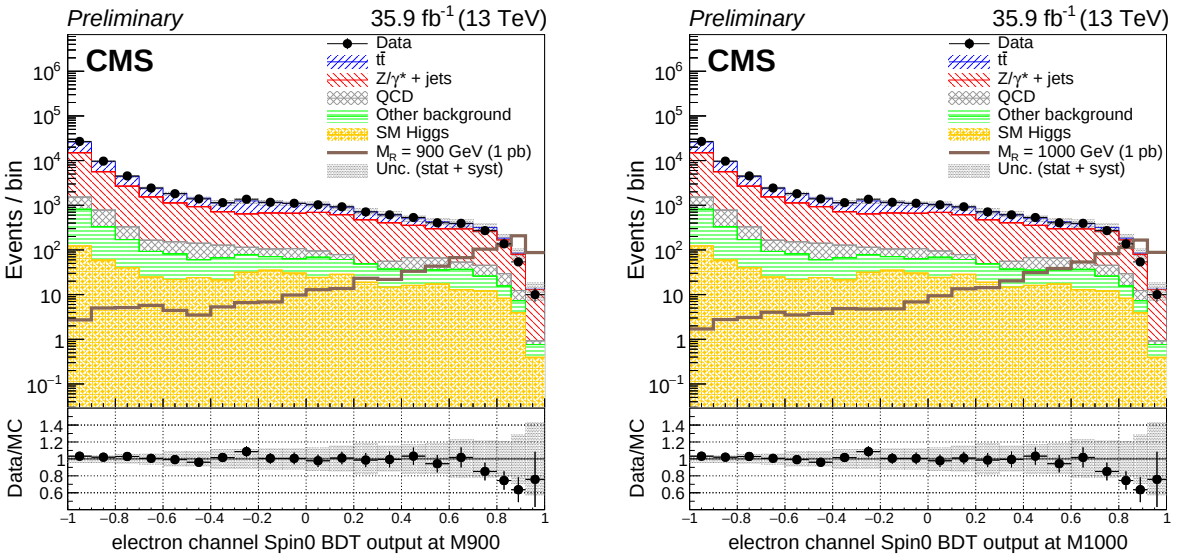


FIGURE B.33: Comparison of BDT discriminant for m_X=900 GeV and m_X=1000 GeV at final selection level in the electron channel. The signals of radion with mass of 900 (left) and 1000 GeV (right) are normalized to 1 pb for the HH$\rightarrow$ b$\bar{\text{b}}$ZZ $\rightarrow$ b$\bar{\text{b}}\ell\ell$jj process.

B.5 Upper Limits

The 95% CL upper limits on $\sigma(pp \to X \to HH)$ as a function of radion mass are shown in Fig. B.34, and as a function of graviton mass in Fig. B.35, together with the NLO predictions for the production cross section. The limits are derived in either the muon channel (left) or the electron channel (right).

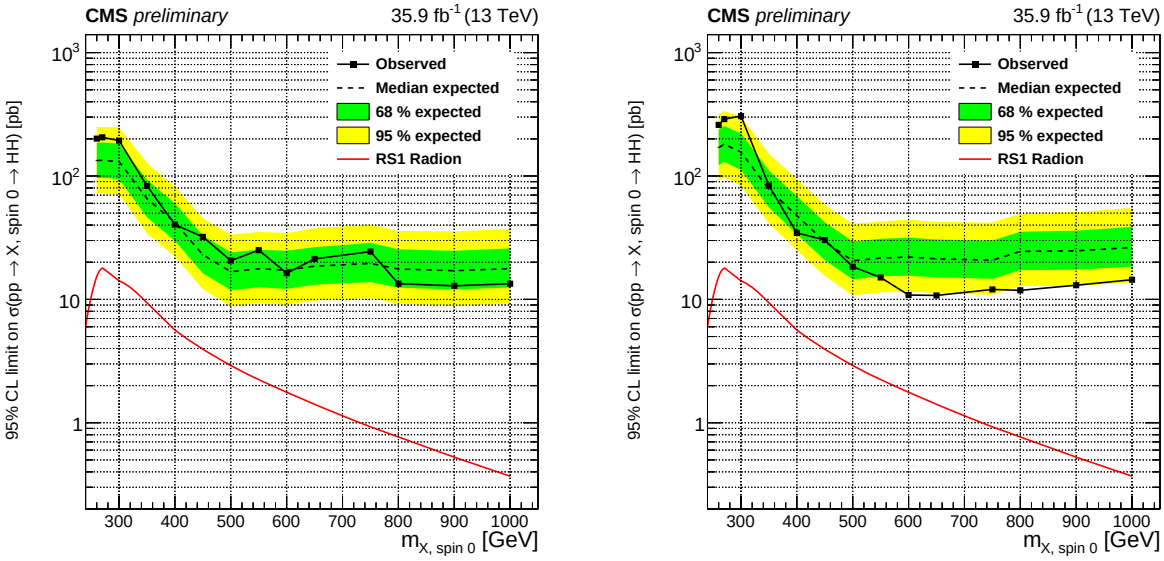


FIGURE B.34: Expected (black dashed line) and observed (black solid line) limits on the cross section of resonant HH production as a function of the mass of the resonance derived in the muon channel (left) and electron channel (right) for the RS1 radion case. The red solid lines show the theoretical prediction for the cross section of an RS1 radion with $\lambda_R = 1\,\text{TeV}$ and $kL = 35$.

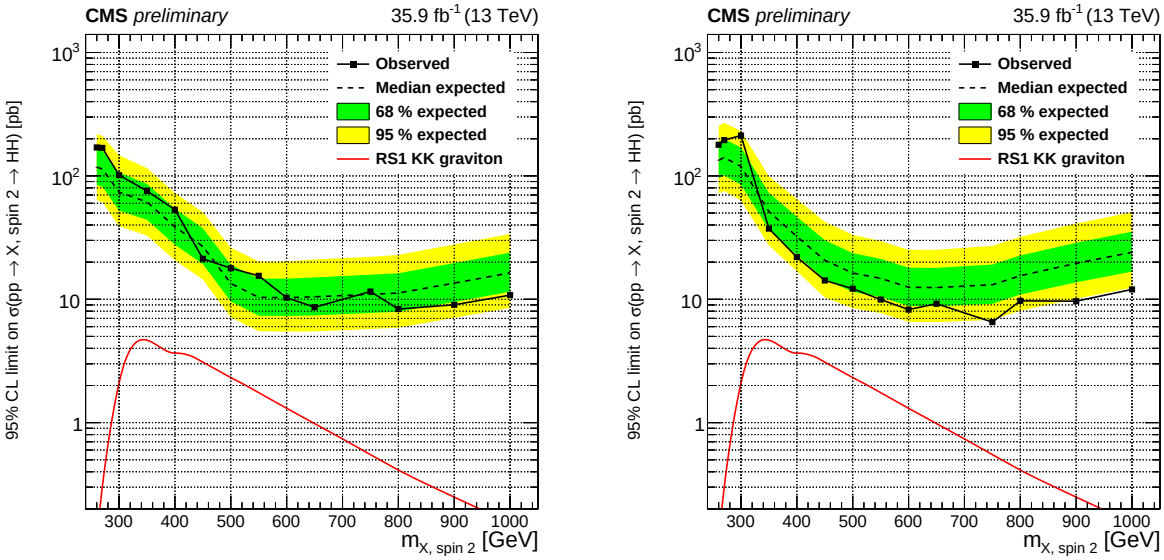


FIGURE B.35: Expected (black dashed line) and observed (black solid line) limits on the cross section of resonant HH production as a function of the mass of the resonance derived in the muon channel (left) and electron channel (right) for the RS1 KK graviton case. The red solid lines show the theoretical prediction for the cross section of an RS1 KK graviton with $\tilde{k} = 0.1$.