

**Jets, Substructure, and Searching for Dark Matter  
at the Large Hadron Collider**

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Author .....  
Department of Physics  
August 32, 2018

Certified by .....  
Christoph M. E. Paus  
Professor of Physics  
Thesis Supervisor

Accepted by .....  
Somebody  
Chairman, Department Committee on Graduate Theses



# **Jets, Substructure, and Searching for Dark Matter at the Large Hadron Collider**

by

Siddharth Madhavan Narayanan

Submitted to the Department of Physics  
on August 32, 2018, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## **Abstract**

Astrophysical observations of gravitational interactions provide strong evidence for the existence of dark matter (DM). Many theories propose and experiments test the hypothesis that DM may have a particle physics origin, but this remains unproven. One such experiment is the Compact Muon Solenoid (CMS) at the Large Hadron Collider (LHC). If DM couples to particles present in protons, it is possible that DM is produced in collisions at the LHC. Because DM is effectively invisible to CMS, we must look for collisions in which DM is produced in association with one or more Standard Model (SM) particles. This thesis describes two different scenarios for the SM particle hypothesis: a single top quark or two light quarks. Both cases result in complicated detector signatures due to the hadronization of final-state quarks. Improved jet substructure techniques to identify some of these unique signatures are presented. Since the observed data is consistent with SM backgrounds in all searches, we translate this result into the most stringent constraints to date on the relevant beyond-SM models.

Thesis Supervisor: Christoph M. E. Paus

Title: Professor of Physics



# Contents

<b>1</b>	<b>The CMS experiment at the LHC</b>	<b>7</b>
1.1	The Large Hadron Collider . . . . .	7
1.2	The Compact Muon Solenoid . . . . .	10
1.2.1	Silicon tracker . . . . .	11
1.2.2	Electromagnetic calorimeter . . . . .	16
1.2.3	Hadronic calorimeter . . . . .	18
1.2.4	Muon chambers . . . . .	20
1.2.5	Online trigger system . . . . .	21
1.3	Particle reconstruction and identification . . . . .	24
1.3.1	Particle flow algorithm . . . . .	24
1.3.2	Muons . . . . .	24
1.3.3	Electrons and photons . . . . .	26
1.3.4	Hadrons and jets . . . . .	28
1.3.5	Hadronic taus . . . . .	33
1.3.6	Missing momentum . . . . .	33
1.3.7	Pileup mitigation . . . . .	34
1.4	Simulation of CMS . . . . .	36
<b>2</b>	<b>Hadronic Resonance Identification</b>	<b>37</b>
2.1	Reconstruction . . . . .	37
2.2	Identification . . . . .	40
2.2.1	Substructure . . . . .	42
2.2.2	A combined tagger . . . . .	48

2.3	Data validation . . . . .	53
<b>3</b>	<b>The Search for <math>t + p_T^{\text{miss}}</math></b>	<b>59</b>
3.1	Signal selection . . . . .	63
3.1.1	Online trigger selection . . . . .	64
3.1.2	Offline signal selection . . . . .	65
3.2	Background estimation . . . . .	68
3.2.1	Visible final states to constrain invisible final states . . . . .	68
3.2.2	Theoretically-limited extrapolations . . . . .	82
3.3	Results . . . . .	88
3.3.1	Constraints on mono-top models . . . . .	92
<b>4</b>	<b>The Search for Invisible Decays of the Higgs Boson</b>	<b>97</b>
4.1	Signal selection . . . . .	98
4.1.1	Online trigger selection . . . . .	98
4.1.2	EW and QCD production of electroweak bosons . . . . .	106
4.1.3	Sensitivity optimization . . . . .	109
4.2	Background estimation . . . . .	110
4.3	Results . . . . .	112

# Chapter 1

## The CMS experiment at the LHC

### 1.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [1]<sup>1</sup> is a circular particle accelerator, 27 km in circumference and between 40 and 175 m below the surface of the French-Swiss border. Designed to collide protons at a maximum center-of-mass energy  $\sqrt{s} = 14$  TeV, the LHC has delivered collisions at  $\sqrt{s} = 7, 8$  TeV (Run 1) and  $\sqrt{s} = 13$  TeV (Run 2); the target energy  $\sqrt{s} = 14$  TeV will be reached in Run 3. In addition to protons, the LHC accelerates and collides heavy nuclei (Pb and Xe) at lower values of  $\sqrt{s}$ . In this thesis, we focus exclusively on data recorded from proton collisions during Run 2.

Protons are brought to the LHC by the multi-stage process [3] depicted in Figure 1.1. Hydrogen atoms are stripped of electrons and accelerated by LINAC2 (a linear accelerator) to a kinetic energy of 50 MeV. LINAC2 then feeds the protons into the Booster ring (final energy of 1.4 GeV), followed by the Proton Synchrotron (26 GeV). From the PS, the protons are injected into the Super Proton Synchrotron (450 GeV). Protons exit the SPS and enter the LHC at one of two places, corresponding to two different beams traveling in opposite directions. The two beams intersect in eight places along the LHC, four of which are instrumented by a detector experiment: CMS, ATLAS, LHCb, and ALICE.

Each proton beam in the LHC is accelerated by eight superconducting cavities

---

<sup>1</sup>Unless otherwise specified, all technical specifications of the LHC are derived from Reference [1]

□

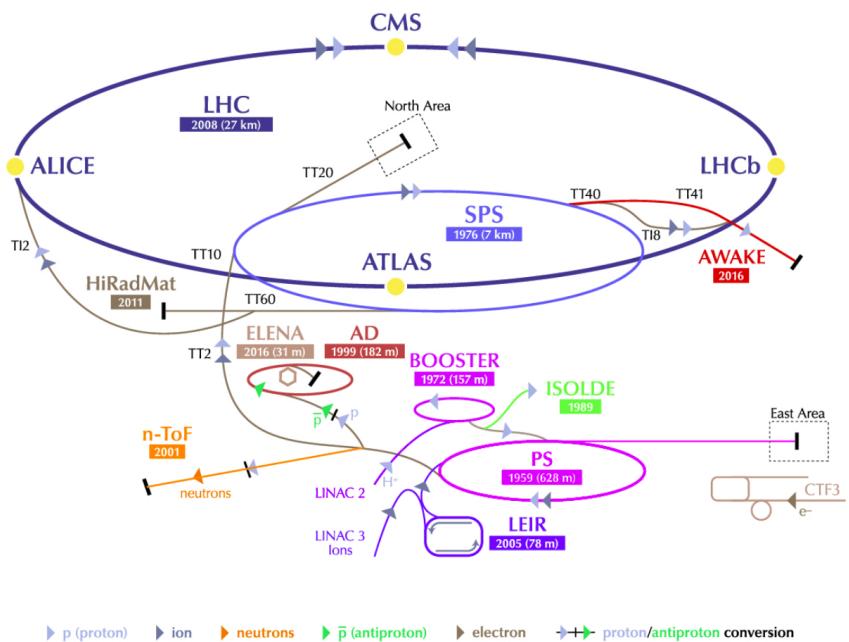


Figure 1.1: Diagram of the CERN accelerator complex. The LHC (dark blue) is fed protons (and heavy ions) by a chain of intermediate accelerators, beginning with LINAC2 (dark pink). Reprinted from Reference [2].

exerting radio frequency longitudinal (i.e. parallel to beam direction) electric fields with a frequency of 400 MHz, The maximum RF voltage seen by each beam is 16 MV per revolution. The physical and temporal design of the RF system creates bunches of protons (corresponding to nodes of the oscillating field) approximately 7.5 cm in length and separated by 25 ns. Superconducting NbTi dipole magnets bend the two proton beams in opposite directions as they travel around the ring. Each of the 1232 dipoles is 14 m long and exerts a transverse  $B$  field between 0.54 and 8.33 T. To achieve such high  $B$  fields, the magnets are cooled to 2 K by superfluid helium. In addition, a number of quadrupole magnets are used to focus and match the beams between the dipoles<sup>2</sup>.

In addition to the center-of-mass energy  $\sqrt{s}$ , the other figure of merit is the number of events producing interesting physics processes, which is defined as:

$$N(pp \rightarrow X) = \int dt L\sigma(pp \rightarrow X) \quad (1.1)$$

where  $\sigma$  is the cross section of the relevant process and  $L$  is the instantaneous luminosity of the LHC. The cross section is fixed by nature, and so increasing the luminosity is the only handle to increase  $N$ . The instantaneous luminosity of two Gaussian beams is given by [1]:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma F}{4\pi \epsilon \beta^*} \quad (1.2)$$

where:

$N_b$  = particles per bunch

$n_b$  = bunches per beam

$f_{\text{rev}}$  = frequency of revolution

$\gamma = E/m$  of beam

$\epsilon$  = emittance of beam

---

<sup>2</sup>Full details on the various quadrupoles can be found in Table 3.7 of Reference [1].

□

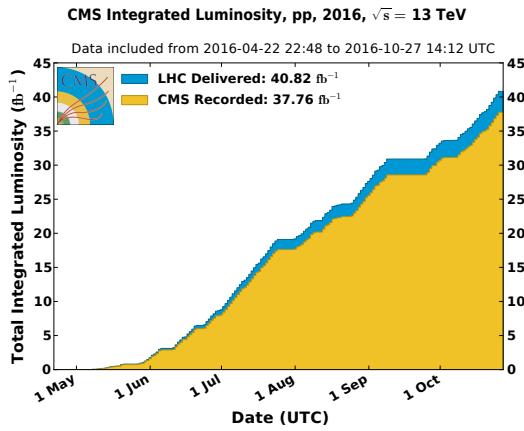


Figure 1.2: Integrated luminosity of the LHC during proton collisions during the 2016 data-taking period [4].

$\beta^*$  = beta function at collision point

$F$  = factor accounting for beam intersection geometry

The instantaneous luminosity evolves as a function of time, primarily due to  $n_b$  and  $N_b$  being modified by collisions. The total integrated luminosity after time  $T$  is:

$$L_{\text{int}} = \int_0^T dt L(t) = L(0) \tau_L \left(1 - e^{-T/\tau_L}\right) \quad (1.3)$$

where  $\tau_L \approx 15$  h is the characteristic beam loss timescale and  $L(0)$  is the instantaneous luminosity at  $T = 0$ . The LHC is designed to deliver  $L(0) \sim \mathcal{O}(10^{34}) \text{ cm}^{-2}\text{s}^{-1}$ . Figure 1.2 shows the total luminosity delivered by the LHC and recorded by CMS during the 2016 portion of Run 2.

## 1.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) [5] is one of two general purpose LHC detectors (the other being ATLAS). It is designed to detect and measure stable hadrons, photons, electrons, and muons produced in proton and ion collisions at LHC interaction point 5. From these event descriptions, a number of physics processes can be probed,

including SM measurements [?], BSM searches [?], and the discovery of the Higgs boson [?]. In what follows, we will use the  $(r, \phi, \eta)$  coordinate system with respect to the  $z$  axis:

$z$  = distance along beam axis, with  $z = 0$  defined to be at the center of the detector

$r$  = distance from the  $z$  axis

$\phi$  = azimuthal angle in the plane orthogonal to the  $z$  axis

$\eta$  = pseudorapidity ( $-\log \theta/2$ ), with respect to the polar angle  $\theta$

In this coordinate system, we define  $x$  and  $y$  to lie in the plane perpendicular to  $z$ , with  $x$  pointing from the center of the detector to the center of the LHC. As with the pseudorapidity, it is convenient to use quantities invariant under  $z$ -boosts, and so we define the transverse momentum:

$$\vec{p}_T = \begin{pmatrix} p_x \\ p_y \end{pmatrix} \quad (1.4)$$

We will frequently make use of the magnitude of this vector,  $p_T$ . CMS can detect collision products that are within the fiducial volume of  $0 \leq \phi < 2\pi$  and  $-5 \leq \eta \leq 5$ . Several detector subsystems (Figure 1.3) are used to identify and reconstruct muons, electrons, photons, and charged and neutral hadrons.

### 1.2.1 Silicon tracker

Starting from the beam pipe, the first of these subsystems is the silicon tracker [6], used to identify charged particles and measure their momenta. The tracker consists of silicon detector geometries: pixels (providing 3D position measurement) and strips (2D). The arrangement of the pixel and strip layers are shown in Figure 1.4. A near-uniform 3.8 T magnetic field, produced by a superconducting NbTi solenoid, envelopes the tracker. The field lines in the tracker volume are approximately parallel to the beam direction.

[]

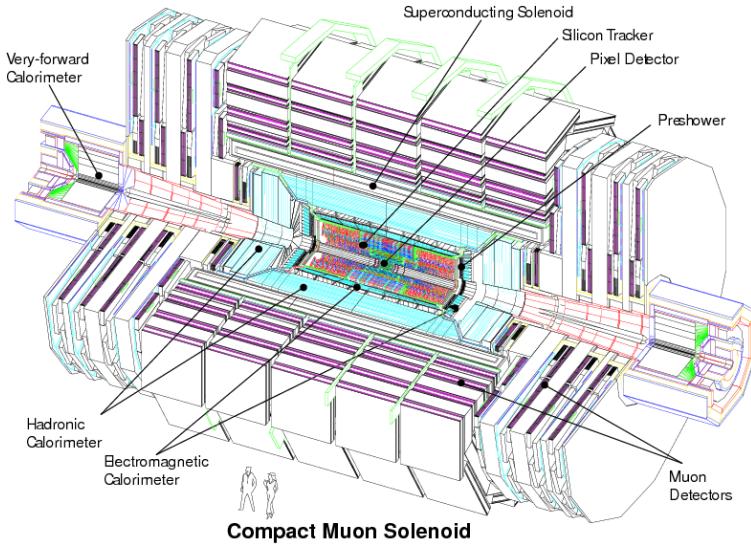


Figure 1.3: Cut-away view of the CMS detector and its subsystems. Reprinted from Reference [5].

A single silicon pixel has dimensions  $285 \times 100 \times 150 \text{ } (\mu\text{m})^3$  (in  $r \times r\phi \times z$ ), leading to a position resolution of  $\sim 10 \times 30 \text{ } (\mu\text{m})^2$  (in  $r\phi \times z$ ). The 66 million pixels are arranged into 7 layers: 3 cylindrical “barrels” (at  $r = 4.4, 7.3, 10.2 \text{ cm}$ ) and  $2 \times 2$  “endcap” annuli (at  $z = \pm 34.5, \pm 46.5 \text{ cm}$ ). Outside the pixel layers are the strip layers, consisting of 9.3 million silicon strips arranged into barrels and endcaps. The resolution in  $r\phi$  varies between 10 and  $50 \text{ } \mu\text{m}$ , depending on the location and pitch of the given strip. Certain strip layers contain two layers of strips, rotated through a “stereo” angle (100 mrad) with respect to each other. By matching adjacent hits, the stereo measurement can add a third dimension ( $z$  for barrel,  $r$  for endcap) to the strip’s 2D measurement, with resolution 100-500  $\mu\text{m}$ . There are a total of 10 barrel layers ( $0.2 < r < 1 \text{ m}$ ) and 24 endcap layers ( $0.6 < |z| < 2.8 \text{ m}$ ).

Pixels with a signal greater than a tuneable readout threshold (typically around  $3000Q_e$ ) are read out. These pixels are then aggregated with adjacent signals to form pixel clusters, which are further subjected to readout thresholds ( $\sim 4000Q_e$ ). The exact position of the particle in this layer (known as a “hit”) is inferred by fitting the charge distribution of the pixels in this cluster to pre-determined templates. A similar

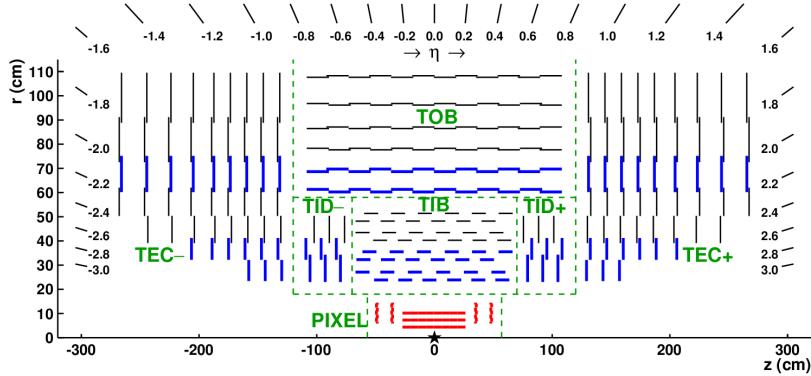


Figure 1.4: Diagram of a slice of the CMS tracking system. The pixel layers are shown in bold red lines. Single-strip (double-strip) layers are indicated by thin black (bold blue) lines. The double-strip modules each consist of two back-to-back strips, rotated with respect to each other, that can provide 3D localization of the hits. Reprinted from Reference [6].

method is employed to determine the strip hit positions, with some modifications to account for Lorentz drift of the charges in the silicon detector due to the  $B$ -field. The efficiency of reconstructing hits varies with the detector type, location, and particle momentum, but is generally greater than 99% (99.5% if defective modules are not considered).

## Tracking

Tracks are found using an iterative “inside-out” process, where each iteration has five steps:

1. Define seeds using pixel hits, double-strip hits (i.e. hits with 3D information), and an estimate of the beam spot (collision point). At least 3 hits are needed for the seed.
2. Use a Kalman filter [7, 8] to evolve track seeds through the rest of the tracker and find hits, accounting for the  $B$ -field and energy loss.
3. Estimate trajectory parameters after finding all hits.
4. Decide whether to keep found tracks based on quality requirements (e.g. number

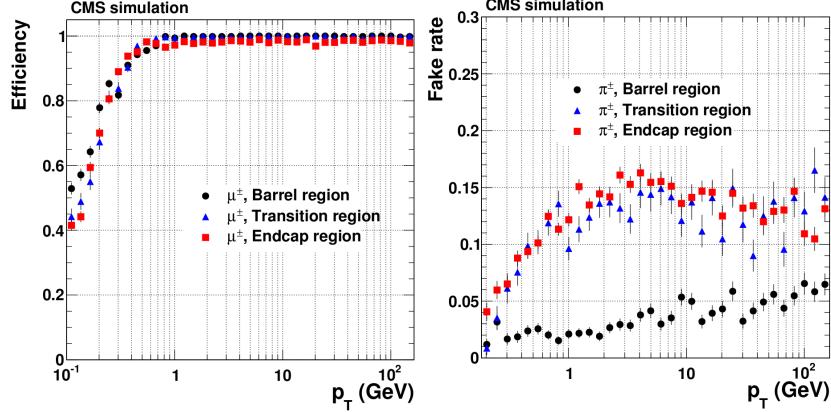


Figure 1.5: Efficiency (fake rate) of the CMS track fit algorithm, evaluated using simulation of muons (charged pions). Reprinted from Reference [6].

of missing hits, track  $\chi^2$ )

5. Remove hits associated with tracks from hit collection and repeat.

The trajectory parameters referred to in step 3 are the 5 parameters of a helix:  $\rho$  (curvature),  $\phi_0$  (azimuthal angle),  $\lambda$  ( $\cot\theta$ ),  $d_0$  (“impact parameter”, minimum  $r$  of track),  $z_0$  (minimum  $|z|$  of track). The CMS track fit typically has 5-7 iterations, with each successive iteration loosening the seed and track fit requirements to look for more difficult tracks (e.g. missing hits, large  $d_0$ ). The efficiency and fake rate of this reconstruction, as a function of track  $p_T$ , are shown in Figure 1.5. For muons with  $|\eta| < 1.5$  and  $p_T > 1$  GeV, the tracking efficiency is over 98%, with a combinatorial fake rate of 2-6%.

## Vertexing

The excellent position resolution of the pixel detector is used to accurately measure the position of primary vertices, as well as any secondary vertices from the decays of long-lived particles. In the former case, tracks are first clustered together on the basis of the likelihood that the tracks in a cluster arise from a single primary vertex. This is done using a deterministic annealing algorithm [9], which has as free parameters the number of clusters and the probability of each track belonging to each cluster. Having determined the clusters, an adaptive fit algorithm [10] is used to determine the vertex

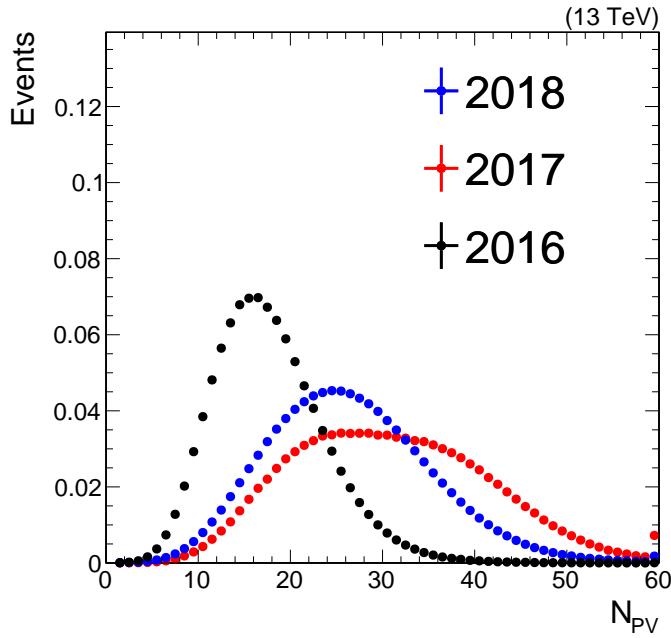


Figure 1.6: The distribution of the number of reconstructed primary vertices in data recorded by CMS during Run 2 of the LHC. While the results in this thesis only concern 2016 data, we show the evolution of  $N_{\text{PV}}$  as a function of time, as this correlates directly with increased instantaneous luminosity.

for each cluster. The free parameters of this fit are the three spatial coordinates of the vertex. As the LHC collides bunches of  $\mathcal{O}(10^{11})$  protons, we expect multiple primary vertices in a single collision, and this is reflected in Figure 1.6. The vertex defined to be the hard scattering interaction (known as *the* primary vertex<sup>3</sup>) is the vertex which maximizes:

$$\sum_{j \in \text{track jets}} (p_{\text{T}}^j)^2 + (p_{\text{T}}^{\text{miss}})^2 \quad (1.5)$$

where “track jets” refer to jets (Section 1.3.4) clustered from the vertex’s tracks, and  $p_{\text{T}}^{\text{miss}}$  is defined in Section 1.3.6.

---

<sup>3</sup>This nomenclature is indeed confusing, defining the singular primary vertex to be one of many primary vertices. However, it is standard terminology in CMS publications, so we will continue to use it. In what follows, the distinction will be clear

## Secondary vertexing

The last reconstruction algorithm concerning the tracker alone is the identification of secondary vertices, which arise from the decay of long-lived particles (e.g.  $B$  mesons). The inclusive vertex fitter (IVF) [11] reconstructs such secondary vertices by the following steps:

1. Select a track as a seed if it satisfies  $\sqrt{d_0^2 + d_z^2} > 50 \mu\text{m}$  and  $d_0 > 1.2\delta d_0$ .
2. Choose nearby tracks based on their closest distance to and opening angle with the seed track.
3. Fit the tracks to a displaced vertex using the adaptive fitter [10].
4. Decide which tracks belong to the candidate secondary vertex and which belong to the primary vertex.
5. Re-fit the secondary vertex position only using the former set of tracks from the previous step.

It is important not only to properly determine the location of the secondary vertex, but also to properly assign tracks. Observables that are a function of the *tracks* (e.g. vertex mass) will be critical for  $b$  jet tagging.

### 1.2.2 Electromagnetic calorimeter

The CMS electromagnetic calorimeter [12] (ECAL) is a homogenous detector with good energy and angular resolution, composed of 76,000  $\text{PbWO}_4$  crystals. The crystals are arranged in two sections: a cylindrical barrel (EB) covering  $|\eta| < 1.44$  and two endcap annuli (EE) extending to  $|\eta| < 3$ . This provides slightly more coverage than the tracking volume. Each crystal in the EB (EE) has dimensions  $2.2 \times 2.2 \times 23$  ( $2.68 \times 2.68 \times 22$ ) ( $\text{cm}^3$ ), with the long dimension pointing towards the beam. This can be compared to a Moli  re radius  $r_M = 2.19 \text{ cm}$  and a radiation length of  $X_0 = 0.89 \text{ cm}$ . A cross-sectional area comparable to  $r_M \times r_M$  facilitates the differentiation of different electromagnetic (EM) showers arising from electrons and photons. The depth

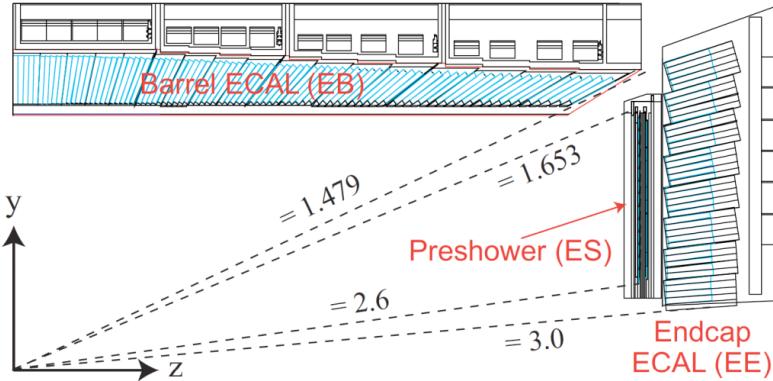


Figure 1.7: One quadrant of the CMS ECAL (symmetric with rotation around  $z$  and reflection across  $z = 0$ ). The dashed lines indicate values of  $\eta$ . Reprinted from Reference [12].

of the crystal (in units of  $X_0$ ) drives the excellent energy resolution, which is determined using a electron beam:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\% \quad (1.6)$$

Scintillation photons from the  $\text{PbWO}_4$  crystals are collected by avalanche photodiodes (APDs) in the EB and vacuum phototriodes (VPTs) in the EE, which provide amplification factors of 50 and 10, respectively.

At high momenta, the two photons from a  $\pi^0$  decay may merge into a single ECAL crystal. This primarily occurs at high  $|\eta|$  due to the  $z$ -boost of the intitial state. To differentiate one- and two-photon deposits, a “preshower” detector sits in front of the EE ( $1.6 < |\eta| < 2.5$ ). The preshower detector consists of a lead absorber and silicon strips. A photon (or photon pair) initiates a shower in the lead. The shower can be resolved in the silicon strips, which have resolution  $\mathcal{O}(1\text{--}10)$  mm.

The physical placement of all three ECAL components is shown in Figure 1.7.

Due to the bending of a charged particle’s trajectory in the solenoidal  $B$ -field, bremsstrahlung photons will be emitted at similar values of  $\eta$ , but spread along  $\phi$ . A “supercluster” (SC) is defined by clustering nearby ECAL energy depositions, allowing for a wider spread in  $\phi$  than in  $\eta$  (Figure 1.8). The particle’s EM energy is defined to

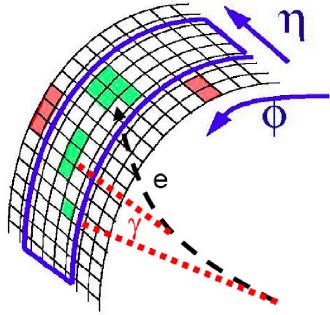


Figure 1.8: The combination of multiple ECAL crystals into a single supercluster, intended to capture energy depositions from bremsstrahlung photons. Reprinted from Reference [14].

be the weighted sum of the energies of all crystals in the SC, where the coefficients account for crystal-specific calibration effects [13]. For an electron or photon, the EM energy is typically the energy of the particle, whereas for other particles (charged hadrons and some muons), it is only a fraction of the total energy.

### 1.2.3 Hadronic calorimeter

The hadronic calorimeter (HCAL) [15, 16, 17] is used to identify and measure the energy of hadrons. It consists of 4 calorimeters: barrel (HB;  $|\eta| < 1.4$ ), endcap (HE;  $1.3 < |\eta| < 3$ ), forward (HF;  $3 \lesssim |\eta| < 5$ ), and outer (HO;  $|\eta| < 1.3$ ). Their arrangement is shown in Figure 1.9.

The HB and HE are both composed of alternating absorber and plastic scintillator layers. The absorber is a non-magnetic brass alloy with an interaction length  $\lambda_I = 1.5$  cm. The absorber layers range in thickness from 40 to 75 mm in the HB and HE, providing a total of  $5.8\text{-}10.6\lambda_I$  of material, depending on the  $\eta$  of the particle. These dimensions are limited by the constraint that the HB and HE be inside the solenoid. To augment the number of interaction lengths, additional layers of plastic scintillator sit outside of the solenoid. These comprise the HO, which use the magnet as an absorber, providing an additional  $\sim 1.1\lambda_I$ . The light from the scintillator tiles is read out by hybrid photodiodes (HPDs) in the HB and HE and by silicon photomultipliers (SiPMs) in the HO.

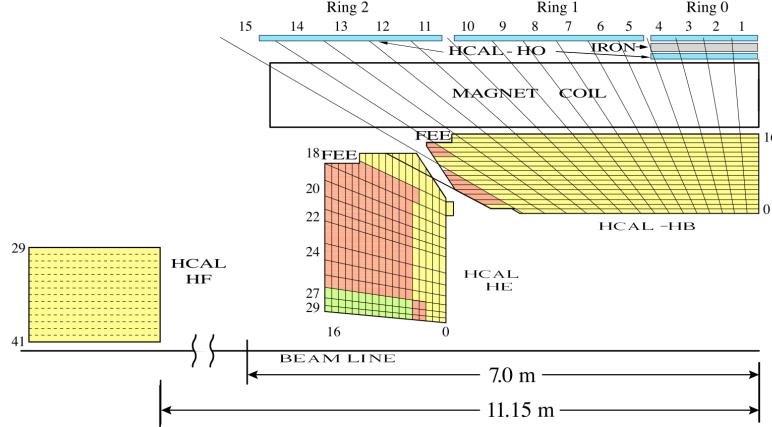


Figure 1.9: One quadrant of the CMS HCAL (symmetric with rotation around  $z$  and reflection across  $z = 0$ ). Note that the slight overlap of the detectors in  $\eta$  ensures the hermeticity of the detector. Reprinted from Reference [17].

Beyond the HE, at 11 m from the interaction point, sits the HF. The HF is also a sampling calorimeter, made of steel absorbers instrumented with quartz fibers. Charged particles from the nuclear shower in the steel emit Cerenkov radiation in the quartz fibers, which transports the light to photomultiplier tubes (PMTs). The HF uses more radiation-hard materials than the rest of the HCAL, as this  $\eta$  range is subject to significantly more radiation from collisions than the central part of the detector.

Each HCAL subsystem is read out in “towers”, corresponding to sections in  $\eta$  and  $\phi$ . For  $|\eta| < 1.74$ , the segmentation is  $\eta \times \phi = 0.087 \times 0.087$ . Beyond this, the towers vary in size from  $0.09 \times 0.175$  to  $0.35 \times 0.175$ .

The energy resolution of the HCAL must be considered in conjunction with the ECAL, as energy can be deposited in both detectors. This is calibrated using test beams of various charged particles. For the HB and HE, the resolution is:

$$\frac{\sigma_E}{E} = \frac{0.847}{\sqrt{E/\text{GeV}}} \oplus 0.074 \quad (1.7)$$

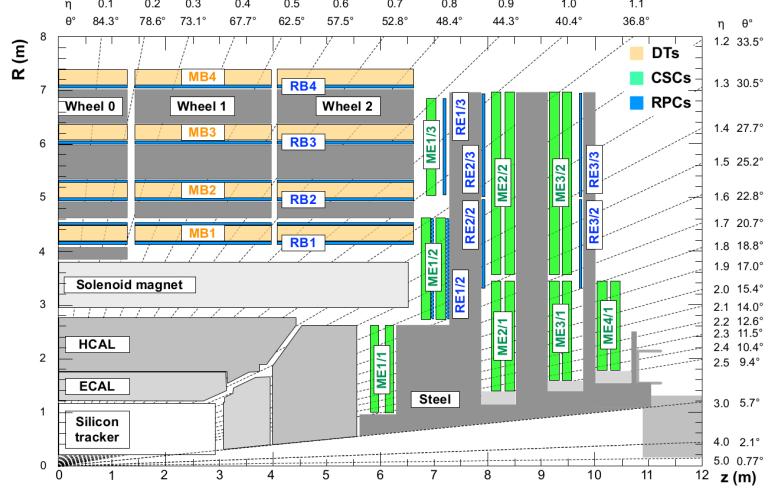


Figure 1.10: One quadrant of the CMS muon detection system, with the DTs (MBs), CSCs (MEs), and RPCs (RBs/REs) labeled. Reprinted from Reference [18].

whereas for the HF, it is:

$$\frac{\sigma_E}{E} = \frac{1.98}{\sqrt{E/\text{GeV}}} \oplus 0.09 \quad (1.8)$$

### 1.2.4 Muon chambers

The muon detectors are gas ionization chambers and are the outermost component of CMS [18]. The chambers are interleaved with the steel return yoke of the CMS magnet, ensuring a  $B$ -field which runs anti-parallel to the field inside the solenoid. This results in the characteristic  $S$ -shape of muon trajectories, as the bending changes direction across the solenoid. 3 types of ionization detectors are used: drift tubes (DTs; barrel), cathode strip chambers (CSCs; endcaps), and resistive plate chambers (RPCs; barrel and endcaps). The physical placement of the detectors is shown in Figure 1.10.

The DT cells are filled with a 85%/15% mix of argon and carbon dioxide, with a gold/steel anode wire held at a voltage of 3600 V. The cell is a rectangular prism, with transverse dimensions of  $42 \times 13 \text{ mm}^2$  and a longitudinal dimension ranging from 1.9 to 4.1 m. The dimensions and the drift speed of  $55 \mu\text{m}/\text{s}$  result in a maximum response (drift) time of 400 ns. The DTs are organized into cylindrical “stations”

(MB). With the exception of the outermost barrel station (MB4), each MB consists of 3 “superlayers” (SLs). Each SL has 4 parallel drift cells, and so measure position in a particular plane. Of the 3 SLs, two measure  $r\text{-}\phi$  position and one measures  $r\text{-}z$  position. MB4 does not have an  $r\text{-}z$  SL.

The endcap is instrumented with CSCs as these have a faster response time and better spatial resolution than DTs. This is needed in the forward region, where both the muon and background fluxes are higher. While the CSC dimensions vary depending on the position of the chamber, each is instrumented with 80 cathode strips, held at voltages (relative to the anode) of 2.9-3.6 kV. The CSC wires have a separation of 2.5-3.16 mm, which governs the position resolution. A 50%/40%/10% mix of CO<sub>2</sub>/Ar/CF<sub>4</sub> is used to fill the chambers.

RPCs are interspersed among the DTs and CSCs in both the barrel and endcap. These serve as a very fast muon detector ( $\sim 1$  ns) for the online trigger system. The spatial resolution of the RPC hits is worse than the DTs and CSCs.

The hit position resolution of the DTs is 78-120  $\mu\text{m}$  (140-390  $\mu\text{m}$ ) in  $r\text{-}\phi$  ( $r\text{-}z$ ). For CSCs, the resolution varies from 40 to 152  $\mu\text{m}$ . The efficiency of individual reconstructed hits (“rechits”) is over 95%.

### 1.2.5 Online trigger system

Figure 1.11 shows various cross sections of  $pp$  collisions as a function of  $\sqrt{s}$ . While it is clear that interesting SM processes (e.g. production of  $W, Z, t$ ) increase as a function of  $\sqrt{s}$ , they still sit several orders of magnitude below the inclusive cross section  $\sigma_{\text{tot}} \sim 10^8$  nb. To produce an appreciable number of rare events, the LHC has bunch crossings every 25 ns (40 MHz). However, it is not possible to read out the entire detector at this rate, much less reconstruct the data and store it. CMS uses a two-stage trigger system [19] to refine the events to keep for permanent storage and analysis. First, a Level 1 (L1) hardware-based trigger selects interesting events at a rate of 100 kHz. These events are then fed to the high level trigger (HLT) to partially reconstruct the events on a CPU farm. The final selected data rate from the HLT is 400 Hz.

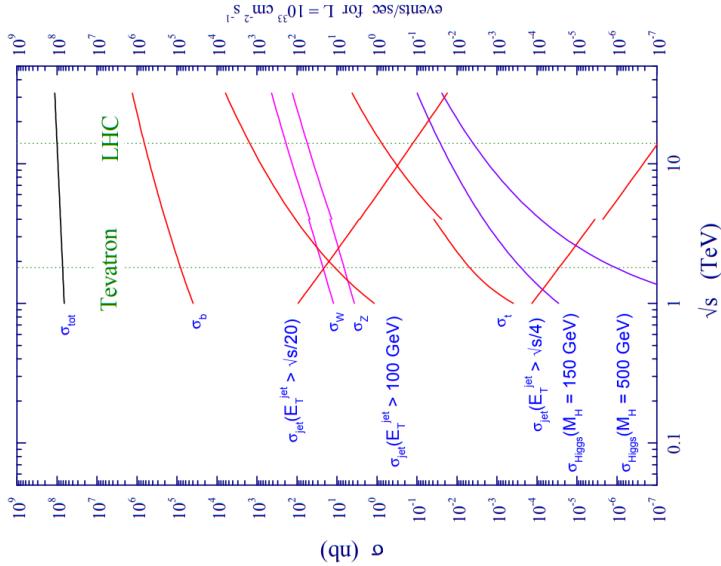


Figure 1.11: Various cross sections of interesting SM processes compared to the inclusive cross section of  $pp$  collisions, at various values of  $\sqrt{s}$ . Reprinted from Reference [20].

## Level 1

The L1 trigger makes decisions within  $4 \mu\text{s}$  of collisions using field programmable gate arrays (FPGAs) and application specific integrated circuits (ASICs), which offer significant speed advantages as compared to CPUs for certain tasks. Individual detector systems (ECAL, HCAL, muon detectors) feed simple reconstructed objects (“trigger primitives” (TPs)) to a series of regional trigger decisions. Quality selections are placed on calorimeter towers, and they are aggregated into clusters of energy deposits. A simple segment-finding and tracking algorithm is run on hits in the muon chambers to produce muon tracks. Note that the inner tracker is not included in the L1 decision: at the time of the construction of the CMS detector, the detector readout and reconstruction algorithms were not fast enough for the L1’s requirements.<sup>4</sup> Regional trigger decisions are sent to the global trigger (GT), which correlates the TPs it receives. Some GT trigger decisions are localized, such as requiring a high-energy ECAL deposit (e.g.  $e, \gamma$ ). Others require computing event-wide observables, such as

---

<sup>4</sup>It should be noted that implementing tracking for hardware triggers is a significant goal for the community in the next few years.

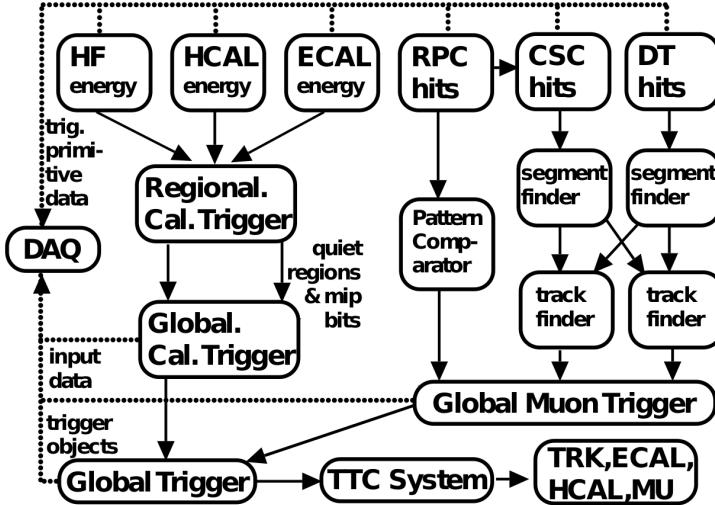


Figure 1.12: Schematic diagram of the CMS L1 trigger system. In addition to the region and global trigger decisions (described in the text), the flow of data also includes the trigger, timing, and control system (TTC). Reprinted from Reference [19].

a large momentum imbalance (e.g. neutrinos, DM). The GT also computes simple jets. If the GT decides to select the event, the detector is read out and forwarded to the data acquisition system (DAQ). Figure 1.12 shows a schematic description of this process.

### High level trigger

The full detector readout is picked up by the HLT. The HLT computing farm consists of 20k CPU cores (this number has grown from 13k in 2012 and continues to grow). Unlike the L1, the HLT has access to the entire event information. Therefore, a full reconstruction of the event is performed, using slightly algorithms similar to (in some cases, simplified) the algorithms used in offline reconstruction. To be selected by the HLT, an event must pass an HLT “path”. A path consists of a series of filters, each making simple decisions that can be chained into complex decisions. For example, three filters may be: (1) two forward jets, (2) large momentum imbalance, and (3) an electron and a muon. These filters could be chained into two separate paths, targeting vector boson fusion production of the Higgs, where the Higgs decays to DM (1&2) or two taus (1&3). Simple decisions (calorimeter and muons) are computed before

complex decisions (tracking). Events that are selected by the HLT are passed on to the Tier 0 computing farm for full offline reconstruction, and then onwards to disk and tape resources for analysis and storage, respectively.

## 1.3 Particle reconstruction and identification

### 1.3.1 Particle flow algorithm

Because of the excellent angular granularity of the CMS detector and the momentum resolution of the tracker, a particle flow [16] algorithm is used to correlate information from all detector subsystems to build a global description of each event. Particle flow (PF) algorithms date back to ALEPH [?], and are in contrast to detector-specific physics object-based algorithms used at other experiments [?].

The key feature of the PF algorithm is to “link” multiple detector signals together into a single PF candidate. This linkage combines inner tracks, ECAL clusters, HCAL clusters, and muon tracks based on their proximity in the  $(\eta, \phi)$  plane. Inner track helices are extended into the calorimeters, searching for clusters compatible with the trajectory. Similarly, clusters from the ECAL, the ECAL preshower, and the HCAL can be linked without a track present. The remainder of this section is organized according to “blocks” in the PF algorithm. At the end of each block, any detector signature (track, calorimeter cell) which has been assigned to a PF candidate in that block is removed from the set of objects passed to the next block. For example, tracks associated with muons will not be considered when reconstructing charged hadrons.

### 1.3.2 Muons

The first block of the PF algorithm links inner tracks and muon chamber hits to identify muons. While “standalone” muons can rely solely on muon chamber hits, we also consider muons that use the inner tracker. To construct outside-in “global muons”, standalone muon tracks are first reconstructed using a Kalman filter fit using muon chamber rechits [21, 18]. These standalone tracks are extrapolated inwards

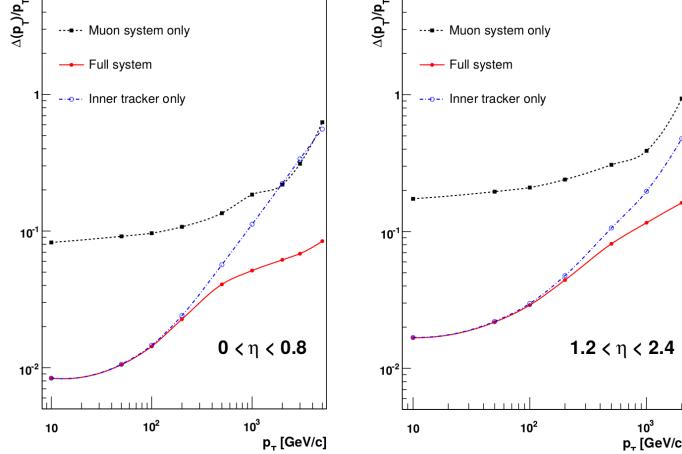


Figure 1.13: Resolution of muon  $p_T$ , compared between the different reconstruction algorithms. The salient feature is that the muon systems become important to the momentum measurement at approximately 200 GeV. Reprinted from Reference [18].

to the inner tracker, accounting for effects from the magnetic field and the material budget between the tracker and the muon chambers. A second Kalman filter fit is run to combine the inner track with the standalone track to form a global muon. To reject backgrounds, requirements are placed on the  $\chi^2/N_{\text{dof}}$  (poorly fit tracks, charged hadrons) and  $d_0, d_z$  (muons from hadron decays and cosmic rays). The total efficiency of global muon reconstruction is 99%. In addition, “tracker” muons can be reconstructed from the inside-out, extrapolating inner tracks to the muon chambers and re-fitting with muon rechits. The PF algorithm uses both global and tracker muons, but in this thesis, we will only use the former. Figure 1.13 compares the momentum resolution of muons reconstructed using the different algorithms. The PF selection consumes all three types of reconstructed muons: global, tracker, and standalone.

We define two muon identification (ID) criteria that are slightly stronger than the PF selection. The first is a loose ID, which is used to veto muons when defining muon-free data samples. A loose muon is required to be either a global or tracker muon and have PF isolation less than 0.25. PF isolation is defined as:

$$\left( \sum_{i \in \text{PV charged had.}} p_T^{(i)} + \max \left\{ 0, \sum_{i \in \text{neut. had.}} p_T^{(i)} + \sum_{j \in \gamma} p_T^{(j)} - \frac{1}{2} \sum_{k \in \text{PU charged had.}} p_T^{(k)} \right\} \right) / p_T^{(\mu)} \quad (1.9)$$

Table 1.1: Observables used in identifying muons and rejecting backgrounds

Observable	Notes
$p_T$	Backgrounds grow at low $p_T$ .
$\chi^2/N_{\text{dof}}$	Ensure a good track fit.
$N_{\text{hit}}^{\mu\text{on}}$	At least one hit in the muon chamber in the global fit.
$N_{\text{stations}}$	At least two muon chamber stations contain segments of the track.
$d_0$	Track impact parameter, remove cosmic rays and hadron decays.
$d_z$	Track impact parameter, as above but also for pileup.
$N_{\text{hit}}^{\text{pixel}}$	At least one pixel hit.
$N_{\text{hit}}^{\text{tracker}}$	At least 5 hits in the tracker.

where the sums are over PF candidates within  $\Delta R < 0.4$  of the muon. PV and PU refer to the primary vertex and pileup vertices, respectively. The tight muon ID requires a global muon, PF isolation less than 0.15, and has selections on the criteria in Table 1.1. The efficiency

To account for differences between data and simulation in the performance of the muon IDs, corrections (known as scale factors) are derived:

$$\text{SF} = \frac{\epsilon_{\text{Data}}}{\epsilon_{\text{MC}}} \quad (1.10)$$

While  $\epsilon_{\text{MC}}$  can be computed directly from MC truth information,  $\epsilon_{\text{Data}}$  must be computed from  $Z \rightarrow \mu\mu$  events, using one muon as a reference tag, and the other to probe the efficiency of the ID being tested.

### 1.3.3 Electrons and photons

Both electrons and photons are seeded using ECAL SCs, with electrons also being associated with an isolated track. Due to the significant material budget of the silicon tracker, electrons will lose a significant amount of energy through bremsstrahlung. Although this energy is partially recovered through the supercluster algorithm, the standard Kalman filter tracking fit does not properly account for the non-Gaussian hit uncertainties induced by bremsstrahlung. Therefore, a modified algorithm based on a

Table 1.2: Observables used in identifying electrons and rejecting hadron and photon backgrounds

Observable	Notes
$p_T$	Backgrounds grow at low $p_T$ and brem. photons can make low- $p_T$ tracking difficult.
$\sigma_{i\eta i\eta}$	Energy-weighted width of cell $\eta$ in SC. Small for electrons.
$ \Delta\eta(\text{track, SC}) $	$\Delta\eta$ between SC seed crystal and GSF track at PV. Small for electrons.
$ \Delta\phi(\text{track, SC}) $	$\Delta\phi$ , as above.
$E_H/E_{\text{EM}}$	Ratio of HCAL and ECAL energies. Large for hadrons.
PF isolation	Sum of energies of other PF candidates near electron. Large for particles in jets.
$ 1/E - 1/p $	Checks if ECAL and tracker agree ( $m_e \sim 0$ )
$N_{\text{hit}}^{\text{miss}}$	Conversions or bad tracks will have multiple missing hits in the inner tracker.
Conversion veto	Check for a pair of tracks originating at a displaced vertex.

Gaussian Sum Filter (GSF) [6, 22] is used to fit the track. Hits in the pixel and strip endcap layers consistent with the ECAL SC are used to seed the GSF fit. GSF differs from a Kalman filter by modeling the uncertainties as a Gaussian mixture model instead of a single Gaussian. The primary backgrounds for electron identification are (1) the overlap of a charged hadron with a neutral hadron or photon and (2) a photon which converts into a  $e^-/e^+$  pair in the tracker. Table 1.2 lists the observables used to reject these backgrounds; two cut-based IDs are defined. The “loose” ID selects electrons with  $\sim 90\%$  signal efficiency and  $\sim 0.5\%$  background acceptance (strongly dependent on the electron phase space); it is used to *veto* electrons. The “tight” ID selects electrons with  $\sim 70\%$  signal efficiency and  $\sim 0.1\%$  background acceptance; it is used to *select* electron-pure samples.

Photons are defined as ECAL SCs without matched tracks. Table 1.3 describes the selection variables used to define the loose veto ID ( $\epsilon_{\text{sig}} \approx 90\%$ ,  $\epsilon_{\text{bkg}} \approx 17\%$ ) and the tight selection ID ( $\epsilon_{\text{sig}} \approx 82\%$ ,  $\epsilon_{\text{bkg}} \approx 12\%$ ). The direction and energy of a photon are defined by the ECAL SC position and energy, respectively.

To calibrate ECAL SCs, we define the  $R_9 = E_{3\times 3}/E_{\text{SC}}$ .  $E_{3\times 3}$  is the sum of the

Table 1.3: Observables used in identifying photons and rejecting hadron backgrounds

Observable	Notes
$p_T$	Backgrounds grow at low $p_T$ .
$\eta$	EB resolution is better than EE; less hadron background.
$\sigma_{i\eta i\eta}$	Defined in Table 1.2.
$E_H/E_{EM}$	Defined in Table 1.2.
PF isolations	Defined in Table 1.2. For photons, separate isolation criteria are placed on each PF type (photon, charged hadron, neutral hadron).

energies of the crystals in a  $3 \times 3$  square centered on the most energetic crystal in the SC. Much like  $\sigma_{i\eta i\eta}$ , it is sensitive to the width of the shower shape. A regression to correct the energy scale [13] is trained as a function of SC energy,  $\eta$ ,  $R_9$ , and the width of the SC in  $\phi$ . Differences between data and MC in the efficiencies of the electron and photon IDs are corrected using scale factors using  $Z \rightarrow ee$  events, as was described for muon IDs. In the case of the photon ID, the electron SC is used as a proxy for the photon.

### 1.3.4 Hadrons and jets

The remaining final states considered by the PF algorithm are hadrons. These are primarily found in jets, along with nonisolated photons (from  $\pi^0$  decays) and muons (from heavy hadron decays). If a calorimeter cluster is within the tracker acceptance (and not already consumed by an earlier PF block), and it is not linked to a track, then it is assumed to have arisen from a photon or neutral hadron. If the cluster is in the ECAL, then the constructed PF candidate is a photon; if in the HCAL, it is a neutral hadron. Unlike for charged hadrons (described next), proximal ECAL and HCAL clusters are not linked together. This is because neutral hadrons are expected to leave very little energy in the ECAL, and the tracking efficiency for charged hadrons is over 90%. To find charged hadrons, HCAL clusters, ECAL clusters, and tracks are linked together. The calorimeter energy of the clusters is estimated using a calibrated sum of the HCAL and ECAL deposits. If this energy is significantly (500 MeV) above the

momentum estimate from the tracks, then it is assumed there are additional neutral hadrons or photons in the cluster. The number of charged hadrons is the number of reconstructed tracks. Outside of the tracker acceptance, ECAL clusters not linked to HCAL clusters are assigned to PF photons. Linked HCAL and ECAL clusters are assumed to arise from a shower containing neutral and charged hadrons.

## Jet clustering

Jets arise from the hadronization and fragmentation of colored particles, as described in Section ???. The precise definition of a jet is dependent on the algorithm used to cluster particles, and so we will use “jet” to refer to the final states of hadronization/fragmentation and to the output of a jet finding algorithm. LHC experiments typically use “sequential recombination” algorithms [23, 24, 25]. Given a set of particles (in CMS, PF candidates) in the event  $E$ , we compute two metrics:

$$\begin{aligned} d_{iB} &= p_{\mathrm{T},i}^2 \\ d_{ij} &= \min\{p_{\mathrm{T},i}^{2q}, p_{\mathrm{T},j}^{2q}\} \frac{\Delta R(p_i^\mu, p_j^\mu)^2}{R} \end{aligned} \quad (1.11)$$

where  $i, j$  are two elements of  $E$ ;  $B$  represents the beam; and  $p, R$  are tunable parameters. Let  $\tilde{E} = E \cup \{B\}$  and find  $i, j \in \tilde{E}$  that minimize  $d_{ij}$ . If  $j = B$ , then we remove particle  $i$  from  $\tilde{E}$  and add it to the set of candidate jets. Otherwise, we combine  $i$  and  $j$  into a new particle  $k$  by defining  $p_k^\mu = p_i^\mu + p_j^\mu$ . This “pseudojet”  $k$  is added to  $\tilde{E}$ , while  $i$  and  $j$  are removed. This process is repeated until  $\tilde{E}$  is exhausted. The exact value of  $R$  is an approximate measure of the  $\Delta R$  radius of the jet. The value of  $q$  defines the relationship between the momentum and angular factors;  $q < 0$  enforces clustering circular jets around hard seeds. For standard single-parton jets, CMS uses  $q = -1$  (referred to as anti- $k_{\mathrm{T}}$ ) and  $R = 0.4$  (AK4). In Section 2.1, we will discuss the use of  $q = 0$  and  $R \gg 0.4$  for multi-parton jets. CMS software uses the FastJet library [26] for efficient implementations of sequential jet clustering. In particular, FastJet reduces the  $\mathcal{O}(N^2)$  computation of Equation 1.11 to  $\mathcal{O}(N \log N)$ , where  $N \sim \mathcal{O}(10^2 - 10^3)$ .

## Jet calibration

While individual PF candidates receive some level of energy calibration from the detector subsystems, it is still necessary to calibrate each jet as a whole [27]. There are 3 steps in the additional calibration:

1. A pileup correction (L1) is derived from MC truth information and randomly-triggered data events. It subtracts energy from jets, as a function of the jet  $p_T$ ,  $\eta$ ,  $A$  (area), and  $\rho$ .  $A$  is computed by adding a uniform distribution of infinitesimally soft particles (“ghosts”) to the event prior to the jet clustering, and counting how many are clustered with each jet. A wide jet will have many such particles and will be more susceptible to pileup contamination. The event-wide quantity  $\rho$  is a measure of the median  $p_T$  per unit area in the event, also computed using ghosts.
2. The L2L3<sup>5</sup> correction accounts for detector response biases by using the GEANT4 simulation of the detector (see Section 1.4). Two jet collections are computed in the simulation: PF jets (using PF candidates) and truth jets (using the particle collection from the hard scattering and showering simulation). The energy scale of PF jets is corrected (as a function of  $p_T, \eta$ ) to match that of truth jets.
3. The L2L3 residual correction is only applied to data and corrects for small differences between the real detector and the simulation used in MC. It is derived using  $Z(\rightarrow \ell\ell) + \text{jet}$ ,  $\gamma + \text{jet}$ , and dijet events. In each case, one well-measured object ( $Z$ ,  $\gamma$ , jet) is used to calibrate the recoiling jet. While the  $Z$  events provide the best reference object ( $\mu$  energy uncertainties are very small), the small cross section introduces large statistical uncertainties at high  $p_T$  and  $\eta$ .

The  $p_T$  dependence of the corrections and the associated uncertainties are shown in Figure 1.14.

---

<sup>5</sup>The terminology is historical, referring to previously-factorized corrections L2 and L3

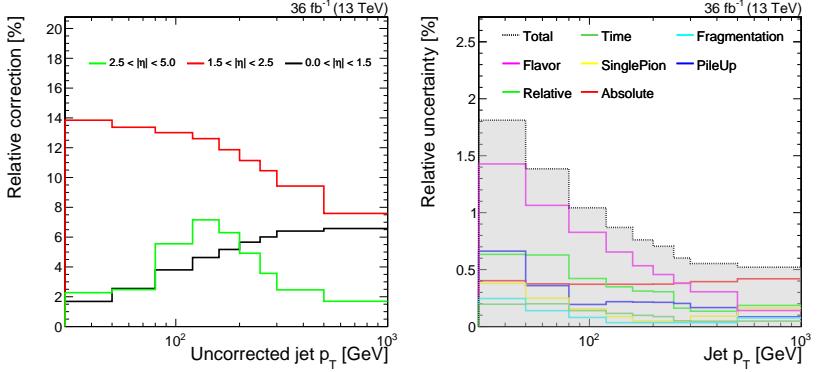


Figure 1.14: Left: the total effect of the L1 and L2L3 jet energy scale corrections in MC, as a function of  $p_T, \eta$ . Right: the corresponding uncertainties on the JES corrections, in the same  $p_T$  binning and averaged over  $\eta$ .

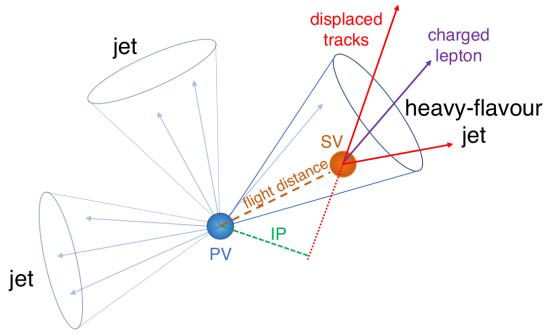


Figure 1.15: Location of a secondary vertex inside a  $b$  jet. Reprinted from Reference [11].

## Heavy flavor jet ID

It is frequently useful to identify the type of parton that induces a jet. In the context of the results presented in this thesis, identifying  $b$  quarks ( $b$ -tagging) is important, but identifying other initial states is also possible ( $c$  quarks, gluons vs quarks). The hadronization of a  $b$  quark involves the production of a  $B$  meson, which has a lifetime of  $\sim 1.5$  ps. This translates into a lab frame displacement of several millimeters, which is well within the vertexing resolution of the pixel detector. Therefore, the signature of a  $b$  jet is the identification of a “secondary vertex” (SV) displaced from the PV, with properties consistent with a heavy hadron (Figure 1.15).

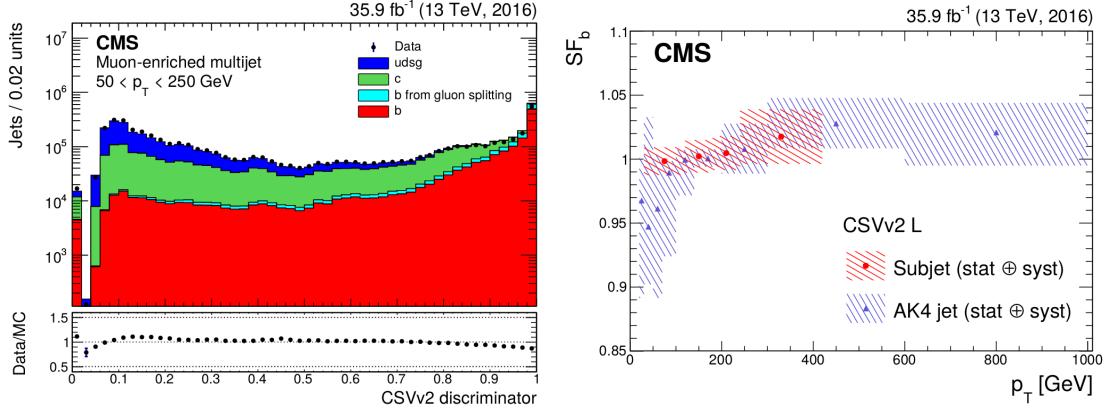


Figure 1.16: Left: The distribution of the CSVv2 response of AK4 jets in a dijet sample enriched with  $g \rightarrow b\bar{b}$  events. Right: data/MC scale factors to correct for discrepancies in the simulation of the  $b$  jet ID. AK4 jet SFs and uncertainties are shown in blue; overlaid in red are the SFs for subjets (see Chapter 2). Reprinted from Reference [11].

The reconstruction of the SV is described in Section 1.2.1. The Combined Secondary Vertex (CSVv2) tagger [11] is an artificial neural network trained to distinguish  $b$  jets from  $u/d/c/s/g$  jets. Nineteen characteristics of the jet (e.g. mass of tracks associated with the SV, presence of soft leptons from the SV,  $d_0$  of the SV) are used to train the NN<sup>6</sup>. For the analyses described in this thesis, a jet is considered  $b$ -tagged if it satisfies  $\text{CSVv2} > 0.54$ . Discrepancies between data and MC in the efficiency of the  $b$  jet ID are corrected for using SFs, derived separately for  $b$  and non- $b$  jets. Two  $b$ -enriched samples are used for the former task. The first is a relatively pure, but statistically limited, sample of  $t\bar{t}$  events, in which the  $bs$  from the  $t$  decays are used. The second is a much more statistically powerful sample of  $g \rightarrow b\bar{b}$  events, which is selected by triggering on and identifying a non-isolated muon from the decay of one of the  $B$  mesons. Figure 1.16 compares the CSVv2 response in data and simulation, as well as the SFs used to correct this distribution.

<sup>6</sup>See Section 4.1.2.1 of Reference [11] for a full list.

### 1.3.5 Hadronic taus

A hadronic tau ( $\tau_h$ ) is defined as a tau lepton which decays to one  $\nu_\tau$  and one or more hadrons. A dedicated algorithm [16, 28] is used to extract  $\tau_h$  candidates from the AK4 jet collection. Only jets within the tracker acceptance and with  $p_T > 13$  GeV are used. Combinations of particles within the jet are considered for compatibility with one of the relevant decay modes: a tau neutrino plus  $h^-$ ,  $h^-\pi^0$ ,  $h^-\pi^0\pi^0$ , or  $h^-h^+h^-$ . Here,  $h^\pm$  refers generically to some charged hadron. While other hadronic final states exist, they have a fairly small branching ratio and are not considered. The  $\pi^0$  will present in the jet as two photons; in some cases, the photon may convert into an  $e^-e^+$  pair. In the case of multi-hadron final states, compatibility with intermediate resonance masses is checked. The highest  $p_T \tau_h$  candidate in the jet is selected as a PF  $\tau_h$ . To reduce the large combinatorial background from quark/gluon jets, stringent requirements are made of the PF isolation. A true  $\tau_h$  should be fairly well-isolated, whereas a combinatorial fake will sit in the middle of a parton shower.

### 1.3.6 Missing momentum

The PF missing momentum ( $p_T^{\text{miss}}$ ) is defined as the magnitude of the missing momentum vector:

$$\vec{p}_T^{\text{miss}} = - \sum_{i \in \text{PF cands.}} \begin{pmatrix} p_x^{(i)} \\ p_y^{(i)} \end{pmatrix} \quad (1.12)$$

In a well-balanced event (without reconstruction error, neutrinos, or other non-interacting particles), the vectorial sum will cancel out perfectly, resulting in  $p_T^{\text{miss}} \sim 0$ . For the DM searches described in this thesis, we will be interested in  $p_T^{\text{miss}} \gg 0$ .

Large anomalous  $p_T^{\text{miss}}$  can be reconstructed if detector noise or algorithm failures induce large, localized energy deposits in the event description [29]. Such events are removed through a series of filters:

- HCAL and ECAL filters, identifying events with calorimeter clusters caused by noise. The shape and timing of the energy distribution can be used to identify noise, as well as detector-specific information (such as known problematic ECAL

crystals).

- Beam halo filter, identifying energy deposits from muons traveling parallel to the beam. These muons are produced from pre-collision interactions between the beam and the machine. They are identified by their localization in  $\phi$  and longitudinal signature in the ECAL/CSCs.
- Reconstruction filters, identifying failures of the PF algorithm to properly reconstruct particles. In some cases, muons can be double-counted or mis-identified as charged hadrons and muons.

These filters remove effectively all anomalous  $p_T^{\text{miss}}$  events while rejecting less than 1% of events with real  $p_T^{\text{miss}}$ .

As the  $p_T^{\text{miss}}$  indirectly depends on the momentum of each PF candidate that enters the sum, any *ad hoc* calibrations must be propagated. Most energy calibrations are intrinsic to the PF algorithm, but the jet energy scale corrections are not. Therefore, the missing momentum is accordingly updated:

$$\vec{p}_T^{\text{miss}} \mapsto \vec{p}_T^{\text{miss}} + \sum_{j \in \text{jets}} (\vec{p}_T^{j,\text{corr.}} - \vec{p}_T^{j,\text{raw}}) \quad (1.13)$$

### 1.3.7 Pileup mitigation

Two algorithms are used to mitigate the effects of pileup in jets: charged hadron subtraction (CHS) and pileup per-particle identification (PUPPI) [30]. CHS simply removes charged particles not from the primary vertex from the set of PF candidates fed into the jet clustering. This is only able to remove charged pileup contamination for jets within the tracker acceptance.

The PUPPI algorithm defines a local shape  $\alpha_i$  for every particle  $i$  in the event that is independent of particle location or charge. The distribution of  $\alpha_i$  is determined for central, charged particles from the PV and from PU. This is then extrapolated to neutral particles and forward particles, to assign a probability  $P(\text{PU}|\alpha_i)$ . More

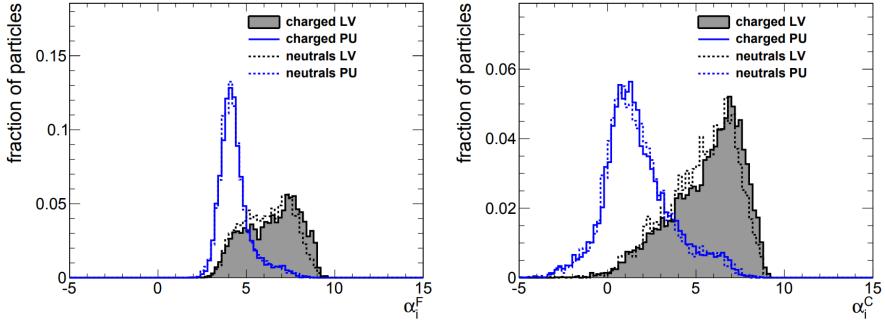


Figure 1.17: Distribution of PUPPI local shapes for charged and neutral particles from the PV and PU vertices. Reprinted from Reference [30].

concretely, the local shape is defined:

$$\alpha_i = \log \sum_{j \neq i} \frac{p_{T,j}}{\Delta R_{ij}} H(\Delta R_{ij} - R_{\min}) H(R_0 - \Delta R_{ij}) \quad (1.14)$$

where  $R_{\min}, R_0$  are tunable parameters and  $H$  is the Heaviside step function. Up to rescaling,  $\alpha_i$  is the sum of  $p_{\text{TS}}$  of particles in an annulus around particle  $i$ . It is expected to be larger for PV particles than PU particles, as PU radiation is uniformly distributed, whereas PV radiation is centered around hard partons. This is illustrated in Figure 1.17, which shows  $\alpha_i^F$  (same as  $\alpha_i$ ) and  $\alpha_i^C$  (same as  $\alpha_i$ , except sum only charged particles).

Then, we define:

$$x_i = H(\alpha_i - \bar{\alpha}) \frac{(\alpha_i - \bar{\alpha})^2}{\sigma^2} \quad (1.15)$$

where  $\bar{\alpha}$  and  $\sigma$  are the median and RMS, respectively, of the charged PU  $\alpha_i$  distribution, with some extrapolation for  $\eta_i$  and  $q_i$ . It is found that  $\alpha_i$  has a Gaussian distribution for pileup, and so  $x_i$  should have a  $\chi^2$  distribution with 1 degree of

freedom. Therefore, we define the probability of pileup as:

$$w_i = 1 - P(\text{PU}|\alpha_i) = \begin{cases} 1 & \text{if } |\eta| < 2.5, q_i \neq 0, \text{ from PV} \\ 0 & \text{if } |\eta| < 2.5, q_i \neq 0, \text{ not from PV} \\ P(\chi^2 < x_i | N_{\text{dof}} = 1) & \text{otherwise} \end{cases} \quad (1.16)$$

The four momenta of particles are scaled by the PUPPI probability, i.e.  $p_i^\mu \mapsto w_i p_i^\mu$ .

In the context of the results described in this analysis, it is found that CHS provides sufficient pileup mitigation for AK4 jets, but as described in Chapter 2, PUPPI will be necessary for CA15 jets.

## 1.4 Simulation of CMS

The CMS offline software suite (CMSSW [31]) uses Geant4 [32, 33] to simulate the detector response to particles produced in collisions. As described in Section ??, particle-level MC simulates the hard scattering and parton shower. Multiple simulated proton-proton collisions are overlaid into a single event to mimic the effect of in-time pileup. The generated particles are then interfaced to Geant4, which simulates the passage of a particle through the magnetic field, the energy deposited as the particle interacts with the detector material, and the evolution of any additional particles produced (e.g. EM showers). The readout electronics' response to the detector signature is then simulated. These signals are finally fed into the same software used to reconstruct real collision data, minimizing the differences between simulation and data.

# Chapter 2

## Hadronic Resonance Identification

In this chapter, we describe the reconstruction and identification of heavy ( $\gtrsim 100$  GeV) resonances that decay to two or more quarks. Within the Standard Model, the only such resonances are the massive vector bosons ( $W, Z \rightarrow q\bar{q}'$ ), the Higgs boson (typically  $H \rightarrow b\bar{b}$ ), and the top quark ( $t \rightarrow bW(\rightarrow q\bar{q}')$ ). These quarks hadronize into jets (described in Chapter ??), which are typically reconstructed at the LHC using the anti- $k_T$  algorithm (described in Chapter 1.3.4). The focus of this chapter is on the cases in which the resonance is boosted and the decay products merge, such that they cannot be identified as 2 or 3 distinct jets. In preparation for Chapter 3, we will take the top quark as a concrete example. The studies presented here can (and in some cases have been) applied to other heavy resonances, both within and beyond the Standard Model.

### 2.1 Reconstruction

The approximate angular separation between the quarks from a heavy resonance decay is[?]:

$$\Delta R \sim \frac{2M}{p_T} \tag{2.1}$$

where  $M$  is the resonance mass and  $p_T$  is the resonance transverse momentum. Setting  $M = m_t$  and  $\Delta R = 1.2$  (i.e. the radius at which three  $R = 0.4$  jets start to overlap), we

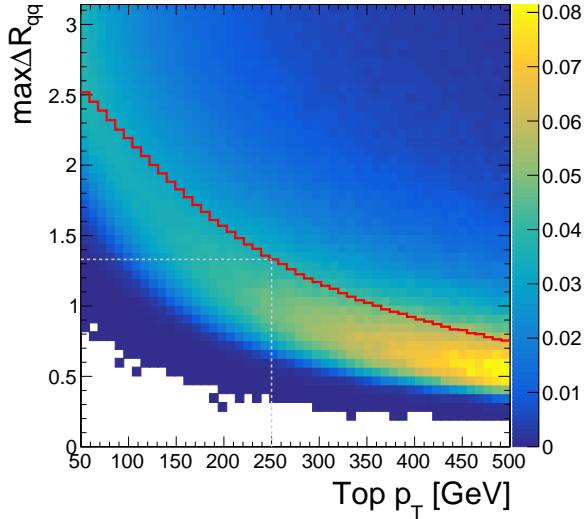


Figure 2.1: Distribution of top quark momenta versus decay radii in a simulated top quark pair sample. The events are weighted such that the inclusive momentum distribution is uniform. The  $z$ -axis units are arbitrary, but proportional to the distribution of jets. The solid red line marks the 50% quantile of jets at each value of  $p_T$ .

extract a “merging scale” of 300 GeV. This is be verified by checking the distribution of the “decay radius” in top quark simulation. Here, we define decay radius as:

$$\max \Delta R_{qq} \equiv \max_{0 \leq i < j \leq 2} \{\Delta R(q_i, q_j)\}, \text{ where } t \rightarrow q_0 q_1 q_2 \quad (2.2)$$

Using a broad spectrum of generated top quark  $p_T$ , Figure 2.1 shows the dependence of the decay radius on the top quark  $p_T$ , where we restrict the resonance to satisfy  $|\eta| < 2.5$ . If we are interested in top quarks with  $p_T > 250$  GeV (motivated by the trigger selection (Section 3.1)), then over half of top quarks will be fully contained within a jet of radius 1.5. That is, at  $p_T \approx 250$  GeV, it is equally likely that a top quark’s decay products will fall within a single large-radius jet or that they will be resolvable as three separate jets. However, past this threshold momentum, the large-radius jet becomes the preferred reconstruction option. This motivates the use of  $R = 1.5$  jets to reconstruct hadronic top quarks with  $p_T > 250$  GeV.

There are two tunable parameters in jet reconstruction. We have specified the jet radius, but we must also choose the jet algorithm. The anti- $k_T$  algorithm tends

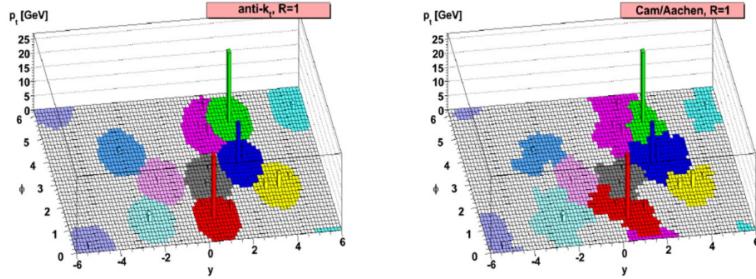


Figure 2.2: Jets clustered using the anti- $k_T$  (left) and CA (right) algorithms. Shown is the  $y$ - $\phi$  plane of a hypothetical calorimeter, unrolled onto a flat surface. The height of each cell represents the  $p_T$  of the particle. The anti- $k_T$  jets tend to be more circular when compared to the CA jets. Figures are adapted from [?].

to pick circular jets, whereas the Cambridge-Aachen (CA) algorithm allows for more geometric shapes (Figure 2.2). As the top jets we seek to reconstruct are the sum of three light quark jets, we do not necessarily expect the  $R = 1.5$  jet to be circular. Figure 2.3 compares the jet mass distribution for top and light quark/gluon (LQG) jets, where the jets are clustered using both algorithms. CA produces a top jet mass distribution with a narrower peak that sits closer to  $m_t$  than anti- $k_T$ . Because of this, and the general improvement in  $S/B$  near the top mass peak, we choose the CA algorithm. Hereafter, we will refer to Cambridge-Aachen  $R = 1.5$  jets as CA15 jets.

The distance parameter of  $R = 1.5$  corresponds approximately to a maximal azimuthal angle separation of  $\pi/2$ , which can cover half of the detector’s fiducial volume. As the jet is so large, particles from pile-up interactions may accidentally be clustered into a jet from the primary vertex. Fundamental quantities (like top quark momentum) are uncorrelated with the number of primary vertices ( $N_{PV}$ ), but reconstructed quantities acquire such a dependence due to the extra radiation. These additional particles bias the energy scale of the jet (e.g. the mass) as well as geometric observables (described in Section 2.2). To mitigate these effects, we scale the particles’ 4-momenta by their corresponding PUPPI scores (described in Chapter 1) prior to clustering the jet. Jets clustered using all particles (without PUPPI filtering) have a jet mass and  $\tau_{32}^{\text{SD}}$  distributions (Figure 2.4) in which both the mean and variance have an  $N_{PV}$ -dependence. Adding PUPPI stabilizes the mean and ensures that the

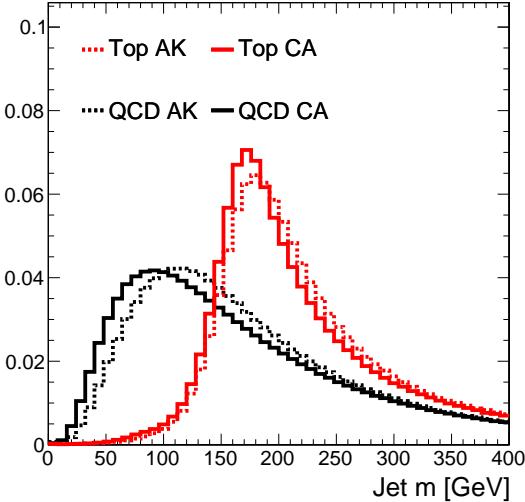


Figure 2.3: Mass distribution for jets clustered using the anti- $k_T$  (dashed) and CA (solid) algorithms. QCD refers to jets originating in QCD multijet events, i.e. from the hadronization of light quarks or gluons.

variance does not grow at large  $N_{\text{PV}}$ .

## 2.2 Identification

Having *reconstructed* the candidate top quark jets, we turn to the problem of *identifying* which CA15 jets originate from top quarks as opposed to light  $q/g$  hadronization. As indicated in Figure 2.3, the jet mass is a powerful observable, but top (LQG) jets do not necessarily have a mass of  $m_t$  ( $m_q, m_g \sim 0$ ). While some of this discrepancy is caused by mismeasurement of the jet energy scale (Chapter 1), a substantial fraction originates from extra radiation being absorbed into the jet. These extra particles arise from pile-up (although this is accounted for by PUPPI), initial state radiation, and underlying event. Many algorithms exist to “groom” such particles from a jet after it has been clustered; here, we will discuss and use the soft drop (SD) method [?]. SD functions by traversing the CA clustering history (a binary tree) in reverse and removing subjets (i.e. branches) of the clustering tree that are deemed to be too soft or wide-angled. More formally, at each node in the clustering tree, the softer subjet

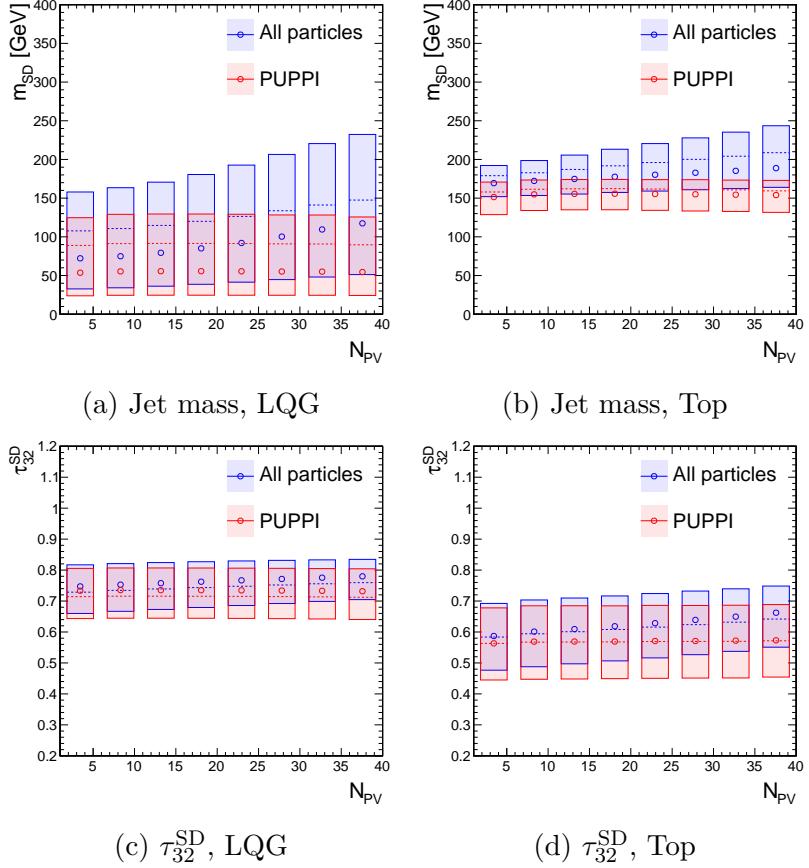


Figure 2.4: Stability of two CA15 jet observables (described in Section 2.2) as a function of  $N_{\text{PV}}$ . The median (mean) of each  $N_{\text{PV}}$  bin is represented by an open circle (dashed line), while the [25%, 75%] percentile range is shown with a box.

of the node will be removed if it satisfies the condition:

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} < \left(\frac{\Delta R_{12}}{R}\right)^\beta \quad (2.3)$$

where  $p_{T,i}$  refers to the  $p_T$  of the  $i$ -th subjet of the node;  $\Delta R_{12}$  is the  $\Delta R$ -distance between the two subjet; and  $R$  and  $\beta$  are tunable parameters. This process starts at the root node of the clustering tree (i.e. the whole jet) and proceeds iteratively to the leaves (i.e. individual particles). This condition is satisfied if the two subjets are very far apart (assuming  $\beta \geq 0$ ) or if the splitting is very asymmetric in momentum. We define the “SD subjets” (or where clear, simply “subjets”) of a jet to be the two branches of the root node, after branches failing the SD condition have been removed. The particles remaining after this grooming procedure are combined to make the “groomed” or SD jet.

We then define  $m_{SD}$  as the mass of the SD jet. Observables may also be defined in terms of the groomed or ungroomed jet. Figure 2.5 compares the ungroomed and groomed mass distributions in top and LQG jets, as a function of jet momentum. It is immediately clear that grooming provides (a) a sharper mass peak in top jets at  $m_t$  and (b) a smoothly falling mass distribution in LQG jets that goes to 0. Furthermore, SD ensures the stability of the mass distribution as a function of jet  $p_T$ , especially in LQG jets. For these reasons,  $m_{SD}$  will be our standard definition of jet mass.

### 2.2.1 Substructure

A substructure observable is any function of a jet’s constituents that is sensitive to the multi-pronged structure of a heavy resonance decay. In addition to jet mass and  $b$ -tagging, substructure is used to reject LQG jets as top decay candidates.

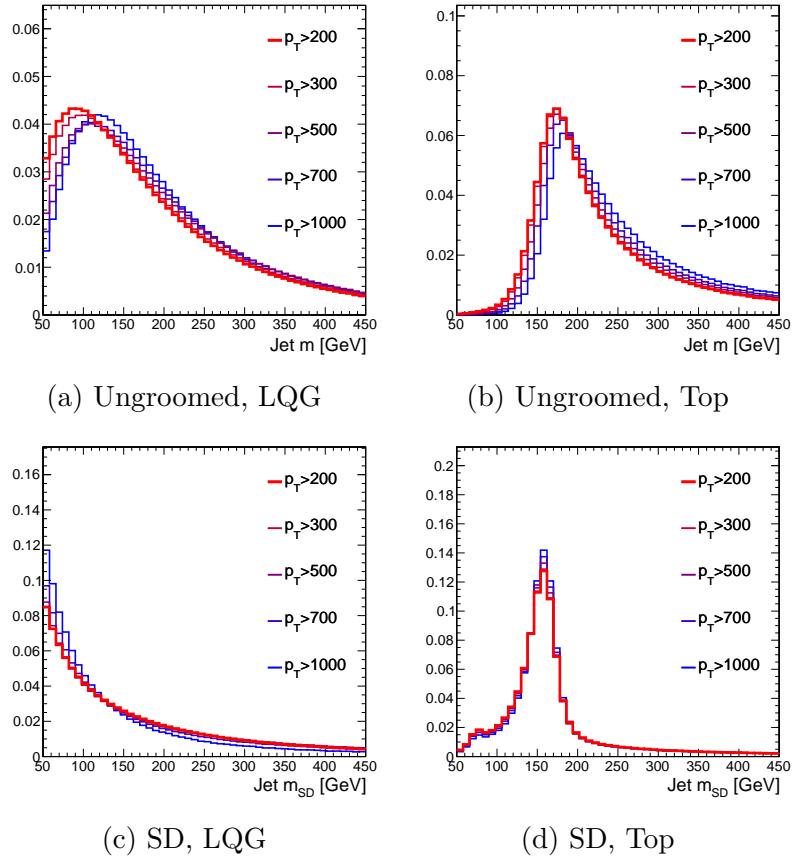


Figure 2.5: Distribution of ungroomed and groomed jet mass in CA15 jets originating from LQG or hadronic top decays. The multiple histograms represent increasingly stringent  $p_T$  requirements on the parton that initiates the jet.

## *N*-subjettiness

The *N*-subjettiness ( $\tau_N$ ) is a measure of the compatibility of a jet with an *N*-axis hypothesis [?]. It is defined as:

$$\tau_N = \frac{\sum_{i \in \text{jet}} p_{T,i} \min\{\Delta R_{ia} | a \in A\}}{\sum_{i \in \text{jet}} p_{T,i} R} \quad (2.4)$$

where  $R = 1.5$  (the jet radius);  $\Delta R_{ia}$  is the  $\Delta R$  distance between the particle  $i$  and the axis  $a$ ; and  $A$  is a set of  $N$  axes. Ideally,  $A$  would be defined to be the set of axes that minimize  $\tau_N$  for each jet, but this minimization problem is computationally difficult. Instead, the exclusive  $k_T$  algorithm is used to partition the jet's constituents into  $N$  subjets (NB: these are not the SD subjets discussed above). Since the  $k_T$  distance metric is proportional to  $\Delta R^2/R^2$ , this approximates the ideal minimization. The set of axes  $A$  is taken to be the directions of the  $N$   $k_T$  subjets. A small  $\tau_N$  indicates a high degree of compatibility with the  $N$ -axis hypothesis. Therefore, we expect a 3-pronged (e.g. top) jet to satisfy  $\tau_3 \ll \tau_2$ , whereas a 1-pronged (e.g. LQG) jet should satisfy  $\tau_3 \lesssim \tau_2$  (for optimal choice of  $A$ ,  $N > M \Rightarrow \tau_N \leq \tau_M$  for any jet). Correspondingly, we take  $\tau_{32} \equiv \tau_3/\tau_2$  to be the tagging observable.

Figure 2.6 shows the distribution of  $\tau_{32}$ . As with jet mass, we may calculate  $\tau_{32}$  either on the whole jet or on the groomed (SD) jet. The discrimination between top and LQG jets is similar in both cases, but as Figure 2.6c demonstrates,  $\tau_{32}^{\text{SD}}$  has the weaker correlation with  $m_{\text{SD}}$  in LQG jets. This feature will be critical to validate any tagger in data, as described in Section 2.3.

## HEPTopTagger

The HEPTopTagger algorithm de-clusters the jet into many subjets and attempts to reconstruct the  $W$  and  $t$  decay products out of these subjets [?]. The computation of the tagging variable  $f_{\text{rec}}$  can be simplified into three steps (a more detailed description is found in the appendix of Reference [?]):

1. Compute subjets of the CA15 jet. This is done in a fashion similar to the SD subjets discussed above, but instead of taking the two subjets of the root node,

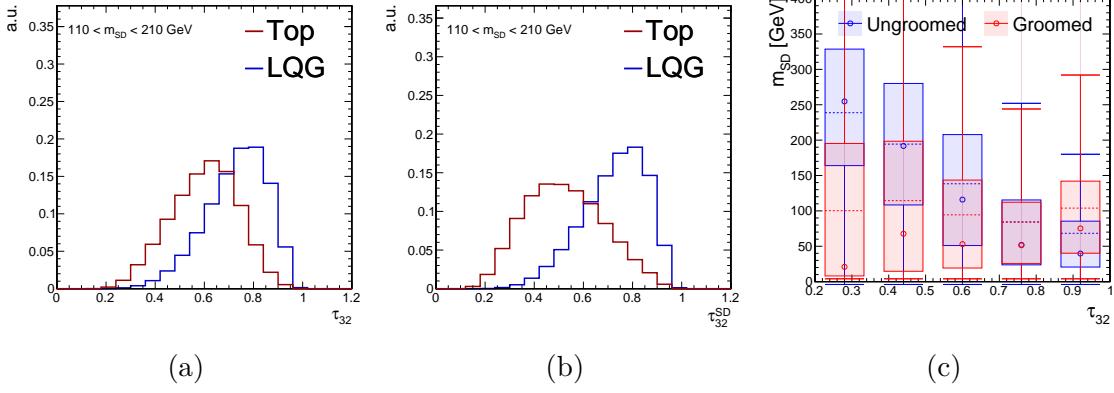


Figure 2.6: Shape of ungroomed (left) and groomed (center)  $\tau_{32}$  distributions in top and LQG jets, with a mass selection consistent with  $m_t$ . Right: the correlation between  $\tau_{32}$  and  $m_{SD}$  in LQG jets, comparing groomed and ungroomed jets.

the tree is traversed until some lower  $p_T$  bound is crossed.

2. Test all triplet combinations of the found subjets and define the  $m_{123}$  as the groomed mass of the trijet system.
3. Choosing the triplet most consistent with a 3-body top decay (see Equation 12 in Reference [?]), define:

$$f_{\text{rec}} = \min_{0 \leq i < j \leq 2} \left| \frac{m_{ij}/m_{123}}{m_W/m_t} - 1 \right| \quad (2.5)$$

where the indices  $i, j$  index elements of the selected triplet.

Figure 2.7 shows the distribution of the selected  $m_{123}$  and  $f_{\text{rec}}$ , although we will only use the latter as a tagging observable. Note that we do not define distinct groomed and ungroomed versions of these observables, as grooming is intrinsically used in defining the subjets.

## Energy Correlation Functions

Energy correlation functions measure the correlation of the positions of hard particles in a jet [?]. Heuristically, an  $N$ -point ECF is small if the hard particles can be grouped into fewer than  $N$  prongs and large if they arise from  $N$  or more prongs. An  $N$ -point

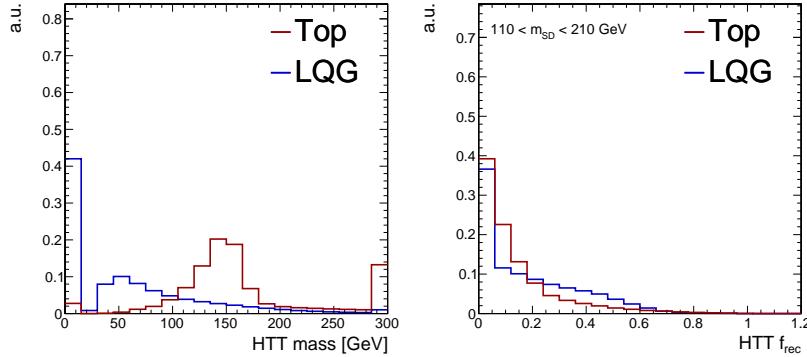


Figure 2.7: Shape of the  $m_{123}$  and  $f_{\text{rec}}$  observables computed by the HEPTopTagger algorithm.

ECF, with angular parameters  $o$  and  $\beta$ , is defined as:

$$e(o, N, \beta) \equiv {}_o e_N^\beta = \sum_{K \subset J, |K|=N} \left[ \prod_{i \in K} \frac{p_T^{(i)}}{p_T^{(J)}} \right] \times \min \left\{ \prod_{i,j \in P} \Delta R_{ij}^\beta \mid P \subset (K^2 \setminus (k, k)), |P| = o \right\} \quad (2.6)$$

where  $K^2 \setminus (k, k)$  indicates all pairs of distinct particles in  $K$ . The proposed tagger in Reference [?] is:

$$N_3^{(\beta)} = \frac{e(2, 4, \beta)}{(e(1, 3, \beta))^2} \quad (2.7)$$

Figure 2.8 shows  $N_3$  for various values of  $\beta$ ; given our desire for stability as a function of jet  $p_T$  and mass, we only consider ECFs computed on the SD jet. The discrimination power of this ECF ratio is roughly comparable to that of  $\tau_{32}^{\text{SD}}$ .  $N_3$  is motivated by the behavior of 3- and 4-point ECFs in top and LQG jets:

- In top jets,  $e(N=4) \ll e(N=3)$ , since 3-point correlation functions are large in a 3-pronged jet
- In QCD jets,  $e(N=3) \sim e(N=4)$ , since both 3- and 4-point ECFs are weak in a 1-pronged jet

Therefore, taking the ratio  $e(N=4)/e(N=3)$  constructs a useful observable.

While  $N_3$  has a strong theoretical motivation, it is possible that other functions of ECFs distinguish between top and LQG jets. In order to construct observables that

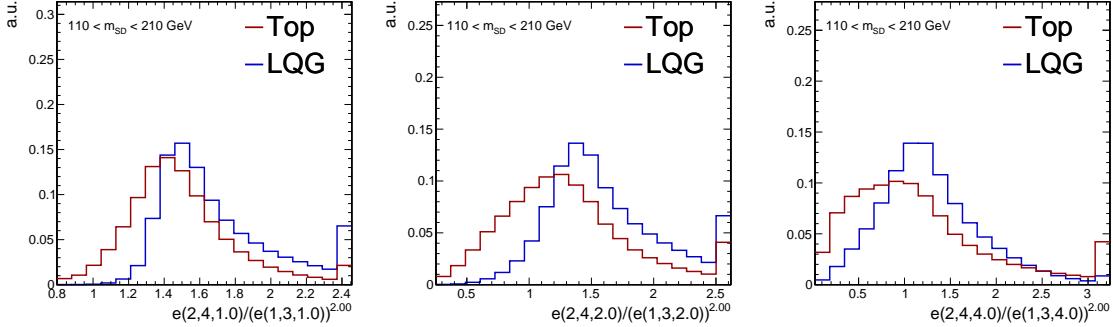


Figure 2.8: Shape of the  $N_3$  observables in top and LQG jets, for various values of  $\beta$ .

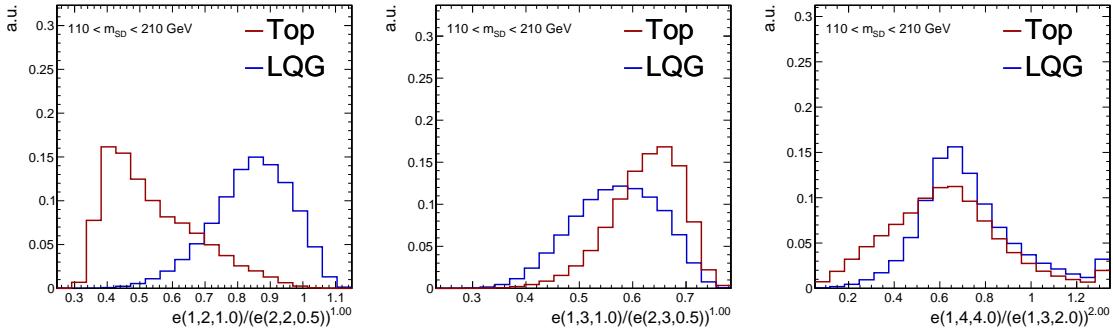


Figure 2.9: Examples of non-trivial ECF ratios other than  $N_3$  that separate top and LQG jet distributions.

do not have a strong dependence on the jet  $p_T$ , we restrict ourselves to ratios of the form:

$$\psi(a, N, \alpha, b, M, \beta) = \frac{e(a, N, \alpha)}{(e(b, M, \beta))^x}, \text{ where } M \leq N \text{ and } x = \frac{a\alpha}{b\beta} \quad (2.8)$$

A large subset of this broader class of ECF observables are found to be useful (Figure 2.9), including ratios not of the form  $e(N = 4)/e(N = 3)$ .

## 2.2.2 A combined tagger

In principle, we have constructed an infinitely large space of substructure observables.

In practice, we only consider a finite sampling of ECF parameters:

$$\begin{aligned} N &\in \{1, 2, 3, 4\} \\ o &\in \{1, 2, 3\} \\ \beta &\in \{0.5, 1, 2, 4\} \end{aligned} \tag{2.9}$$

This grid results in  $\sim 900$   $\psi$  observables.

To build a single optimal observable out of all the  $\{\psi_i\}$ s, we will use a boosted decision tree (BDT). A simplified algorithm to train a single decision tree node  $n$  is as follows:

1. Choose a  $\psi_j$ , either by sampling randomly or selecting the one most optimal for the next step.
2. Based on the training data fed to the node, select a decision boundary  $d_n$  to optimize a loss function. For example, one can use the cross-entropy loss:

$$\ell(X, y; j, d_n) = -\hat{\pi}_B \ln \hat{\pi}_B - \hat{\pi}_S \ln \hat{\pi}_S, \text{ where } \hat{\pi}_c = \hat{P}(y = c | \psi_j < d_n) \tag{2.10}$$

A tree is built iteratively:

1. Train a node  $n$  using the above criteria.
2. If a stopping condition is not met (i.e. maximum number of nodes, minimal improvement in  $\ell(n)$ ), train one node on the samples that pass  $n$  and another on the samples that fail.

Figure 2.10 provides a pictorial example of how a decision tree can be built.

While decision trees can very accurately describe training data provided sufficient complexity, they also pathologically overfit the data. To mitigate this, while retaining descriptive power, a standard method is to *boost* many simple trees [?]. The simplicity

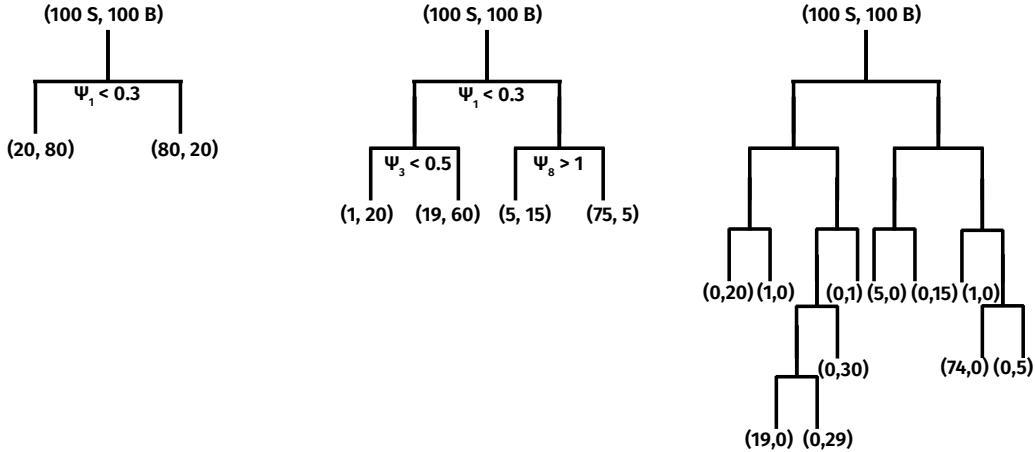


Figure 2.10: Steps in greedily training a simple decision tree.

of the tree prevents overfitting, while boosting many trees allows for a complex model. The result of a BDT is a classifier  $f_n(x) = \sum_{i=0}^n \nu^i T_i(x)$ , where  $\nu \leq 1$  is tunable and each  $T_i$  is a decision tree. A simplified algorithm to train a BDT is as follows:

1. Define a global loss function, e.g.:

$$L(y_i; f_i) = \ln(1 + \exp(-y_i f_i)) \quad (2.11)$$

2. Train a single tree  $T_0$  and initialize classifier  $f_0 = T_0$

3. Until some stopping condition (index  $m = 1, \dots$ ):

- 3.1. Compute the “residual”

$$r_{mi} = -\nabla_f L(y_i; f)|_{f=f_{m-1}(\psi_i)} \quad (2.12)$$

$$L = (y - f_m(\psi))^2 \Rightarrow r_m = y - f_m(\psi) \quad (2.13)$$

- 3.2. Fit a regression tree  $T_m$  to predict  $r_{mi}$  as a function of  $x_i$ :

$$\ell(X, r_m; j, d, \hat{r}) = \sum_{i|\psi_{ji} < d} (r_{mi} - \hat{r})^2 \quad (2.14)$$

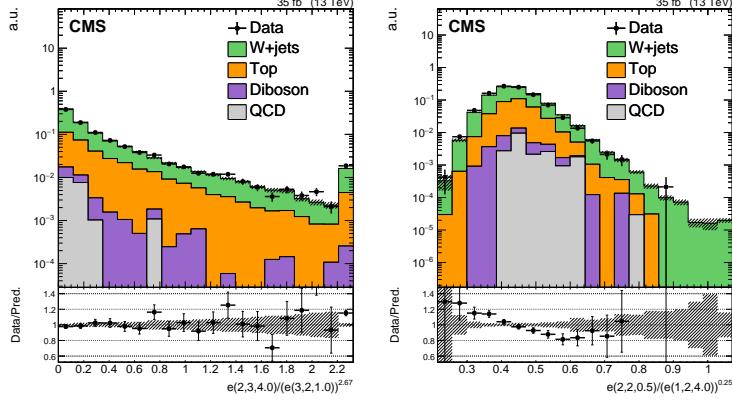


Figure 2.11: Two different ECF ratios in a  $W$ +jets selection, heavily enriched in LQG jets. One is fairly well-modeled, while the other is not.

### 3.3. Update $f_m = f_{m-1} + \nu T_m$

While we would like to train a BDT on the entire space of  $\{\psi\}$ , there are two issues to be solved: poorly modeled ratios and a large feature space. Firstly, not every ECF ratio is well-simulated by our MC (Figure 2.11). More systematically, we can compute the CDFs of each  $\psi$  and define a score:

$$-\log_{10} \text{KS}(F(\psi_i|\text{data}), F(\psi_i|\text{MC})) = -\log_{10} \max |F(\psi_i|\text{data}) - F(\psi_i|\text{MC})| \quad (2.15)$$

where  $F$  represents the CDF and KS denotes the Kolmogorov-Smirnov metric on probability distributions. The score is close to 0 for poorly-simulated distributions. Figure 2.12 parameterizes this as a function of  $N/M$  (ratio of the number of particles) and  $a\alpha/b\beta$  (ratio of the angular powers) and shows an interesting structure. It is found that  $3/2$  and  $4/2$  ratios are uniformly poorly modeled, as are ratios with large  $a\alpha/b\beta$ . While this structure is interesting and suggestive, we do not yet understand its origin, and therefore do not use this parameterization to select ECF ratios. Instead, we simply reject any ECF ratio for which  $-\log_{10} \text{KS} < 1$ .

Even after filtering poorly-modeled ratios, we are left with a sampled  $\psi$  grid of  $\sim 400$  points. By inspection, many of the  $\psi$ s are highly correlated or not useful at all. It is desirable to reduce the size of the feature space, as computing each ECF is somewhat

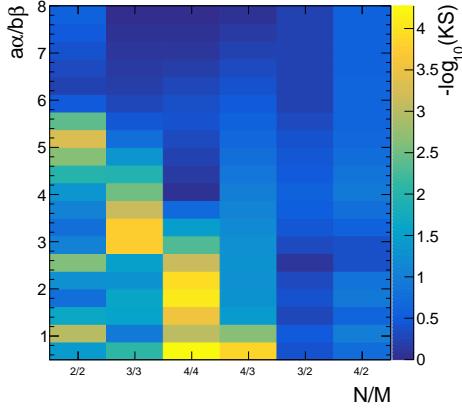


Figure 2.12: The  $-\log_{10}$  KS metric as a function of  $N/M$  and  $a\alpha/b\beta$ , computed using events enriched in LQG jets.

computationally-intensive: an  $N$ -point ECF on a  $p$ -particle jet has  $\binom{N}{p}$  terms. Note that standard pre-processing techniques, like principal component analysis, do not reduce the number of features to be computed. While L1 regularization does attempt such a dimensional reduction, it cannot be trivially applied to BDTs. Therefore, we introduce a targeted iterative training method to solve this problem:

1. Train a BDT with trees  $T_1, \dots, T_n$
2. For each  $\psi_i$ , define a score:

$$s_i = \sum_{m=1}^n \nu^{m-1} \sum_{\text{nodes using } \psi_i \text{ in } T_m} N_{\text{samples}}(\text{node}) \times (\ell(\text{node}) - \ell(\text{parent}))^2 \quad (2.16)$$

3. Remove (one or more)  $\psi_i$  with smallest  $s_i$  and repeat until the global loss  $L$  worsens significantly

Iterative training is expensive and can require the training of  $\mathcal{O}(50)$  BDTs. It is semi-parallelizable, and the entire process typically takes a few hours. However, as the inference samples are 1-2 orders of magnitude larger than the training samples, this method reduces the total CPU time needed to run an analysis. Figure 2.13 shows background acceptance rate at  $\epsilon_{\text{sig}} = 0.5$  (a proxy for the global loss) as a function of feature space size. For illustrative purposes, we only show the range [1, 50]. The

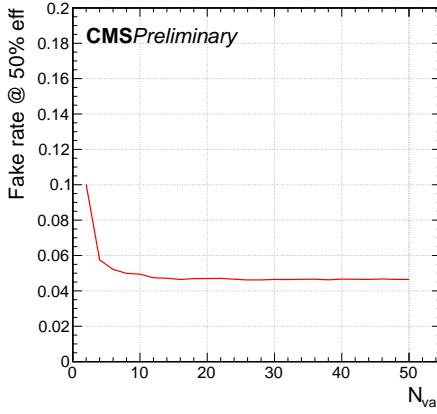


Figure 2.13: Performance of BDTs as a function of the number features used in the training.

inputs for this training are the ECF ratios, as well as  $\tau_{32}^{\text{SD}}$  and  $f_{\text{rec}}$ , which provide additional information. It is clear that beyond 13-15 features, there is little to be gained by adding additional information.

The features selected by this reduction and classification process are:

$$\begin{aligned}
 & \frac{e(1, 4, 20)}{e(1, 3, 10)^2}, \frac{e(1, 4, 40)}{e(1, 3, 20)^2}, \frac{e(2, 4, 05)}{e(1, 3, 05)^2}, \frac{e(2, 4, 10)}{e(1, 3, 10)^2}, \frac{e(2, 4, 10)}{e(2, 3, 05)^2}, \frac{e(2, 4, 20)}{e(1, 3, 20)^2} \\
 & \frac{e(1, 2, 20)}{e(1, 2, 10)^2}, \frac{e(1, 3, 40)}{e(2, 3, 20)}, \frac{e(3, 3, 10)}{e(1, 3, 40)^{3/4}}, \frac{e(3, 3, 10)}{e(2, 3, 20)^{3/4}}, \frac{e(3, 3, 20)}{e(3, 3, 40)^{1/2}} \\
 & \tau_{32}^{\text{SD}}, f_{\text{rec}}
 \end{aligned} \tag{2.17}$$

While a number of  $N_3$  or other 4/3 ratios appear in this list, we find a number of 2/2 and 3/3 ratios to contribute meaningfully to the classification task as well. Figure ?? shows the distributions of all selected features. Figure ?? shows the background acceptance as a function of signal efficiency, comparing the final BDT (“Combined BDT”) to several other taggers. The “11 ECF” BDT refers to a BDT trained using only the 11 ECF ratios from Equation 2.17, indicating that  $\tau_{32}^{\text{SD}}$  and  $f_{\text{rec}}$  are critical to achieving the same performance that can be reached with a much larger ECF ratio set (“50 ECF”). For comparison, also shown are the efficiency curves for  $\tau_{32}^{\text{SD}}$  and a BDT trained using  $\tau_{32}^{\text{SD}}$  and  $f_{\text{rec}}$  only. At fixed signal efficiency  $\epsilon_{\text{sig}} = 0.5$ , the combined

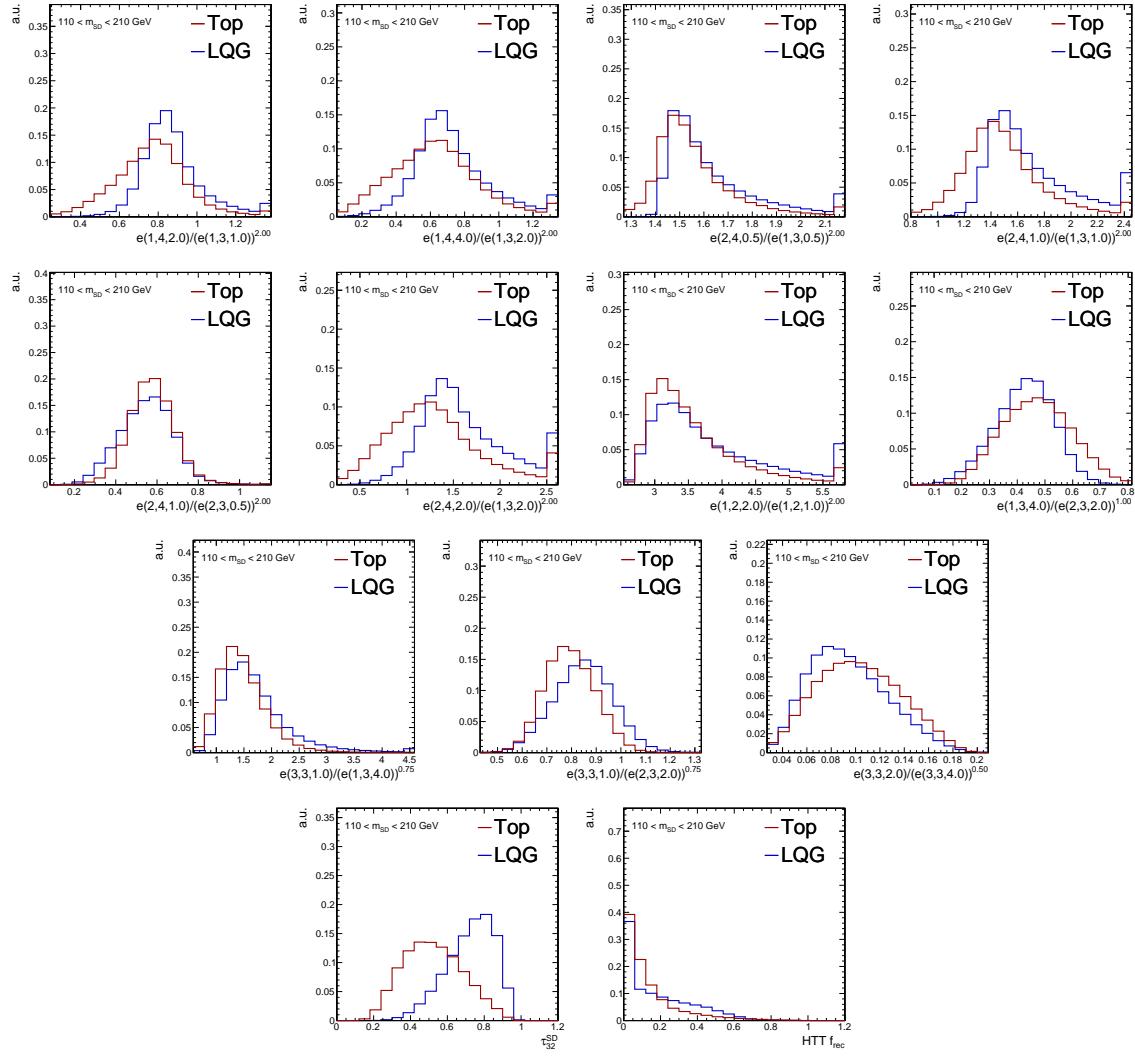


Figure 2.14: Distributions of the 13 features selected by the iterative BDT training

BDT reduces the background acceptance by 30% relative to  $\tau_{32}^{\text{SD}}$ , the standard top ID criterion at the LHC prior to this study.

## 2.3 Data validation

Prior to using the top BDT to identify top jets and reject LQG jets, we must verify that the simulation describes the BDT distribution properly as compared to data, and correct for any residual discrepancies. Figure 2.16 shows the BDT response and  $m_{\text{SD}}$  in top- and LQG-enriched selections. Top quarks are isolated by selecting events

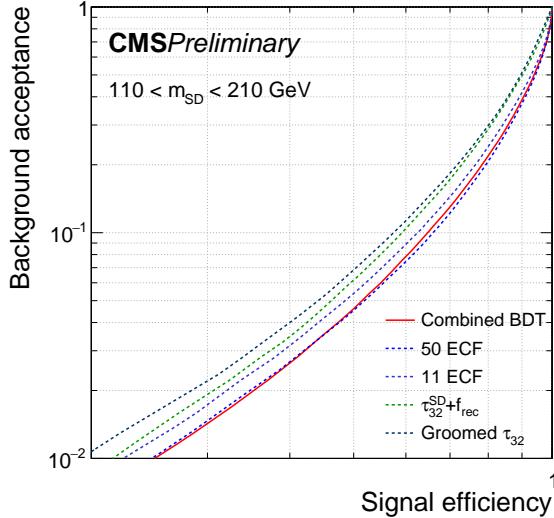


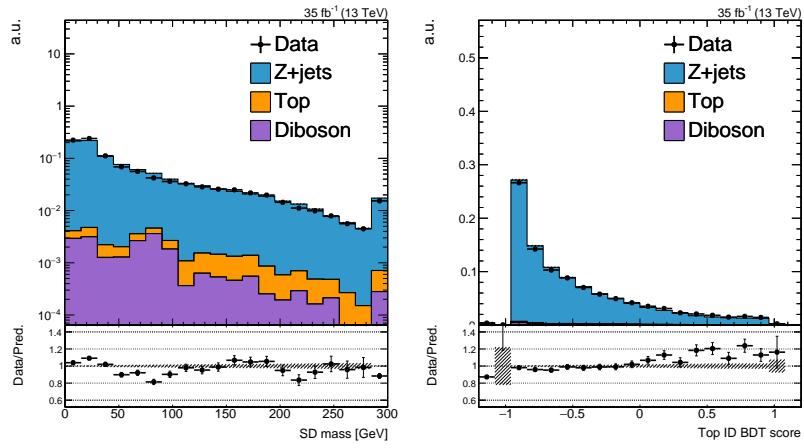
Figure 2.15: Receiver Operating Characteristic (ROC) curve comparing various top identification methods. The “Combined BDT” is the ID method chosen as the final tagger.

that produce  $t\bar{t}$  pairs, in which one top quark decays hadronically (the top jet) and the other decays muonically ( $t \rightarrow b\mu^+\nu_m u$ ). The leptonic  $t$  is selected by identifying the muon and  $b$  jet. We further require that the CA15 jet have  $110 < m_{\text{SD}} < 210$  GeV and at least one SD subjet to be  $b$ -tagged. LQG jets are selected by using  $Z(\rightarrow \mu\mu) + \text{jet}$  events. We require two opposite sign muons, with  $|m_{\mu\mu} - m_Z| < 30$  GeV; this selection selects a  $\gtrsim 95\%$  pure  $Z + \text{jet}$  sample. In both samples, we observe reasonably good agreement between data and simulation.

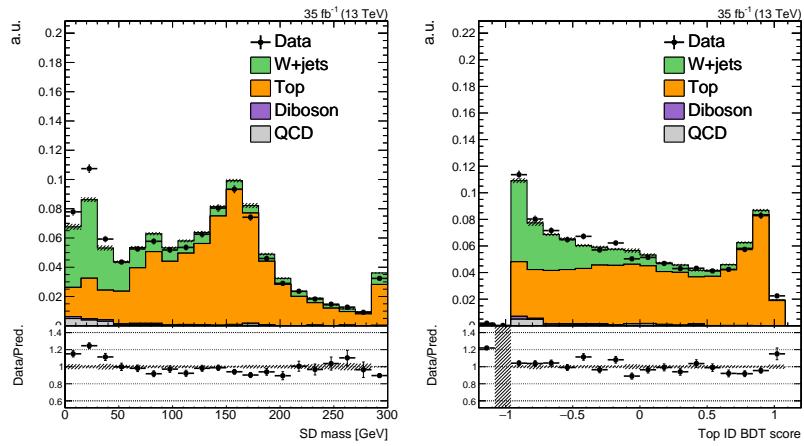
To account for any remaining differences, we define a scale factor (SF):

$$\text{SF}(x) = \frac{\epsilon_{\text{Data}}(\text{BDT} > x \text{ and } 110 < m_{\text{SD}} < 210)}{\epsilon_{\text{MC}}(\text{BDT} > x \text{ and } 110 < m_{\text{SD}} < 210)} \quad (2.18)$$

where  $x$  is a particular decision boundary and  $\epsilon$  is the fraction of data or MC events passing this BDT and mass selection. These are chosen to optimize sensitivity to the mono-top analysis, as described in Chapter 3. The SF is strongly dependent on the type of jet; in particular, we expect different SFs for top and LQG jets. In what follows, we will define two decision boundaries: a loose (0.1) and a tight (0.45)



(a) Dimuon selection



(b)  $t\bar{t}$  selection

Figure 2.16: Comparison of the BDT response and jet mass in data and simulation, in top and LQG jets.

category.

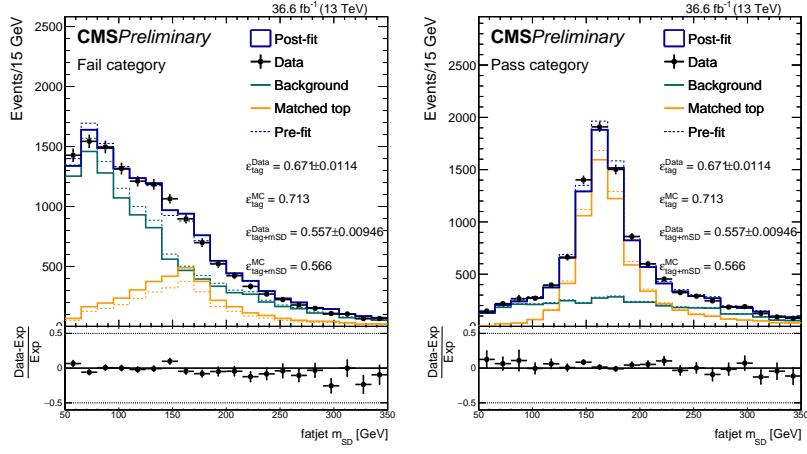
To compute  $SF_{LQG}$ , we use the dimuon selection in Figure 2.16, as this contains an essentially pure selection of LQG jets. Two sources of uncertainty are considered: the statistical uncertainties present in the data and MC, and the uncertainties on the theoretical prediction of the cross section of the small non-LQG backgrounds ( $t\bar{t}$  and diboson events). The measured SFs are:

$$\begin{aligned} SF_{LQG}(0.1) &= 1.02 \pm 0.05(\text{total}) \\ &\quad \pm 0.04(\text{statistical}) \pm 0.03(t\bar{t} + \text{diboson}) \\ SF_{LQG}(0.45) &= 0.97 \pm 0.07(\text{total}) \\ &\quad \pm 0.06(\text{statistical}) \pm 0.03(t\bar{t} + \text{diboson}) \end{aligned} \quad (2.19)$$

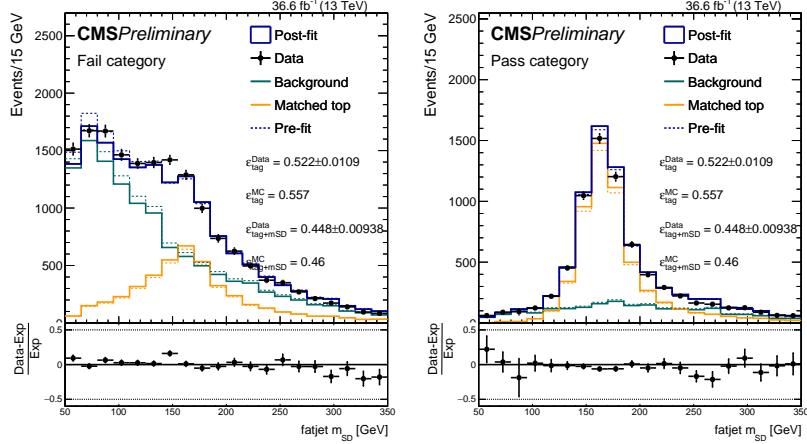
The process for top jets is complicated by the fact that the top pair selection in Figure 2.16 is not sufficiently pure in merged top jets. There is significant contamination from  $W+jets$  events. Furthermore, we cannot ensure that every  $t\bar{t}$  event selected produces a *merged* top jet - some events may contain jets in which only part of the top's decay products are clustered into the CA15 jet. Therefore, we extract the efficiency by means of a template fit to the mass distribution of passing and failing events, which can separate the top and LQG components in the selection. It is for this reason that we only use groomed observables in the BDT: grooming prevents a strong correlation between the observables and  $m_{SD}$ . Such a correlation would cause the mass distribution of passing LQG jets to be indistinguishable from that of passing top jets. Figure 2.17 show the fits in the passing and failing regions for both decision boundaries.

Several sources of uncertainty are considered for this measurement:

- Poisson uncertainties in the data and simulation
- CA15 jet energy scale and resolution
- Definition used to select “merged top” jets, allowing  $\max \Delta R_{qq'}$  to vary between 1 and 1.5 (nominal value is 1.2)



(a) Loose BDT-tagged



(b) Tight BDT-tagged

Figure 2.17: Fits to the  $m_{SD}$  distribution in a  $t\bar{t}$  sample to extract the efficiency in data of the BDT and mass selections. All uncertainties plotted and quoted are statistical in nature.

- Efficiency of selecting  $b$  jets

The resultant SFs and associated uncertainties are:

$$\begin{aligned}
 \text{SF}_{\text{top}}(0.1) &= 1.08 \pm 0.04(\text{total}) \\
 &\quad \pm 0.03(\text{statistical}) \pm 0.02(\text{JES} + \text{JER}) \pm 0.02(\text{merging}) \pm 0.002(b) \\
 \text{SF}_{\text{top}}(0.45) &= 1.07 \pm 0.06(\text{total}) \\
 &\quad \pm 0.03(\text{statistical}) \pm 0.02(\text{JES} + \text{JER}) \pm 0.014(\text{merging}) \pm 0.000(b)
 \end{aligned} \tag{2.20}$$

# Chapter 3

## The Search for $t + p_T^{\text{miss}}$

In this chapter, we discuss the search for dark matter produced in association with a single top quark (“mono-top”). Since the initial state of  $pp$  collisions do not contain any appreciable contribution from top quarks, any process that produces a single top quark must involve some flavor violation. In the Standard Model, any such process is heavily suppressed by off-diagonal elements of the CKM matrix. The SM production mechanism for the mono-top signature (Figure 3.1) involves a  $b$  quark in the final state, and thus does not couple the third generation with the first or second. True production of mono-top must introduce some such coupling as an extension to the SM, in addition to one (or more) invisible particle to serve as a DM candidate.

To illustrate how beyond-SM physics can produce this final state, we introduce two DM models: a flavor-changing neutral current  $V$  and a charged, colored scalar  $\phi$ . These models will also be used to benchmark the sensitivity of the analysis. However, it should be emphasized that the search is motivated and designed without reliance on any specific model; the assumption is that the mono-top final state alone is indicative of new physics, regardless of the specific production mechanism.

The FCNC  $V$  is assumed to couple to a fermionic DM candidate  $\chi$ . A partial Lagrangian of the interaction terms is given by:

$$\mathcal{L}_{\text{int}} = V_\mu \bar{\chi} \gamma^\mu (g_\chi^V + g_\chi^A \gamma_5) \chi + \bar{q}_u \gamma^\mu (g_u^V + g_u^A \gamma_5) q_u V_\mu + \bar{q}_d \gamma^\mu (g_d^V + g_d^A \gamma_5) q_d V_\mu + \text{h.c.}, \quad (3.1)$$

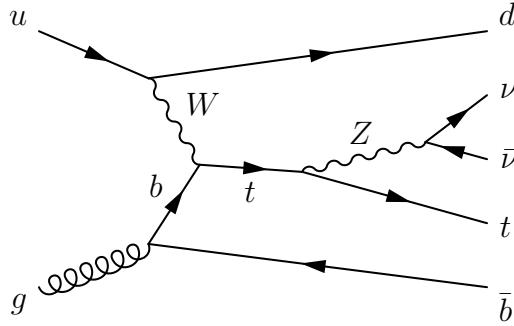


Figure 3.1: Production of mono-top in the SM, in which a top quark is produced in addition to a  $Z$  boson and bottom quark. The  $Z$  decays to neutrinos, providing large  $p_T^{\text{miss}}$ .

The model comes with 22 free parameters, broadly organized in three sets:

- The masses  $m_V$  and  $m_\chi$ . (2)
- The couplings  $g_\chi^V$  and  $g_\chi^A$ . These, respectively, control the strength of the vector and axial interactions between  $V$  and  $\chi$ . (2)
- The four coupling matrices  $g_q^X$ , where  $q = u, d$  and  $X = V, A$ . As before,  $X$  determines the type of spin-1 interaction. In principle, different coupling strengths can be permitted for up- and down-type quarks, so this indexed by  $q$ . Each  $g_q^X$  is a  $3 \times 3$  matrix, cross-coupling the three quark generations. To preserve  $SU(2)_L$  symmetry, we require  $g_u^V - g_u^A = g_d^V - g_d^A$ . ( $3 \times 6 = 18$ )

It is the  $g_{u,d}^{V,A}$  matrices that determine whether the model can produce mono-top, or mono-bottom, or mono-up, etc. If  $g_{u,d}$  is strongly diagonal (i.e. strongest couplings are within generations), then mono-light quark production will dominate, resulting in the mono-jet final state (Figure 3.2a). On the other hand, if we assume the only non-zero elements are those that couple the first and third generations, then mono-top production at the LHC is the best way to probe this model (Figure 3.2b). It is this latter choice that will be made in the rest of this chapter; other choices are best probed using a combination of multiple DM channels, which is left as future work. Furthermore, to respect  $SU(2)_L$  symmetry, we make the assumption that  $g_u^V = g_d^V$  and  $g_u^A = g_d^A$ .

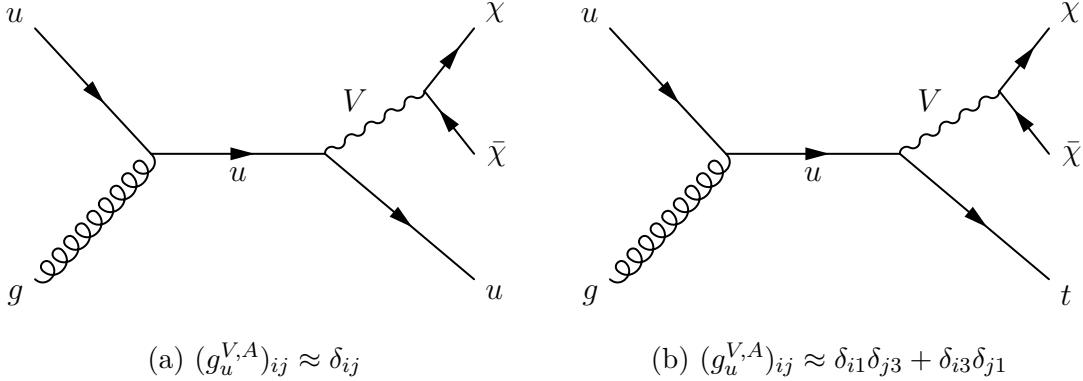


Figure 3.2: Possible DM production at the LHC, assuming a simplified spin-1 extension to the SM.

In the second benchmark model, the charged, colored scalar  $\phi$  couples to down-type quarks, or to a fermionic DM candidate  $\psi$  and a top quark. The interaction terms of the Lagrangian is given by:

$$\mathcal{L}_{\text{int}} = \phi \bar{d}_i^C [(a_q)^{ij} + (b_q)^{ij}\gamma^5] d_j + \phi \bar{t} [a_\psi + b_\psi \gamma^5] \psi + \text{h.c.} \quad (3.2)$$

There are 16 free parameters in this model, broadly organized in three categories:

- The masses  $m_\phi$  and  $m_\psi$ . (2)
- The couplings at the  $\phi \bar{t} \psi$  vertex  $a_\psi$  and  $b_\psi$ , which respectively control the strength of the scalar and pseudoscalar interactions. (2)
- The couplings at the  $\phi \bar{d}_i d_j$  vertex  $a_q^{ij}$  and  $b_q^{ij}$  where  $i, j = 1, 2, 3$ . Again,  $a$  and  $b$  refer the scalar and pseudoscalar couplings, respectively. (12)

In this model, mono-top production primarily occurs through the resonant decay of  $\phi$  to  $\psi$  and  $t$ , as shown in Figure 3.3.

The two benchmark models show markedly different spectra in Figure 3.4, motivating their use to test different modes of mono-top production. The FCNC produces a falling  $p_T^{\text{miss}}$  distribution, whereas the scalar resonance produces a  $p_T^{\text{miss}}$  distribution peaking at approximately  $m_\phi/2$ .

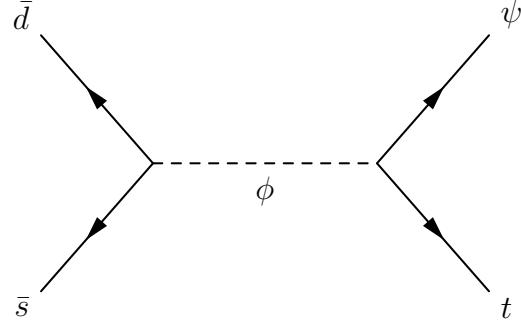


Figure 3.3: Possible DM production at the LHC, assuming the existence of a charged, color scalar that couples to DM and the top quark.

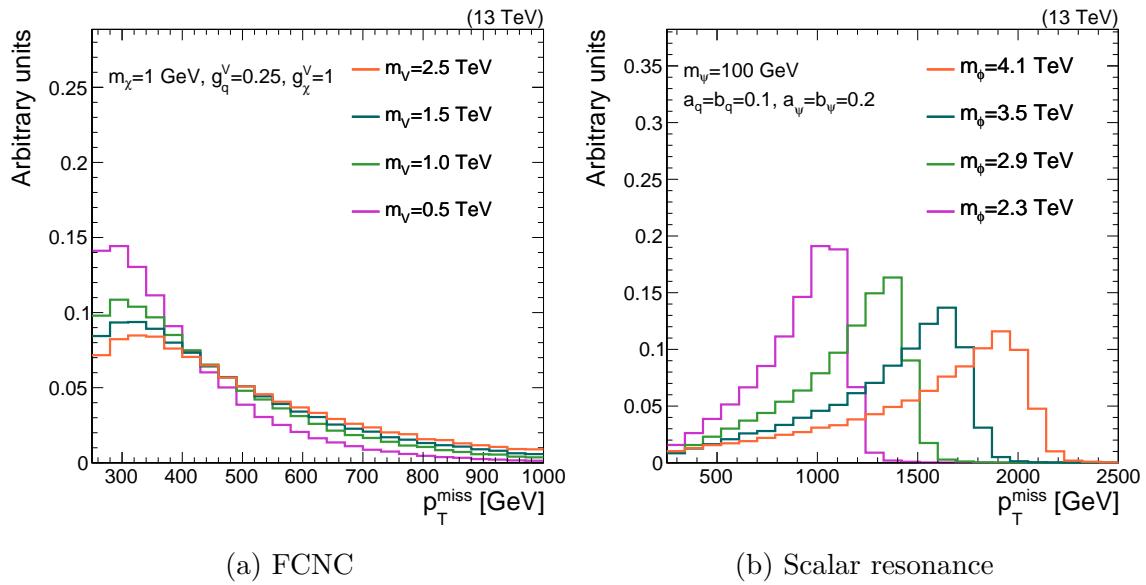


Figure 3.4: Spectra of DM (missing) momentum under various signal hypothesis.

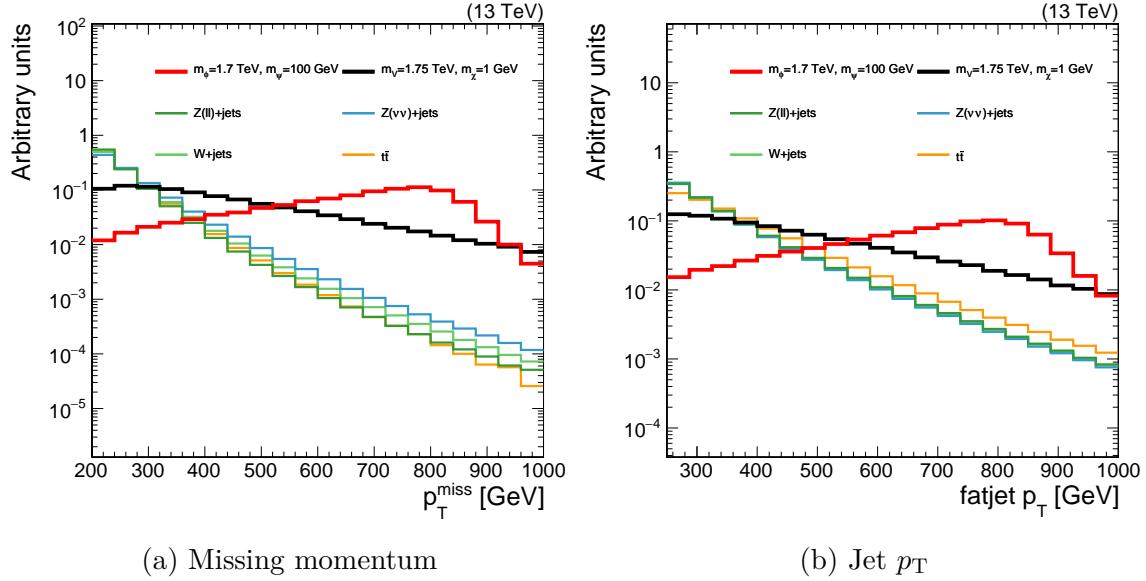


Figure 3.5: Comparison of missing and jet momenta in various backgrounds and signal models.

### 3.1 Signal selection

When looking at events that pass a simple set of criteria (moderate  $p_T^{\text{miss}}$  and one CA15 jet), it is clear (Figure 3.5) that the highest signal sensitivity is found in regions of high  $p_T^{\text{miss}}$  and jet  $p_T$ . The three primary background processes are:

- $Z \rightarrow \nu\nu$ . When the  $Z$  is produced in association with one or more jets, the jet system can (with low probability) pass the criteria used to select a top jet. The neutrinos manifest as  $p_T^{\text{miss}}$ .
- $W \rightarrow \ell\nu$ . As in the case of the  $Z$ , additional jets can mimic the signature of a top jet. Typically, the charged lepton in the final state is vetoed, but if it is out of acceptance ( $e, \mu$ ) or fails ID criteria ( $\tau_h$ ), then it is not identified.
- $t\bar{t} \rightarrow b\bar{q}\bar{q}' + \bar{b}\ell\nu$ . As in the case of the  $W$ , a charged lepton in the final state may not be properly identified. Unlike the previous two processes, a semi-leptonic  $t\bar{t}$  event contains a real hadronic top quark decay.

### 3.1.1 Online trigger selection

Data events are first selected by the L1 trigger system by requiring  $p_{T,L1}^{\text{miss}} > 70 \text{ GeV}$ , where:

$$p_{T,L1}^{\text{miss}} = - \left( \sum_{i \in C} \vec{p}_i \right)_T, \quad C = \{\text{calorimeter deposits with } |\eta| < 3.0\} \quad (3.3)$$

Events that pass this selection are sent to the HLT system, where we place requirements on the both the missing momentum ( $p_{T,\text{HLT}}^{\text{miss}}$ ) and the missing hadronic momentum ( $H_{T,\text{HLT}}^{\text{miss}}$ ). These are defined as:

$$p_{T,\text{HLT}}^{\text{miss}} = - \left( \sum_{i \in \text{particles}} \vec{p}_i \right)_T, \quad \text{all particles except muons} \quad (3.4)$$

$$H_{T,\text{HLT}}^{\text{miss}} = - \left( \sum_{i \in \text{jets}} \vec{p}_i \right)_T, \quad \text{jets passing noise-rejection ID} \quad (3.5)$$

The HLT decides to keep an event if  $\min(p_{T,\text{HLT}}^{\text{miss}}, H_{T,\text{HLT}}^{\text{miss}})$  is higher than a specified threshold. Over the course of the data-taking period considered in this chapter, this threshold varied from 90 to 120 GeV.

Note that in all of the trigger decisions, muons are implicitly (in the L1) or explicitly (in the HLT) excluded from the missing momentum calculations. This means that an event which produces high-momentum muons can be selected using these triggers. This flexibility will be exploited in Section 3.2.

Since the online environment and reconstruction are significantly limited as compared to the offline reconstruction of  $U$ , we do not expect the trigger decision to be a step function at  $U = 120 \text{ GeV}$ . Therefore, we define and measure a trigger efficiency:

$$\epsilon_{\text{trig}}(U) = \frac{N_{\text{pass trig}}(U)}{N(U)} \quad (3.6)$$

This is measured using  $W \rightarrow \mu\nu$  events containing one or more high- $p_T$  jets. The events are triggered using single- $\mu$  triggers (which have lower thresholds and efficiencies  $\sim 1$  in this phase space). We then require events have exactly one well-identified

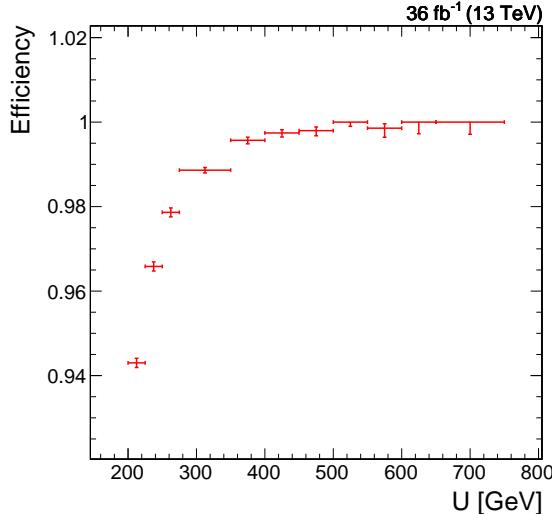


Figure 3.6: Efficiency of the  $p_T^{\text{miss}}$  trigger measured in single-muon events.

muon and at least one jet with  $p_T > 100$  GeV. Figure 3.6 shows the efficiency as a function of  $U$ . To avoid the steeply-rising part of the curve, we choose a threshold of  $U > 250$  GeV in the analysis.

### 3.1.2 Offline signal selection

Events in the signal regions (SRs) are then selected according to Table 3.1, chosen to be consistent with the signal topology while mitigating the aforementioned SM backgrounds. As described in Section ??, two working points (WPs) are defined for the top ID BDT. The signal events (passing all other selection criteria) are partitioned into a “loose” SR and a “tight” SR on the basis of which WP the top candidate jet satisfies.

Figure 3.7 shows the  $p_T^{\text{miss}}$  distributions, as predicted by MC and as observed in collected data, in the two signal regions.

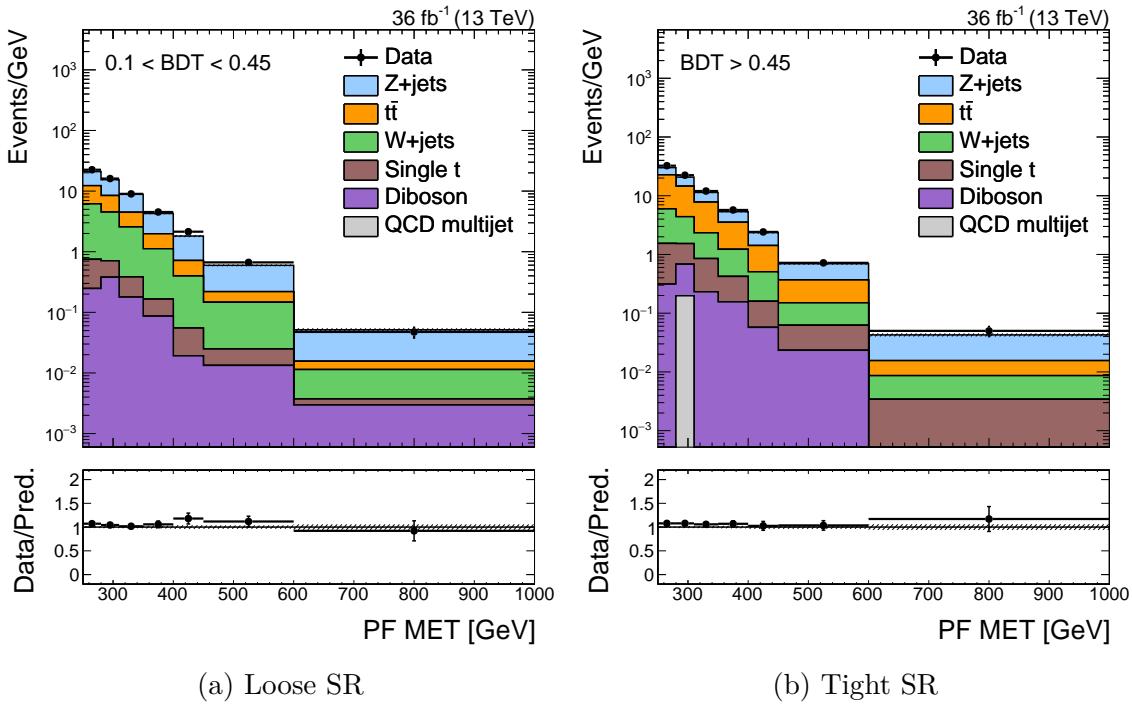


Figure 3.7:  $p_T^{\text{miss}}$  distributions in the two mono-top signal regions. The bottom section of each figure shows the ratio of the data and the prediction. The only uncertainties plotted in these figures are those arising from Poisson fluctuations in data (black bars) and MC (grey band).

Table 3.1: Criteria used to select events for the mono-top search signal regions. Note that two SRs are defined, based on the BDT score.

Criterion	Notes
$p_T^{\text{miss}} > 250 \text{ GeV}$	Signal events should have large missing momentum. Exact threshold is chosen to maximize online trigger efficiency.
1 CA15 jet with $p_T > 250 \text{ GeV}$	Top quark candidate. Recoils against $p_T^{\text{miss}}$ , so threshold is set at 250 GeV.
CA15 jet $110 < m_{\text{SD}} < 210 \text{ GeV}$	Consistency with top quark mass.
At least one $b$ -tagged sub-jet	Identifying $B$ hadron produced from top decay/hadronization.
No $b$ -tagged narrow jets	Rejecting semi-leptonic $t\bar{t}$ decays.
No identified $e, \mu, \tau_h$	Suppress $W+\text{jet}$ and $t\bar{t}$ processes.
No identified $\gamma$	Suppress $\gamma+\text{jet}$ processes.
$\min_{\text{jets}} \Delta\phi(\text{jet}, p_T^{\text{miss}}) > 0.5$	Remove events with large $p_T^{\text{miss}}$ caused by mismeasured jets.
CA15 jet BDT	Identifying top decay structure. If the jet passes the tight WP, it is placed in the “tight” SR. Otherwise, if it only passes the loose WP, it is placed in the “loose” SR.

## 3.2 Background estimation

Searching for DM amounts to looking for an excess of data events over the SM prediction at large values of  $p_T^{\text{miss}}$ . Therefore, the  $p_T^{\text{miss}}$  distribution of the three primary SM backgrounds described in Section 3.1 must be predicted with small uncertainty. The MC simulation provides a reasonable description of the data, but the theoretical uncertainties inherent in the MC (primarily due to higher-order QCD effects) can range up to 20%. To reduce the prediction uncertainty further, a “data-driven” approach is used to estimate the SM processes in the SR. In this context, “data-driven” refers to the use of control data (i.e. data that cannot contain signal) to directly estimate or supplement the estimation of SM processes in the SR.

### 3.2.1 Visible final states to constrain invisible final states

As a starting point, let us tackle the estimation of  $Z \rightarrow \nu\nu$  in the SR. Since the momentum imbalance (up to experimental effects) in a  $Z \rightarrow \nu\nu$  event is just the transverse momentum of the  $Z$  boson ( $p_T^Z$ ), we must estimate  $p_T^Z$ . To good approximation, the  $p_T^Z$  distribution is independent of the decay mode of the  $Z$  boson. Therefore, it is natural to estimate  $p_T^{\text{miss}}(Z \rightarrow \nu\nu)$  by measuring  $p_T^Z(Z \rightarrow \mu\mu)$ , as muons are easily identifiable and reconstructible.

However, there is one important distinction between  $\nu\nu$  and  $\mu\mu$  events. In the latter,  $p_T^Z$  can be directly measured, whereas in the former it must be inferred through a momentum imbalance. Effects like jet energy scale and acceptance can impact  $p_T^{\text{miss}}$ , but not  $p_T^{\mu\mu}$ . Therefore, instead of directly measuring  $p_T^{\mu\mu}$  in  $\mu\mu$  events, we define and use the hadronic recoil  $U$ :

$$\vec{U} = \vec{p}_T^{\text{miss}} + \sum_{\mu} \vec{p}_T^{\mu} + \sum_e \vec{p}_T^e + \sum_{\gamma} \vec{p}_T^{\gamma} \quad (3.7)$$

In the SR (where there are no  $e, \mu, \gamma$ ),  $U = p_T^{\text{miss}}$ . In  $Z \rightarrow \mu\mu$  events,  $U$  mimics the momentum imbalance, if we had pretended the identified muons did not exist when computing  $p_T^{\text{miss}}$ . Therefore,  $U$  is an exact analogy for  $p_T^{\text{miss}}$  in the SR. Figure 3.8

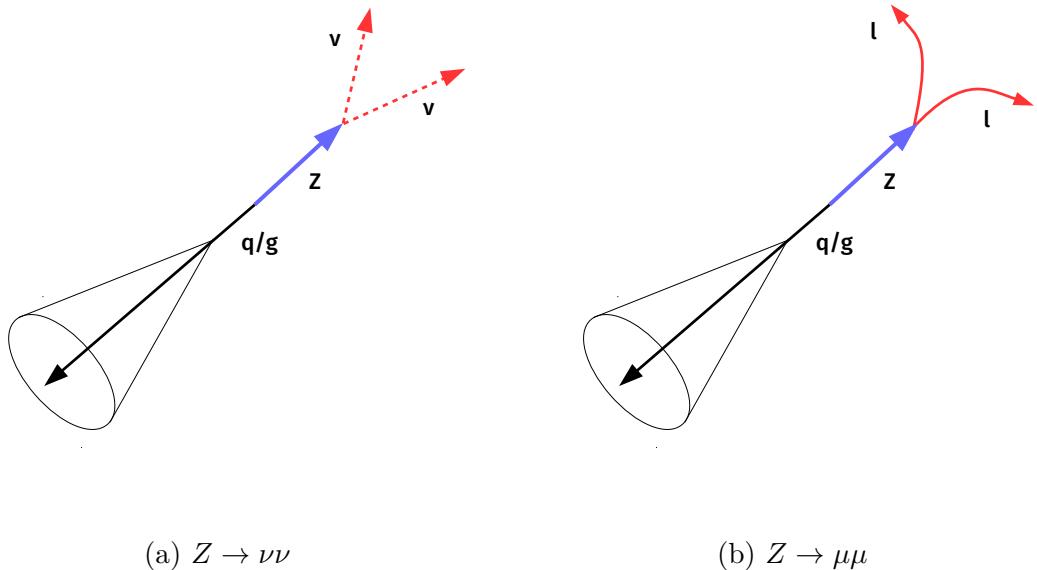


Figure 3.8: Schematic representation of two  $Z$  decay modes: to neutrinos (as in the SR) and to muons (as in the CRs). Note that in both cases,  $U$  is sensitive to the same effects arising from the measurement of the jet recoiling against the  $Z$  boson, whereas  $p_T^{\mu\mu}$  is largely independent of the jet.

makes the same argument in a schematic fashion.

Table 3.2 describes the criteria used to define events in the “ $\mu\mu$ ” control regions (CRs). Figure 3.9 shows the distribution of  $U$  in these CRs, as well as the  $m_{\mu\mu}$  and  $p_T^\mu$  distributions.

The control data is used to constrain the SR prediction by means of “transfer factors”  $T_{Y,i}^X$ , where  $X$  refers to a particular CR (e.g.  $\mu\mu$ ),  $Y$  refers to a particular process (e.g.  $Z$ ), and  $i$  refers to a particular bin in the CR (e.g.  $200 < U < 250$  GeV in the tight category). Formally:

$$T_{Z,i}^{\mu\mu} = \frac{N_i^{\text{SR}}(Z \rightarrow \nu\nu)}{N_i^{\mu\mu}(Z \rightarrow \mu\mu)} \quad (3.8)$$

The transfer factors are estimated using MC simulation. To encode the effects of

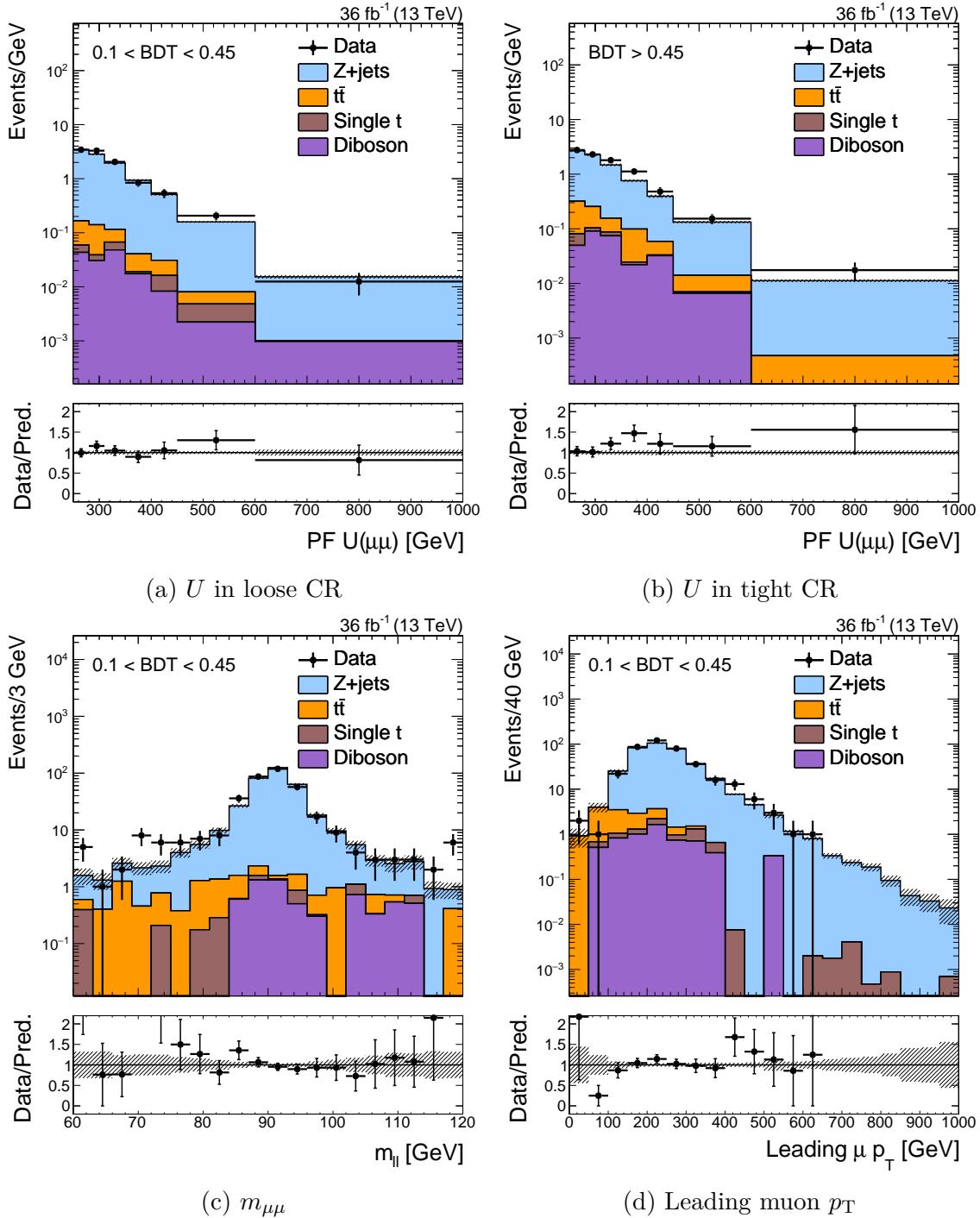


Figure 3.9: Various kinematic distributions in the two mono-top  $\mu\mu$  CRs. Note the clearly discernible peak in the  $m_{\mu\mu}$  distribution near  $m_Z$

Table 3.2: Criteria used to select events for the mono-top  $Z \rightarrow \mu\mu$  CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_T > 250$ GeV	Same as SR
CA15 jet $110 < m_{SD} < 210$ GeV	Same as SR
Well-identified $\mu^-, \mu^+$ pair, with $ m_{\mu\mu} - m_Z  < 30$ GeV	Identifying the $Z \rightarrow \mu\mu$ resonance.
No identified $e, \tau_h$	Same as SR.
No identified $\gamma$	Same as SR
$\min_{\text{jets}} \Delta\phi(\text{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

various uncertainties, we introduce nuisance parameters  $\boldsymbol{\theta}$ . That is:

$$T_{Y,i}^X \rightarrow T_{Y,i}^X(\boldsymbol{\theta}) \equiv T_{Y,i}^X \times \prod_{j=0}^{n_\theta} (1 + \theta_j) \quad (3.9)$$

$$\theta_j \sim p_j(\theta_j) \quad (3.10)$$

where  $n_\theta$  is the number of nuisance parameters and  $p_j(\theta_j)$  is some prior distribution for each nuisance (see below for how the priors are used). The priors are typically chosen to have a central value (e.g. mean, median) at 0, with a finite variance that encodes the uncertainty. In this chapter, we assume  $p_j$  is either a normal distribution centered at 0 or a log-normal distribution (in cases where negative values are undesirable). We will use the terms “uncertainty” and “nuisance parameter” interchangeably.

Let  $\text{Pois}(d|\lambda)$  refer to the Poisson probability of observing  $d$  with an expected mean of  $\lambda$ . In terms of these transfer factors, the likelihood for the data observed in the signal and  $\mu\mu$  control regions is:

$$\begin{aligned} \mathcal{L}(\mathbf{d} \mid \mu, \boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}, \boldsymbol{\theta}) = \prod_{i \in \text{bins}} & \left[ \text{Pois}\left(d_i^{\text{SR}} \mid \mu S_i^{\text{SR}}(\boldsymbol{\theta}) + \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} + B_i^{\text{SR}}(\boldsymbol{\theta})\right) \right. \\ & \times \text{Pois}\left(d_i^{\mu\mu} \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^{\mu\mu}(\boldsymbol{\theta})} + B_i^{\mu\mu}(\boldsymbol{\theta})\right) \left. \right] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \end{aligned} \quad (3.11)$$

where the following notation is used:

$d_i^X$  : The data observed in bin  $i$  of region  $X$ . For now,  $X = \text{SR}, \mu\mu$ .

$S_i^{\text{SR}}$  : The predicted number of signal events in bin  $i$  of the SR, under some fixed signal hypothesis.

$\mu$  : The “signal strength”. Essentially an unconstrained nuisance parameter that scales up or down the total signal yield.

$\mu_{\text{SR},i}^P$  : The expected number of events from process  $P$  in bin  $i$  of the SR. This is also an unconstrained nuisance parameter.

$B_i^X$  : The predicted number of “minor” background events in bin  $i$  of region  $X$ . Here, “minor” refers to all SM processes that are not the signal and are not estimated using a data-driven method.

The signal and background yields  $\mathbf{S}$  and  $\mathbf{B}$  are estimated using MC. Note that the inclusion of the priors in the likelihood constrains the nuisance parameters to be close to their “nominal” values; moving a  $\theta_j$  to fit the data incurs a large cost from the prior.

If we set  $B_i = \mu = 0$  (the null hypothesis, ignoring small minor backgrounds), a simple picture emerges of how the likelihood is maximized. The parameters  $\mu_{\text{SR}}^{Z \rightarrow \nu\nu}$  float freely to satisfy  $d_{\text{SR},i} \sim \mu_{\text{SR},i}^{Z \rightarrow \nu\nu}$  and  $d_{\mu\mu,i} \sim \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} / T_{Z,i}^{\mu\mu}(\boldsymbol{\theta})$ . If both constraints cannot be satisfied simultaneously by scaling  $\mu_{\text{SR}}^{Z \rightarrow \nu\nu}$ , the (constrained) nuisance parameters  $\boldsymbol{\theta}$  modify the transfer factor  $T_Z^{\mu\mu}$ . Table 3.6 shows the relevant uncertainties for  $T_Z^{\mu\mu}$ , and Figure 3.10 shows the shape of uncertainties that evolve as a function of  $U$ .

The transfer factors are shown in Figure 3.11. The exact values of  $T_Z^{\mu\mu}(\boldsymbol{\theta})$  have two salient features:

1. The values are strictly greater than one. This is due to (a)  $\mathcal{B}(Z \rightarrow \nu\nu) > \mathcal{B}(Z \rightarrow \mu\mu)$  and (b) a non-100% efficiency in reconstructing and identifying muons. This implies that the constraining power of the  $\mu\mu$  CR is less than that of the SR, especially at high  $U$  (i.e. the Poisson uncertainties are larger).

Table 3.3: Uncertainties affecting the  $\mu\mu \leftrightarrow \nu\nu$  extrapolation. “Shape” uncertainties have different priors for each bin, but are assumed to be correlated across bins.

Uncertainty	1 s.d.	Notes
$\mu$ ID	2%	
$\mu$ track	1%	
$\tau_h$ veto	3%	
$Z + \text{heavy flavor}$	3%	
Trigger	0-2%	Shape
$b$ -tag	$\sim 0.5\%$	Shape
$udcsg$ -mistag	3-4%	Shape

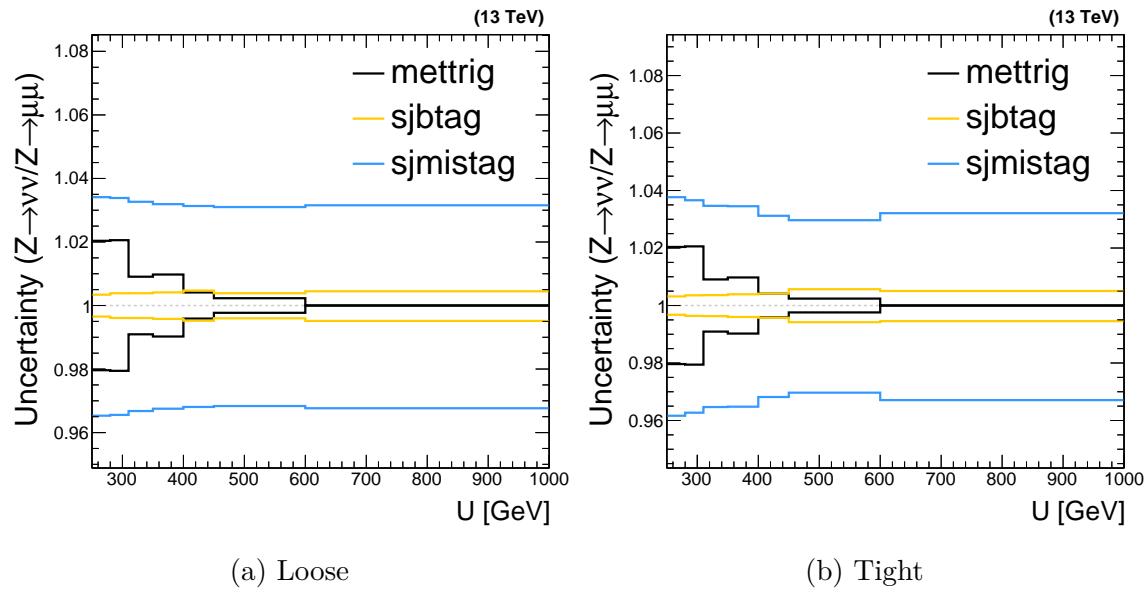


Figure 3.10: Shape uncertainties affecting  $T_i^{\mu\mu}$  in both categories, as a function of  $U$ .

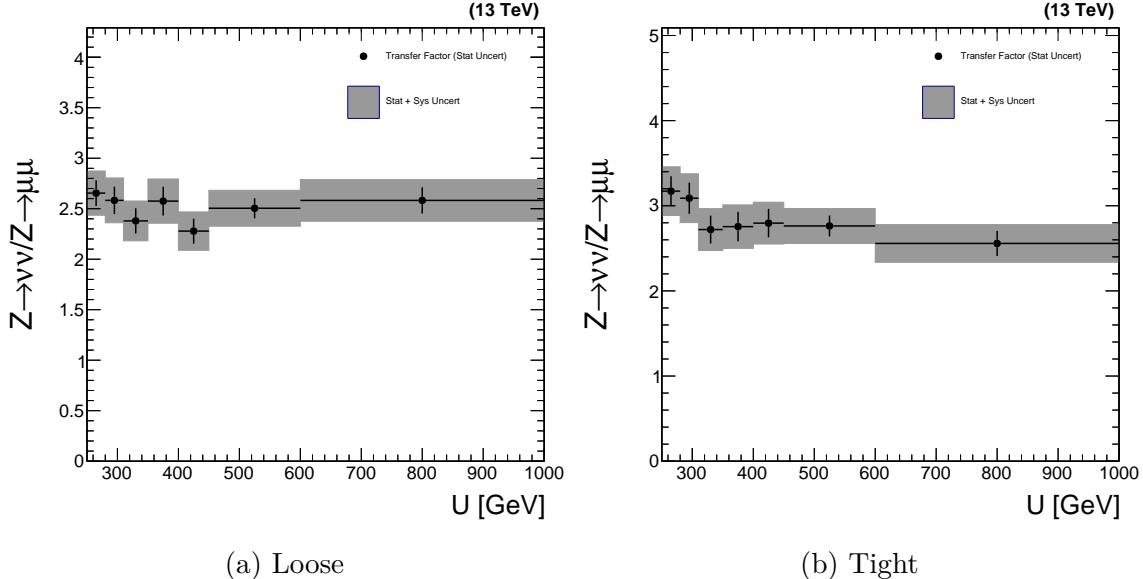


Figure 3.11: The transfer factors  $T_{Z,i}^{\mu\mu}$  as a function of recoil and BDT score. The vertical black bars represent the Poisson uncertainties in the MC simulation, while the grey bands represent the sum of Poisson uncertainties and other, systematic, uncertainties. All uncertainties are represented as one standard deviation.

2. The one standard deviation variation of all uncertainties that impact  $T_{Z,i}^{\mu\mu}$  are contained within a 10% envelope. This is already a factor of two smaller than the inherent  $\sim 20\%$  uncertainties in the MC simulation.

To account for point (1), we can simply add more control data by also looking at  $Z \rightarrow ee$  decays. Essentially all of the arguments used for the  $\mu\mu$  CRs applies to the  $ee$  CRs. Figures 3.12-3.13 show the data/simulation agreement and the transfer factors for the new dielectron regions. A further set of statistical constraints to improve the estimate at high  $U$  (which is where the signals are most enhanced) is described in Section 3.2.2.

Similar methods are used to predict the  $W$ +jets and  $t\bar{t}$  contributions in the SRs; these three backgrounds comprise at least 95% of the SM processes. In both cases, the momentum imbalance in the SR is a proxy for the momentum of the  $W$  boson, since the charged lepton is lost. A sketch of the event topologies is shown in Figure 3.14. Following the same arguments as used for the  $Z \rightarrow \ell\ell$  CRs, we can use the hadronic recoil  $U$  in CRs that measure visible final states of  $W$  and  $t\bar{t}$ .

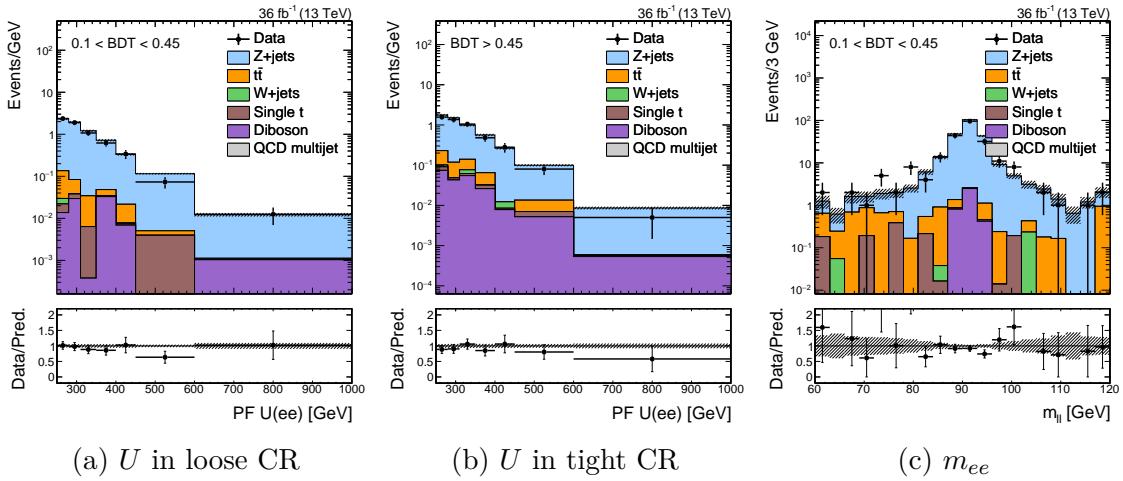


Figure 3.12: Various kinematic distributions in the two mono-top  $ee$  CRs.

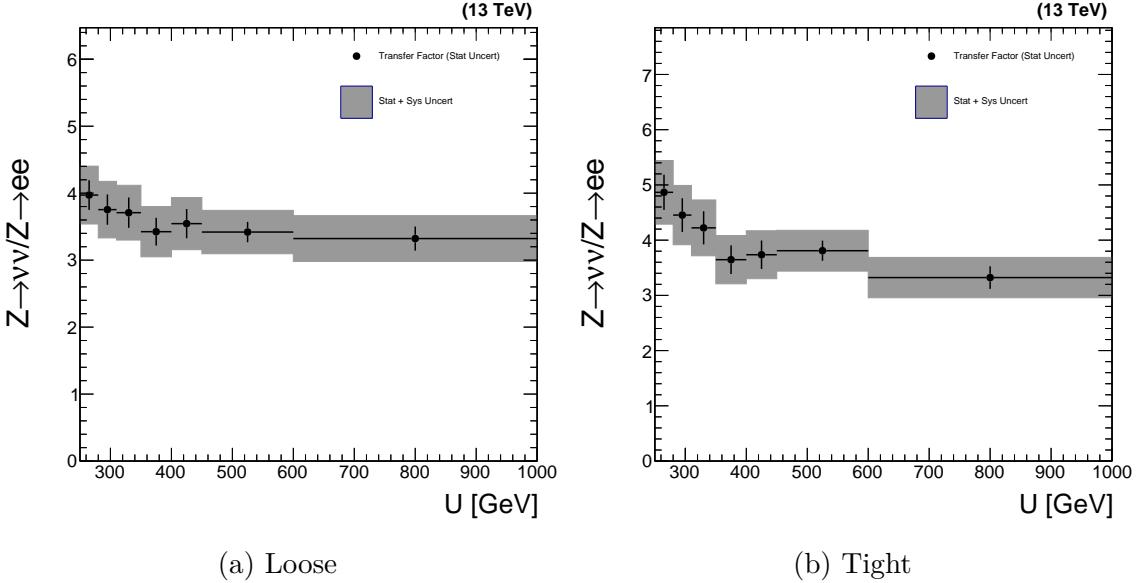


Figure 3.13: The transfer factors  $T_i^{ee}$  as a function of recoil and BDT score. The vertical black bars represent the Poisson uncertainties in the MC simulation, while the grey bands represent the sum of Poisson uncertainties and other, systematic, uncertainties. All uncertainties are represented as one standard deviation.

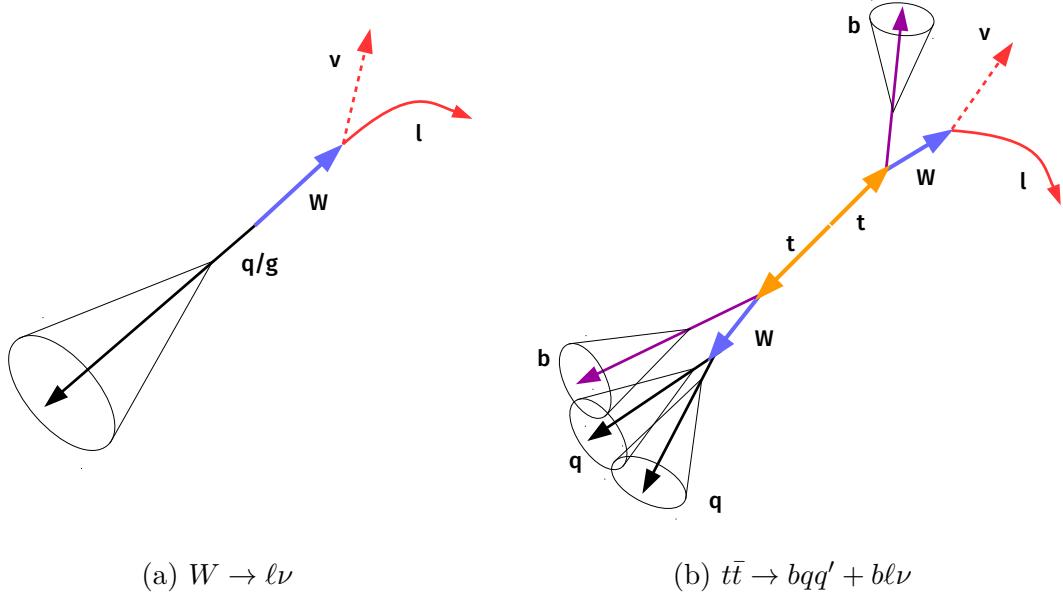


Figure 3.14: Schematic representation of the  $W$  and  $t\bar{t}$  SM processes. In both cases,  $U \approx p_T^W$ . Furthermore, if the charged lepton is lost,  $U = p_T^{\text{miss}} \approx p_T^W$ .

Starting with muon final states (electrons will follow naturally), we define two sets of CRs based on the number of identified  $B$  hadrons. The selection for the  $b\mu$  CRs (to measure  $t\bar{t}$ ) is shown in Table 3.4. The selection for the  $\mu$  CRs (to measure  $W$ ) is shown in Table 3.5. Figures 3.15-3.16 show various kinematic distributions in these regions.

Each CR gets a set of transfer factors to constrain the targeted process in the SR:  $\mathbf{T}_{t\bar{t}}^{b\mu}$  and  $\mathbf{T}_W^\mu$ . In the tight  $\mu$  CR (Figures 3.16b-3.16c), the stringent top ID requirement enhances the  $t\bar{t}$  and suppresses the  $W$  contribution. Since we cannot create a pure  $W$  in the tight category, we introduce an additional set of transfer factors  $\mathbf{T}_{t\bar{t}}^\mu$ . This extra constraint uses the  $b\mu$  CRs to estimate the  $t\bar{t}$  component in the  $\mu$  CRs, thereby leaving only one large degree of freedom in the  $\mu$  CRs. These three sets of transfer factors, and the corresponding uncertainties, are shown in Figure 3.17.

As we added the  $ee$  CRs to complement the  $\mu\mu$  CRs, we also add  $be$  ( $e$ ) CRs to augment the statistical power of the  $b\mu$  ( $\mu$ ) CRs, especially at high recoil. Figures 3.18 and 3.19 respectively show some kinematic distributions and the transfer factors corresponding to these electron constraints.

Table 3.4: Criteria used to select events for the mono-top  $b\mu$  CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250 \text{ GeV}$	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_T > 250 \text{ GeV}$	Same as SR
$\text{CA15 jet } 110 < m_{\text{SD}} < 210 \text{ GeV}$	Same as SR
At least one $b$ -tagged sub-jet	Identifying $B$ hadron produced from hadronic top decay.
Exactly one $b$ -tagged narrow jet	Identifying $B$ hadron produced from leptonic top decay.
Exactly one well-identified $\mu$	Produced from $W \rightarrow \mu\nu$
No identified $e, \tau_h$	Same as SR.
No identified $\gamma$	Same as SR
$\min_{\text{jets}} \Delta\phi(\text{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

Table 3.5: Criteria used to select events for the mono-top  $\mu$  CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250 \text{ GeV}$	Mimicking the selection in the SR; also constrained by trigger thresholds.
1 CA15 jet with $p_T > 250 \text{ GeV}$	Same as SR
$\text{CA15 jet } 110 < m_{\text{SD}} < 210 \text{ GeV}$	Same as SR
No $b$ -tagged sub-jets	Suppressing semi-leptonic $t\bar{t}$ decays
No $b$ -tagged narrow jets	Suppressing semi-leptonic $t\bar{t}$ decays
Exactly one well-identified $\mu$	Produced from $W \rightarrow \mu\nu$
No identified $e, \tau_h$	Same as SR.
No identified $\gamma$	Same as SR
$\min_{\text{jets}} \Delta\phi(\text{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

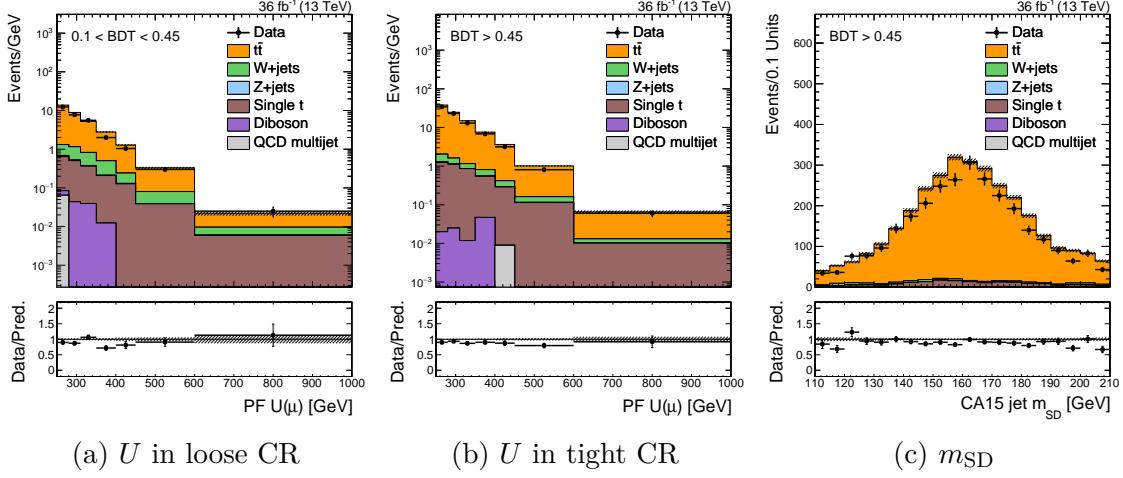


Figure 3.15: Various kinematic distributions in the two mono-top  $b\mu$  CRs.

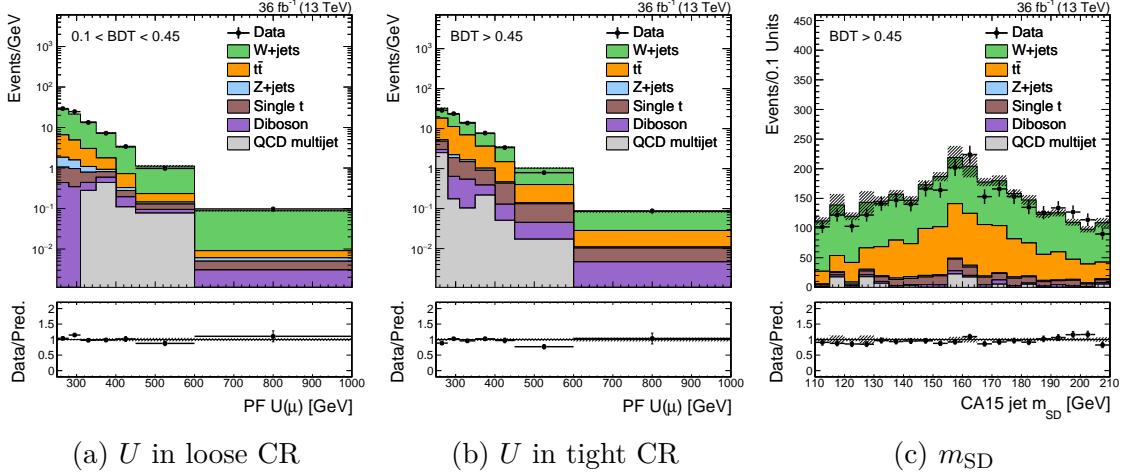


Figure 3.16: Various kinematic distributions in the two mono-top  $\mu$  CRs.

Table 3.6: Uncertainties affecting the various single-muon extrapolations. “Shape” uncertainties have different priors for each bin, but are assumed to be correlated across bins.

Uncertainty	1 s.d. ( $T_{t\bar{t}}^{b\mu}$ )	1 s.d. ( $T_W^\mu$ )	1 s.d. ( $T_{t\bar{t}}^\mu$ )	Notes
$\mu$ ID	1%	1%	1%	
$\mu$ track	0.5%	0.5%	0.5%	
$\tau_h$ veto	3%	3%	3%	
$W$ +heavy flavor		3%		
Trigger	0-2%	0-2%	0-2%	Shape
$b$ -tag	2%	$\sim 0.5\%$	3-6%	Shape
$udcsg$ -mistag	1%	5%	1%	Shape

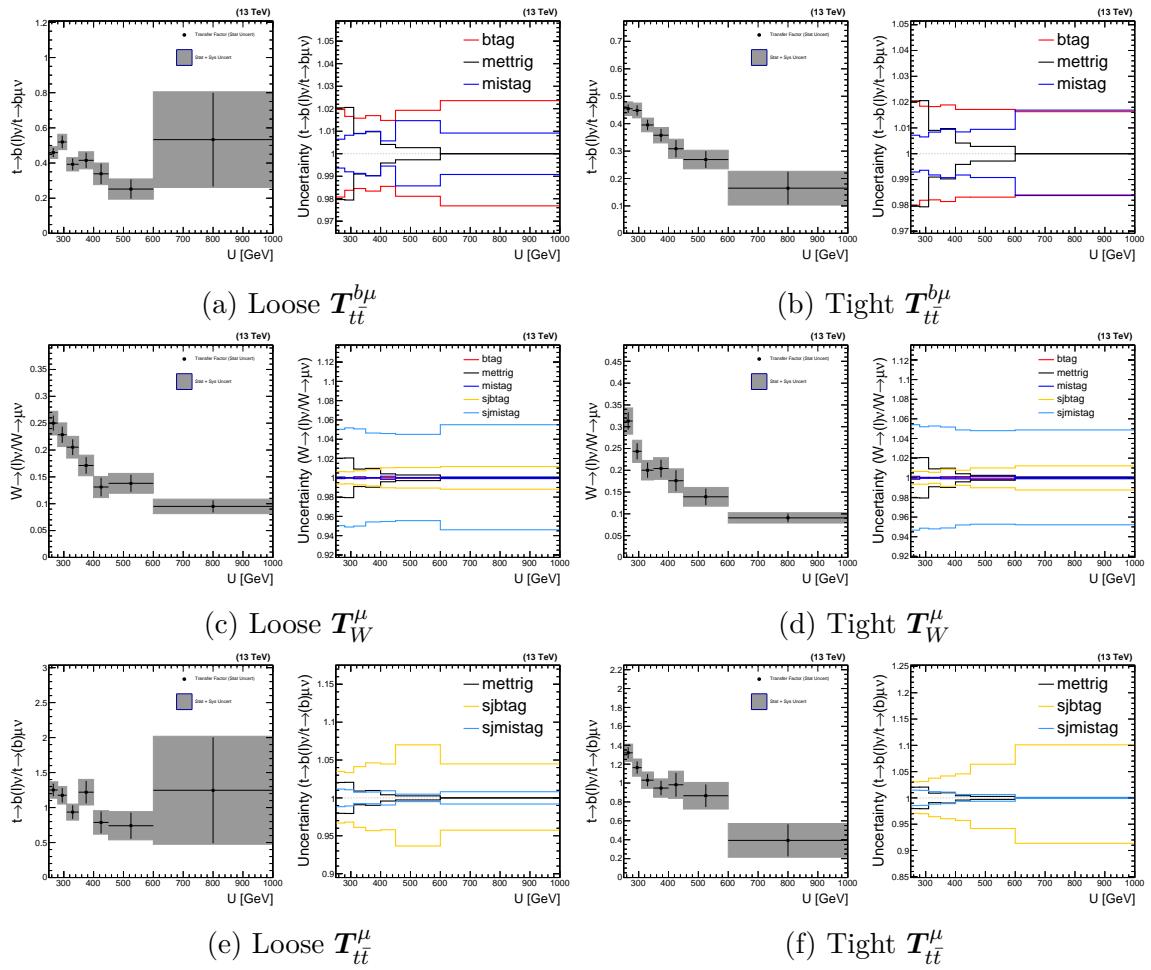


Figure 3.17: The transfer factors  $T_{t\bar{t}}^{b\mu}$ ,  $T_W^\mu$ , and  $T_{t\bar{t}}^\mu$ ; and corresponding shape uncertainties.

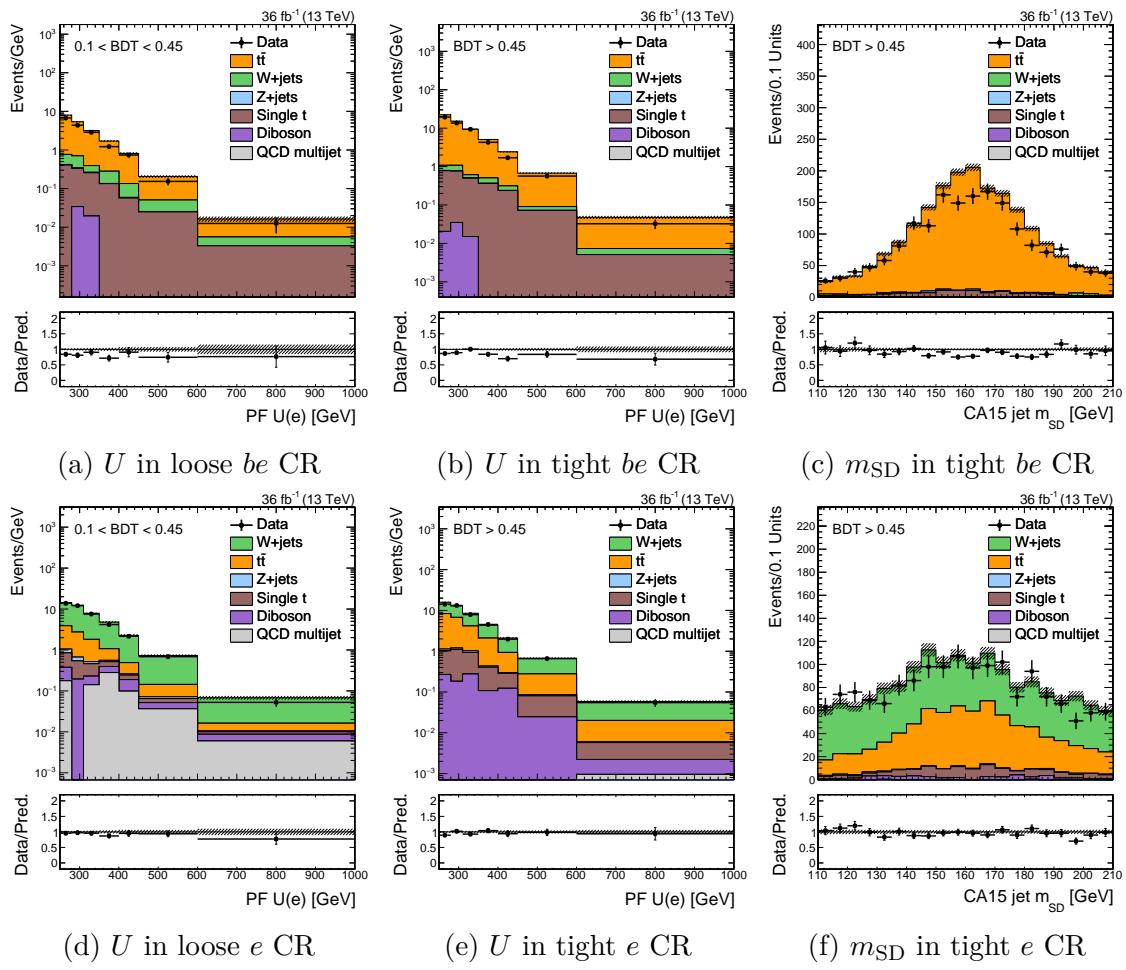


Figure 3.18: Various kinematic distributions in the mono-top  $be$  CRs (top) and  $e$  CRs (bottom).

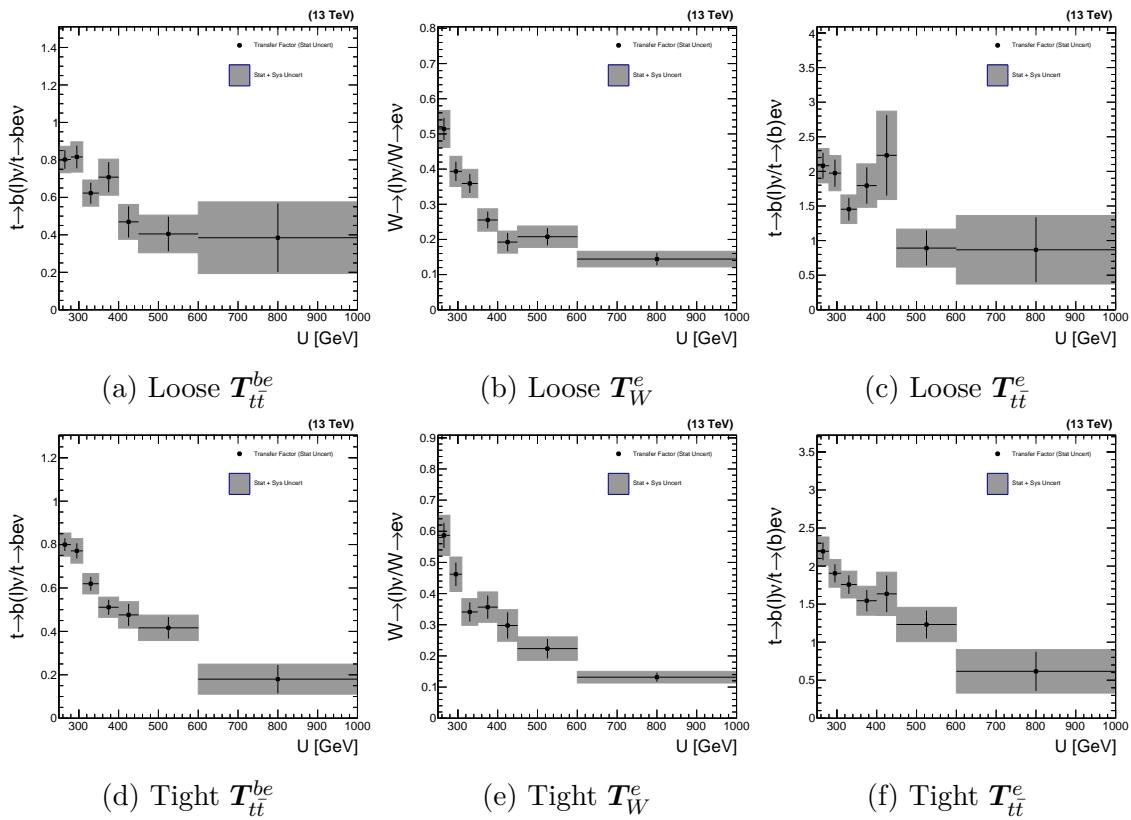


Figure 3.19: The transfer factors  $\mathbf{T}_{tt}^{be}$ ,  $\mathbf{T}_W^e$ , and  $\mathbf{T}_{tt}^e$

Having defined (almost all of) the CRs and transfer factors, we can write down a complete likelihood for the mono-top search:

$$\begin{aligned} \mathcal{L}(\mathbf{d} | \mu, \boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}, \boldsymbol{\mu}_{\text{SR}}^{W \rightarrow \ell\nu}, \boldsymbol{\mu}_{\text{SR}}^{t\bar{t}}, \boldsymbol{\theta}) = \\ \prod_{i \in \text{bins}} \left[ \text{Pois}\left(d_i^{\text{SR}} \mid \mu S_i^{\text{SR}}(\boldsymbol{\theta}) + \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} + \mu_{\text{SR},i}^{W \rightarrow \ell\nu} + \mu_{\text{SR},i}^{t\bar{t}} + B_i^{\text{SR}}(\boldsymbol{\theta})\right) \right. \\ \times \prod_{X=\mu\mu,ee} \text{Pois}\left(d_i^X \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta})\right) \\ \times \prod_{X=b\mu,be} \text{Pois}\left(d_i^X \mid \frac{\mu_{\text{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta})\right) \\ \left. \times \prod_{X=\mu,e} \text{Pois}\left(d_i^X \mid \frac{\mu_{\text{SR},i}^{W \rightarrow \ell\nu}}{T_{W,i}^X(\boldsymbol{\theta})} + \frac{\mu_{\text{SR},i}^{t\bar{t}}}{T_{t\bar{t},i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta})\right)\right] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \quad (3.12) \end{aligned}$$

### 3.2.2 Theoretically-limited extrapolations

Despite the combination of the  $\mu\mu$  and  $ee$  regions, there are still large statistical uncertainties in the estimate of  $Z \rightarrow \nu\nu$  at high  $U$ . This is apparent in Figure 3.12, in which exactly one event is observed in the last bin of the tight CR. The dilepton CRs are limited by  $\sigma(pp \rightarrow Z \rightarrow \nu\nu) > \sigma(pp \rightarrow Z \rightarrow \ell^+\ell^-)$ ; accordingly, to alleviate this limitation, we look to a process with a much bigger cross-section. In similar regions of final-state phase space,  $\sigma(pp \rightarrow \gamma + \text{jets}) \sim 30 \times \sigma(pp \rightarrow Z(\rightarrow \nu\nu) + \text{jets})$ . Therefore, it is natural to use the  $\gamma$ +jet production spectrum as a way to estimate the  $Z$ +jet spectrum. As before, let us define another transfer factor:

$$T_{\gamma,i}^\gamma = \frac{N_i^{\text{SR}}(Z \rightarrow \nu\nu)}{N_i^\gamma(\gamma)} \quad (3.13)$$

However, unlike the  $\mathbf{T}_Y^X$  we have discussed so far (which correlate similar processes, e.g.  $Z \rightarrow \nu\nu/Z \rightarrow \mu\mu$ ),  $T_{\gamma}^\gamma$  is highly sensitive to the theoretical predictions of the  $Z$  and  $\gamma$  spectra.

To reduce the impact of higher-order effects on  $\mathbf{T}_\gamma^\gamma$ , we ensure that the numerator and denominator are predicted to as high an order as possible. Leading-order MC is used to simulate  $V$ +jet processes, because higher-order simulations are much more

computationally intensive. We therefore choose to produce a less accurate LO simulation, as opposed to a more accurate, but statistically-limited, NLO simulation. While producing sufficient NLO simulation for the analysis is prohibitive, we can compute certain inclusive distributions at NLO. Since  $U \approx p_T^V$  is the quantity of interest in this analysis, we want to ensure this distribution is accurately predicted. It is clear from Figure 3.20 that adding an additional QCD order induces large corrections, both at low and high  $p_T^V$ . In the LO simulation, we can obtain an estimate of the uncertainty due to NLO effects by varying the renormalization and factorization scales ( $\mu_R, \mu_F$ ) by factors of two. This is represented by the red envelope and grey band in Figure 3.20 and clearly is insufficient to cover NLO effects. Therefore, we compute a simple correction for the NLO QCD effects, known as a  $k$ -factor:

$$k_{Z,\text{QCD}}(p_T^Z) = \frac{d\sigma_{\text{NLO QCD}}(Z)/dp_T^Z}{d\sigma_{\text{LO}}(Z)/dp_T^Z} \quad (3.14)$$

We include another  $k$ -factor,  $k_{\text{EWK}}$ , to correct for higher-order EWK effects. Unlike  $k_{\text{QCD}}$ ,  $k_{\text{EWK}}$  is derived using a theoretical calculation [34, 35, 36] instead of NLO simulation.  $k_{\text{EWK}}$  covers NLO EWK terms, as well as large Sudakov logarithms that appear at high  $p_T^V$  in the NNLO expansion (NLL).

Figure 3.21 compares the  $k$ -factors for all three  $V + \text{jet}$  processes. While there are similar trends as a function of  $p_T^V$ , it is clear that the corrections are quite different for each process. Therefore, transfer factors like  $T_{\gamma}^{\gamma}$  are strongly sensitive to NLO effects, i.e.:

$$T_{\gamma,i}^{\gamma} = \frac{N_i^{\text{SR}}(Z \rightarrow \nu\nu)}{N_i^{\gamma}(\gamma)} \neq \frac{N_i^{\text{SR,LO}}(Z \rightarrow \nu\nu)}{N_i^{\gamma,\text{LO}}(\gamma)} \quad (3.15)$$

The distributions in Sections 3.1-3.2 are all corrected using these  $k$ -factors. Figure 3.22 shows the equivalent for CRs that target  $\gamma + \text{jet}$  events, and Table 3.7 describes the selection used to define these CRs.

Now that we can describe  $T_{\gamma}^{\gamma}$  at NLO, we must assess the impact of unknown higher-order terms on the transfer factors. We account for variations caused by uncertainties in the PDF model by taking the RMS of the 100 parameter variations prescribed for the NNPDF3.0 set [37]. By varying  $\mu_F$  and  $\mu_R$  by factors of 0.5 and

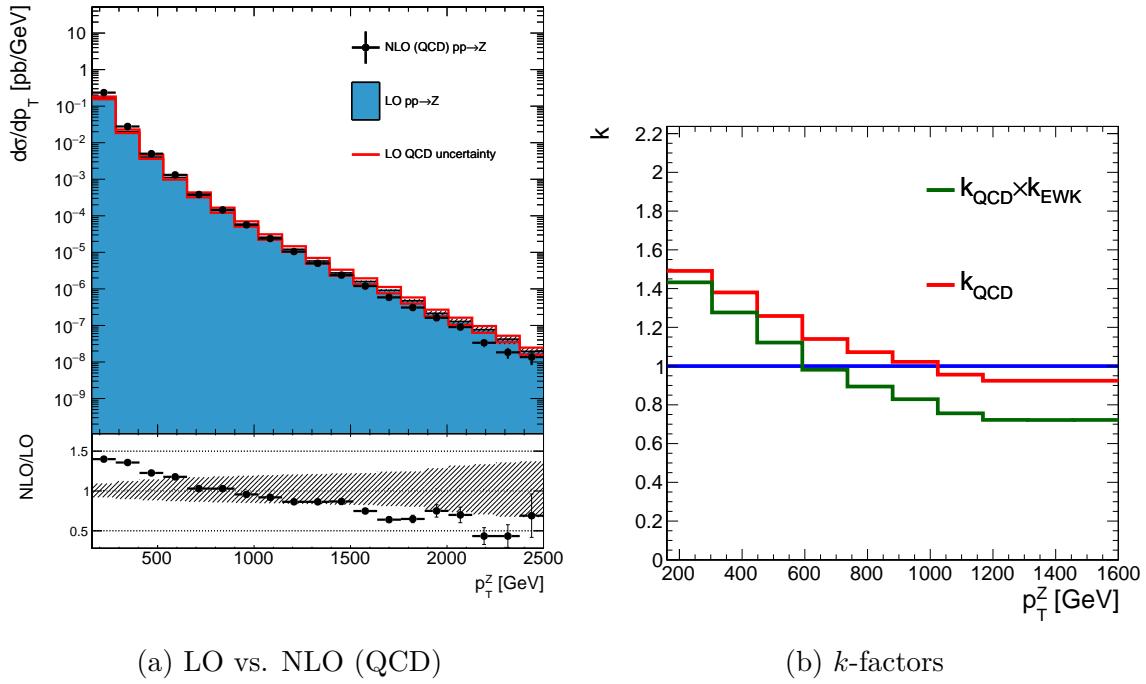


Figure 3.20: Theoretical predictions for  $p_T^Z$  in  $Z$ +jet events and the corresponding  $k$ -factors. No detector simulation is applied in these figures; all quantities are directly from MC simulation of the physics process. “LO QCD uncertainty” refers to an estimate of the effect of the QCD renormalization and factorization scales on the LO simulation. The grey band in the ratio is the quadrature sum of the QCD and statistical uncertainties.

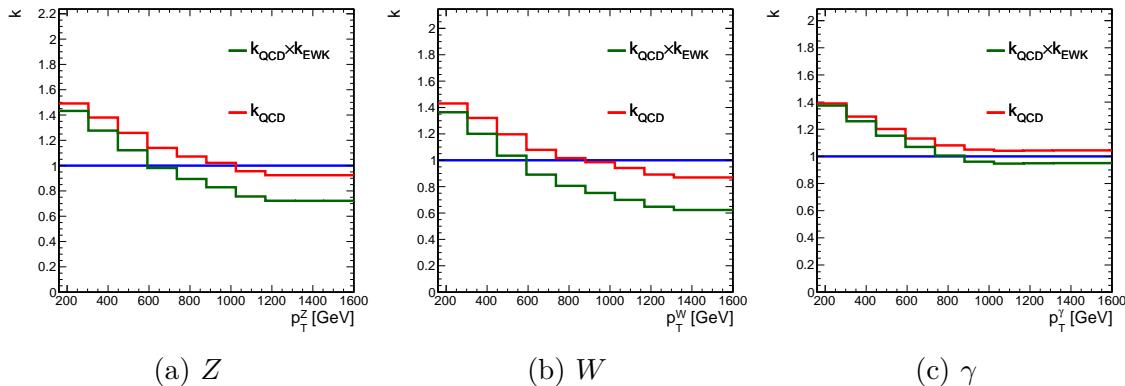


Figure 3.21: Differential  $k$ -factors for each of the  $V$ +jet processes.

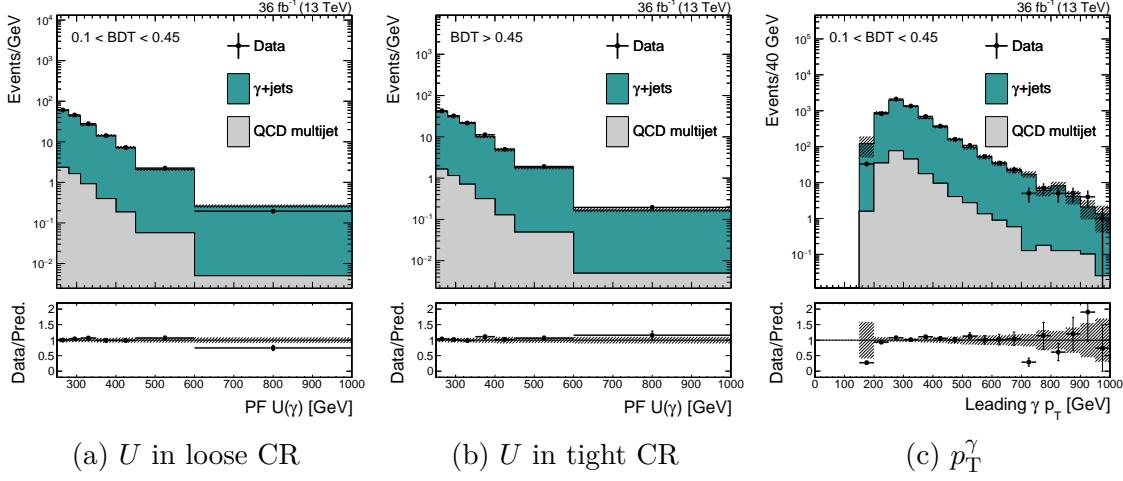


Figure 3.22: Various kinematic distributions in the two mono-top  $\gamma$  CRs.

Table 3.7: Criteria used to select events for the mono-top  $\gamma$  CR. As with the SR, the region is further split based on the jet BDT score.

Criterion	Notes
$U > 250$ GeV	Mimicking the selection in the SR
1 CA15 jet with $p_T > 250$ GeV	Same as SR
CA15 jet $110 < m_{SD} < 210$ GeV	Same as SR
No identified $\mu, e, \tau_h$	Same as SR.
Well-identified $\gamma$ with $p_T^\gamma > 175$ GeV	High- $p_T$ photon, set by trigger threshold
$\min_{\text{jets}} \Delta\phi(\text{jet}, U) > 0.5$	Same as SR
CA15 jet BDT	Same as SR

2, we assess the effect of the integration scale choices on  $\mathbf{T}$ . These scale and PDF uncertainties cover all unknown QCD effects on the production of electroweak bosons. To be conservative, they are assumed to be uncorrelated between processes. However, the uncertainties are correlated between all bins (i.e. as a function of  $p_T^V$ ). A second set of uncertainties is included for higher-order EWK effects, following what is suggested in References [38, 39, 40, 34, 35, 41, 42, 36, 43] and agreed upon in the LHC Dark Matter Working Group. These EWK uncertainties break down into three categories:

- Unknown Sudakov logarithms in the NLL correction. These uncertainties are correlated across processes ( $Z, W, \gamma$ ).
- Missing NNLO EWK effects not covered by the NLL correction. These are not correlated across processes.
- The full difference between the NLL correction and an exponentiation of the NLO correction; also not correlated across processes.

It should be stressed that while these uncertainties apply to the prediction of each  $V$ +jet processes, they do not affect transfer factors that correlate processes differing only in decay mode or acceptance. This is simply because these uncertainties primarily deal with the initial state or the production of an electroweak boson, which is not related to the description of the decay to leptons or the experimental identification of leptons and  $b$ -jets. That is:

$$T_{Z,i}^{\mu\mu} = \frac{N_i^{\text{SR}}(Z \rightarrow \nu\nu)}{N_i^{\mu\mu}(Z \rightarrow \mu\mu)} \approx \frac{N_i^{\text{SR,LO}}(Z \rightarrow \nu\nu)}{N_i^{\mu\mu,\text{LO}}(Z \rightarrow \mu\mu)} \quad (3.16)$$

Now that we have tools to construct transfer factors of the form  $N(V)/N(V')$  with reasonably small uncertainties (i.e. smaller than the statistical uncertainty of the data), it is natural to add another transfer factor to our toolbox:

$$T_{Z/W,i}^{\text{SR}} = \frac{N_i^{\text{SR}}(Z \rightarrow \nu\nu)}{N_i^{\text{SR}}(W \rightarrow \ell\nu)} \quad (3.17)$$

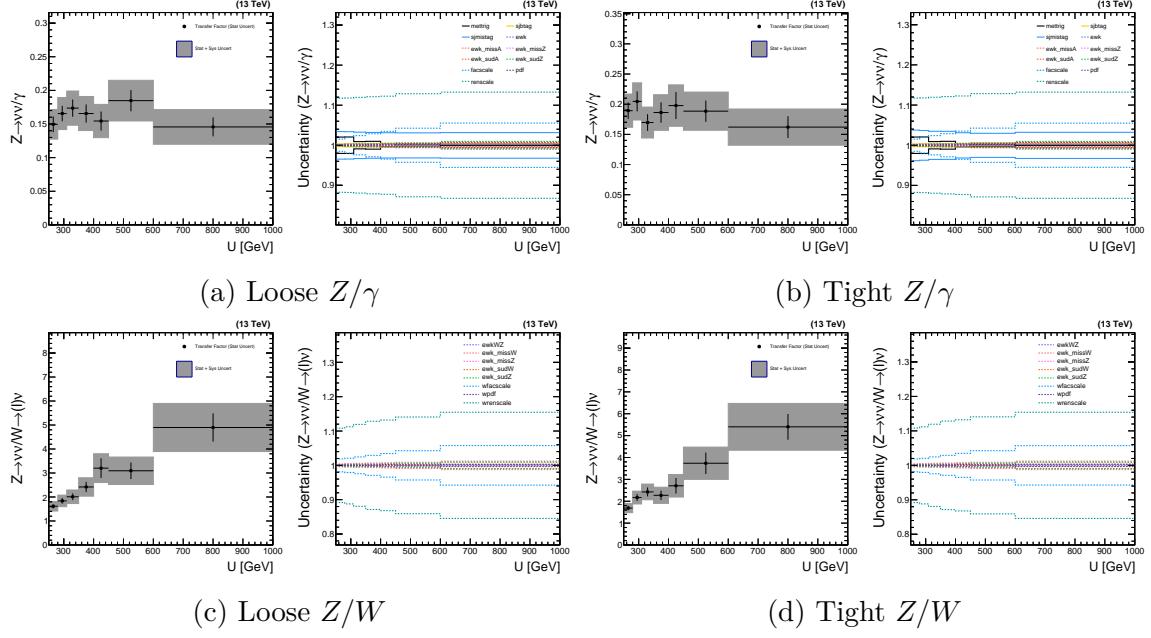


Figure 3.23: The transfer factors  $\mathbf{T}_\gamma^\gamma$  and  $\mathbf{T}_{Z/W}^{\text{SR}}$ ; and corresponding shape uncertainties.

This allows us to use the  $e, \mu$  CRs (which target  $W + \text{jet}$  production) to further reduce the uncertainty in the estimation of  $Z \rightarrow \nu\nu$  in the SR. For technical reasons, the transfer factor is defined as the ratio  $Z/W$  in the SR. However, the SR and the  $\mu$  CR are connected through a product of transfer factors:

$$N_i^{\text{SR}}(Z \rightarrow \nu\nu) = T_{Z/W,i}^{\text{SR}}(\hat{\boldsymbol{\theta}}) \times T_{W,i}^\mu(\hat{\boldsymbol{\theta}}) \times N_i^\mu(W \rightarrow \ell\nu) \quad (3.18)$$

where  $\hat{\boldsymbol{\theta}}$  is the maximum-likelihood estimate of  $\boldsymbol{\theta}$ .

Figure 3.23 shows these additional transfer factors and their shape uncertainties. It is clear from inspection that  $\mathbf{T}_\gamma^\gamma \ll 1$ , and the same holds for the effective transfer factor  $\mathbf{T}_{Z/W}^{\text{SR}} \times \mathbf{T}_W^\mu$ . This indicates that the CR data to which the transfer factor is linked has greater statistical power than the SR data.

Having included all of these components, the likelihood can be written as:

$$\begin{aligned}
\mathcal{L}(\mathbf{d} \mid \mu, \boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}, \boldsymbol{\mu}_{\text{SR}}^{t\bar{t}}, \boldsymbol{\theta}) = & \\
\prod_{i \in \text{bins}} \left[ \text{Pois} \left( d_i^{\text{SR}} \mid \mu S_i^{\text{SR}}(\boldsymbol{\theta}) + \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} + \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z/W,i}^{\text{SR}}(\boldsymbol{\theta})} + \mu_{\text{SR},i}^{t\bar{t}} + B_i^{\text{SR}}(\boldsymbol{\theta}) \right) \right. & \\
\times \prod_{X=\mu\mu, ee} \text{Pois} \left( d_i^X \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right) & \\
\times \prod_{X=b\mu, be} \text{Pois} \left( d_i^X \mid \frac{\mu_{\text{SR},i}^{t\bar{t}}}{T_{tt,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right) & \\
\times \prod_{X=\mu, e} \text{Pois} \left( d_i^X \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{W,i}^X(\boldsymbol{\theta}) T_{Z/W,i}^{\text{SR}}(\boldsymbol{\theta})} + \frac{\mu_{\text{SR},i}^{t\bar{t}}}{T_{tt,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right) & \\
\times \text{Pois} \left( d_i^\gamma \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{\gamma,i}^\gamma(\boldsymbol{\theta})} \right] \times \prod_{j=0}^{n_\theta} p_j(\theta_j) & \tag{3.19}
\end{aligned}$$

The discussion in this section has largely relied on arguments from simulation and calculation. We can, however, validate that our estimates of  $\mathbf{T}$  and the corresponding uncertainties are reasonable by using CR data and appropriate proxies. For example, to test  $\mathbf{T}_\gamma^\gamma \sim N(Z \rightarrow \nu\nu)/N(\gamma)$ , we can look at  $N(Z \rightarrow \mu\mu)/N(\gamma)$ . Up to differences in branching ratio and muon identification, these ratios are identical. Figure 3.24 shows some examples of these proxy ratios. In particular, we see that the  $Z/\gamma$ ,  $Z/W$  and  $W/\gamma$  ratios are well-described the MC, especially as compared to the systematic uncertainties that are assigned.

### 3.3 Results

Having built this likelihood, we perform a maximum likelihood fit to the data in all regions simultaneously. The results of a ML fit under a background-only hypothesis (i.e. setting  $\mu = 0$ ) is shown in Figures

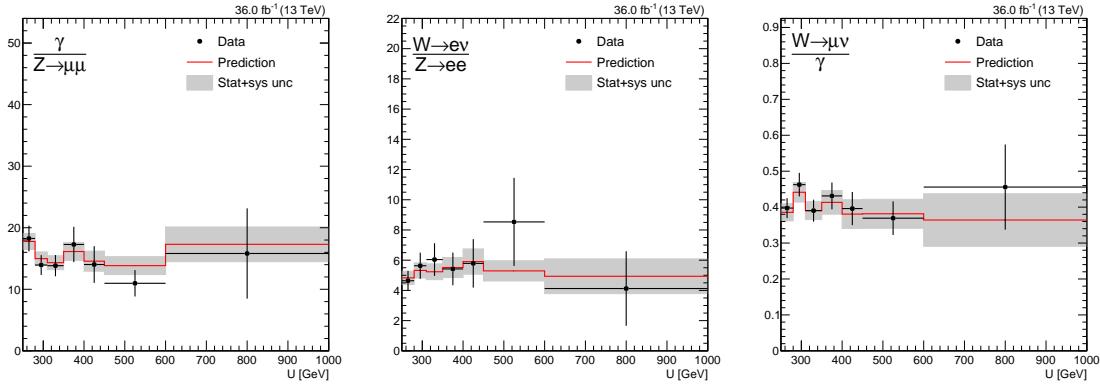


Figure 3.24: Data validation of CR-to-CR transfer factors in the loose category. Only ratios with theoretically-limited systematic uncertainties are shown.

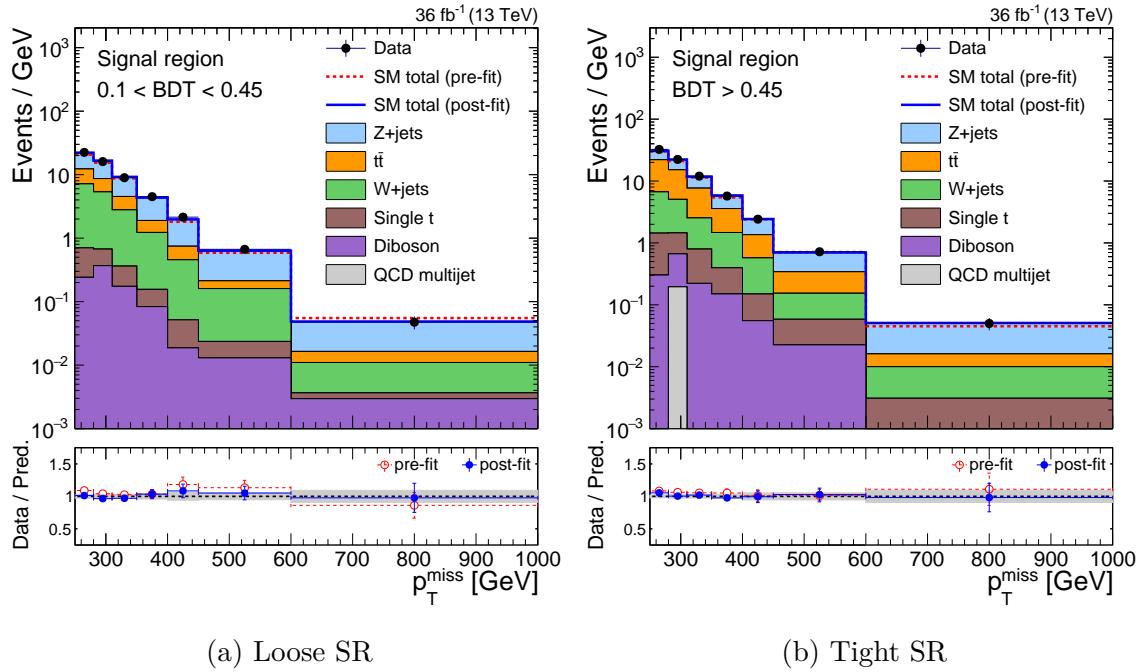


Figure 3.25: Comparison of pre- and post-fit results in the SRs, after simultaneously fitting all channels. The fit is performed having fixed  $\mu = 0$ .

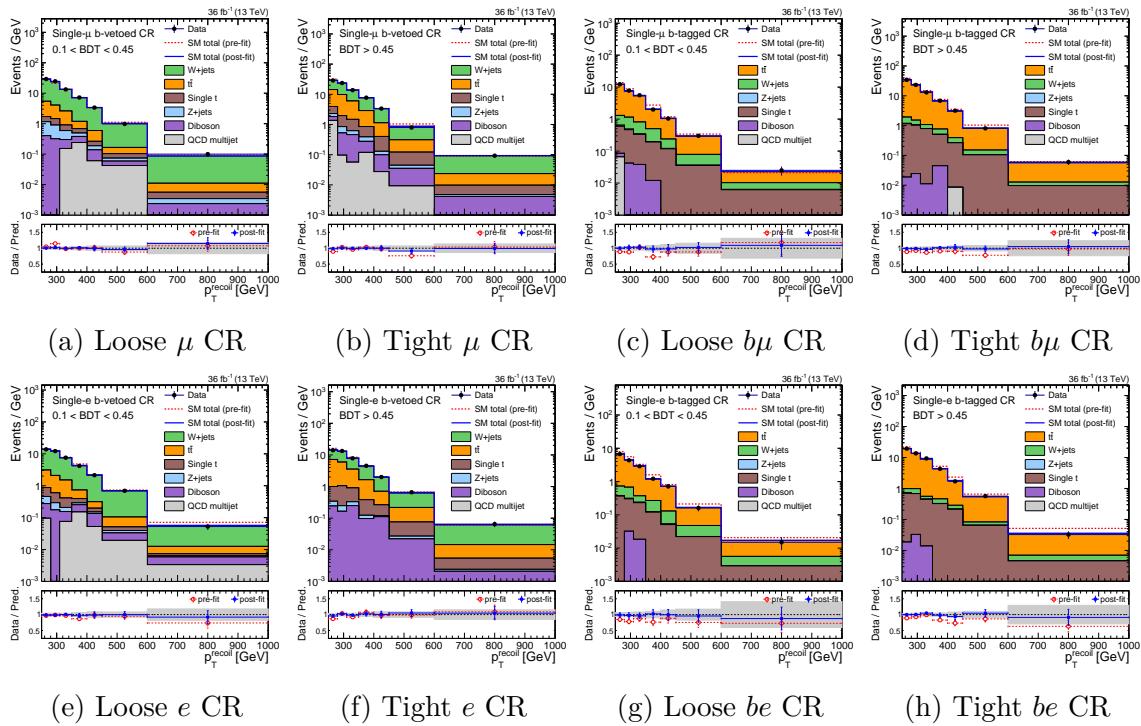


Figure 3.26: Comparison of pre- and post-fit results in the single-lepton CRs, after simultaneously fitting all channels. The fit is performed having fixed  $\mu = 0$ .

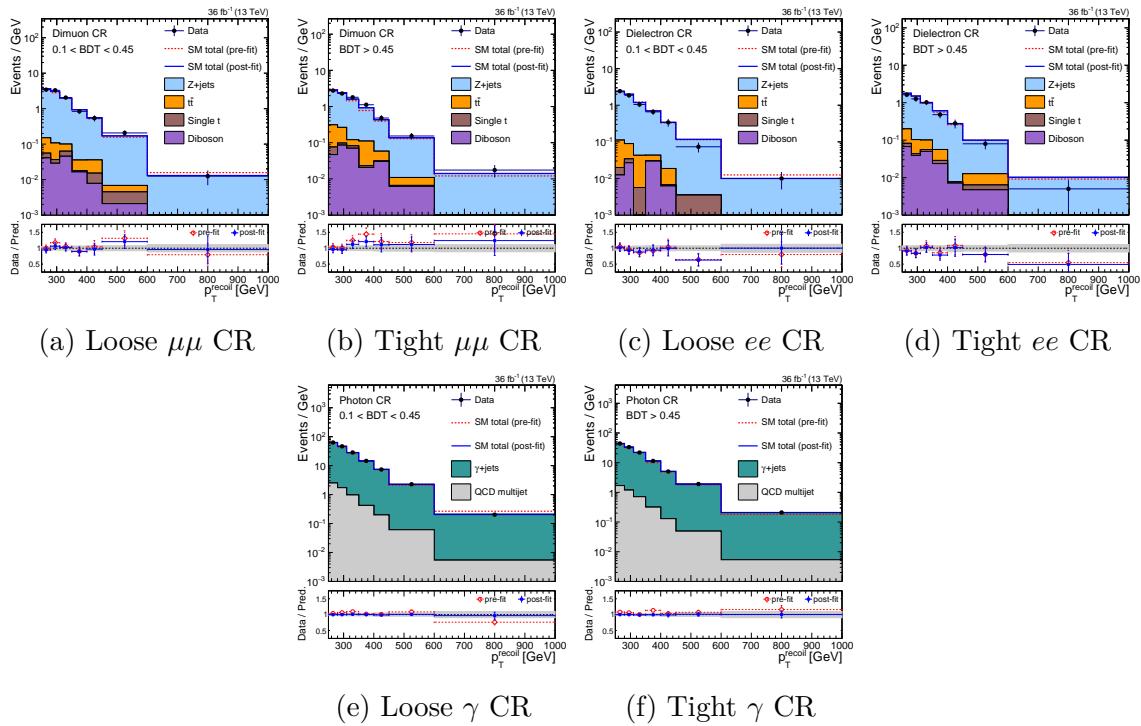


Figure 3.27: Comparison of pre- and post-fit results in the di-lepton and photon CRs, after simultaneously fitting all channels. The fit is performed having fixed  $\mu = 0$ .

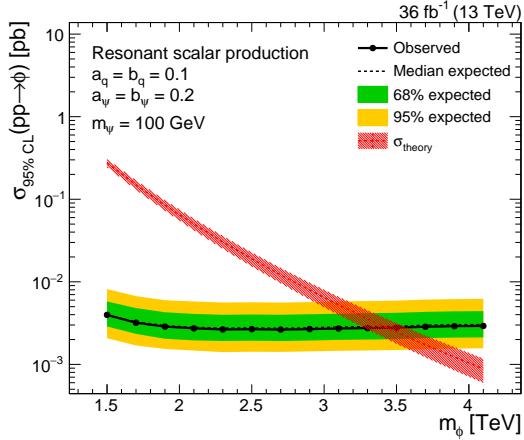


Figure 3.28: 95% CLs upper limits on the cross section of the production of  $\phi$ , where the resonance mass  $m_\phi$  is scanned. Values of  $m_\phi$  for which  $\sigma_{95\% \text{ CLs}} < \sigma_{\text{theory}}$  are excluded at 95% confidence level. The dashed black line represents the median expected exclusion, with the green and yellow bands representing the 1 and 2 standard deviation envelopes on the expected exclusion. The red band represents the 1 standard deviation of the uncertainty on the theory cross section (driven by higher-order QCD terms).

### 3.3.1 Constraints on mono-top models

The results of the previous section show that the background-only (B-only) model is able to describe the data. To quantify the preference of the B-only model to a signal+background model (S+B), we compute 95% CLs upper limits on the signal strength  $\mu$  for each signal hypothesis. A description of CLs upper limits and the asymptotic profile likelihood method is given in Section [?].

We begin with the simpler of the two models: the resonant scalar. Figure 3.28 shows the upper limit on the cross section of  $\phi$  production as a function of  $m_\phi$ . A fixed value of  $m_\psi$  is chosen such that  $m_\psi \ll m_\phi$ . In this regime, the exact value of  $m_\psi$  does not strongly affect the kinematics or cross section; in the opposite regime, the decay  $\phi \rightarrow t\bar{\psi}$  is strongly suppressed. The values of the couplings  $a_q, b_q$  and  $a_\psi, b_\psi$  are similarly fixed to reasonable values. The cross section (and therefore the signal strength  $\mu$ ) roughly scales as  $(a_q + b_q)^2$  and  $(a_\psi + b_\psi)^2$ . Given these assumptions, the observed data excludes scalars with mass  $m_\phi < 3.4$  TeV. This can be compared to the expected exclusion, which is also 3.4 TeV.

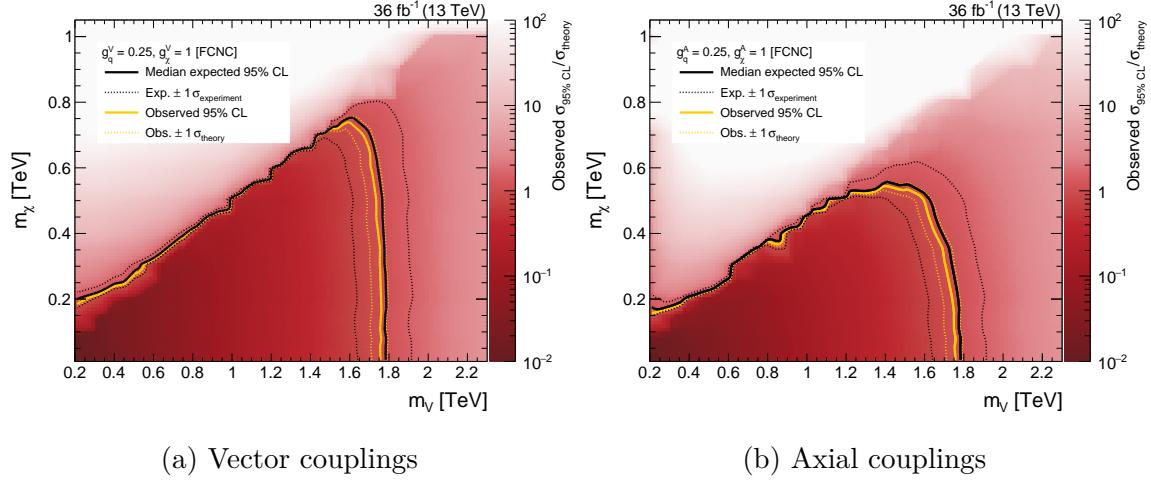


Figure 3.29: 95% CLs upper limits on the signal strength ( $\mu = \sigma/\sigma_{\text{theory}}$ ) of the flavor-violating process  $pp \rightarrow t\chi\bar{\chi}$  as a function of  $m_V$  and  $m_\chi$ . Two hypotheses are tested: assuming  $g_\chi^V, g_q^V \neq 0$  (vector-like) and assuming  $g_\chi^A, g_q^A \neq 0$  (axial vector-like).

The FCNC model contains many more non-trivial parameters, and so we do not simply fix all but one. First, fixing the couplings (at values agreed upon by the LHC Dark Matter Working Group in Reference [44]), we show the upper limits as a function of  $(m_V, m_\chi)$  in Figure 3.29. Both vector-like and axial vector-like couplings are probed independently. In either scenario, assuming  $m_\chi < 50$  GeV, we are able to exclude  $0.2 < m_V < 1.75$  TeV. The lower bound arises from measurements of  $m_t$  and  $\Gamma_t$ ; allowing  $m_V \lesssim 200$  GeV modifies  $\Gamma_t$  beyond measured bounds as the  $t^{(*)} \rightarrow Vu$  channel opens. As  $m_\chi$  crosses the  $m_V/2$  boundary,  $\sigma_{\text{theory}}$  drops off rapidly, reducing the strength of the exclusion. In the vector case, this defines a clear exclusion triangle bounded by  $m_V < 1.75$  TeV and  $2m_\chi < m_V$ . In the axial-vector case,  $\Gamma_V$  is much narrower, modifying the transition to the off-shell region. On the entire plane, the observed exclusion is consistent with the median expected exclusion within one standard deviation.

Sticking to two-dimensional projections of the parameter space, Figure 3.30 shows the excluded regions as a function of mediator mass  $m_V$  and all four free couplings  $g_q^V, g_q^A, g_\chi^V, g_\chi^A$ . As the DM mass is fixed to be strictly on-shell regardless of  $m_V$ , there are no visible differences between the vector and axial-vector scenarios. Assuming

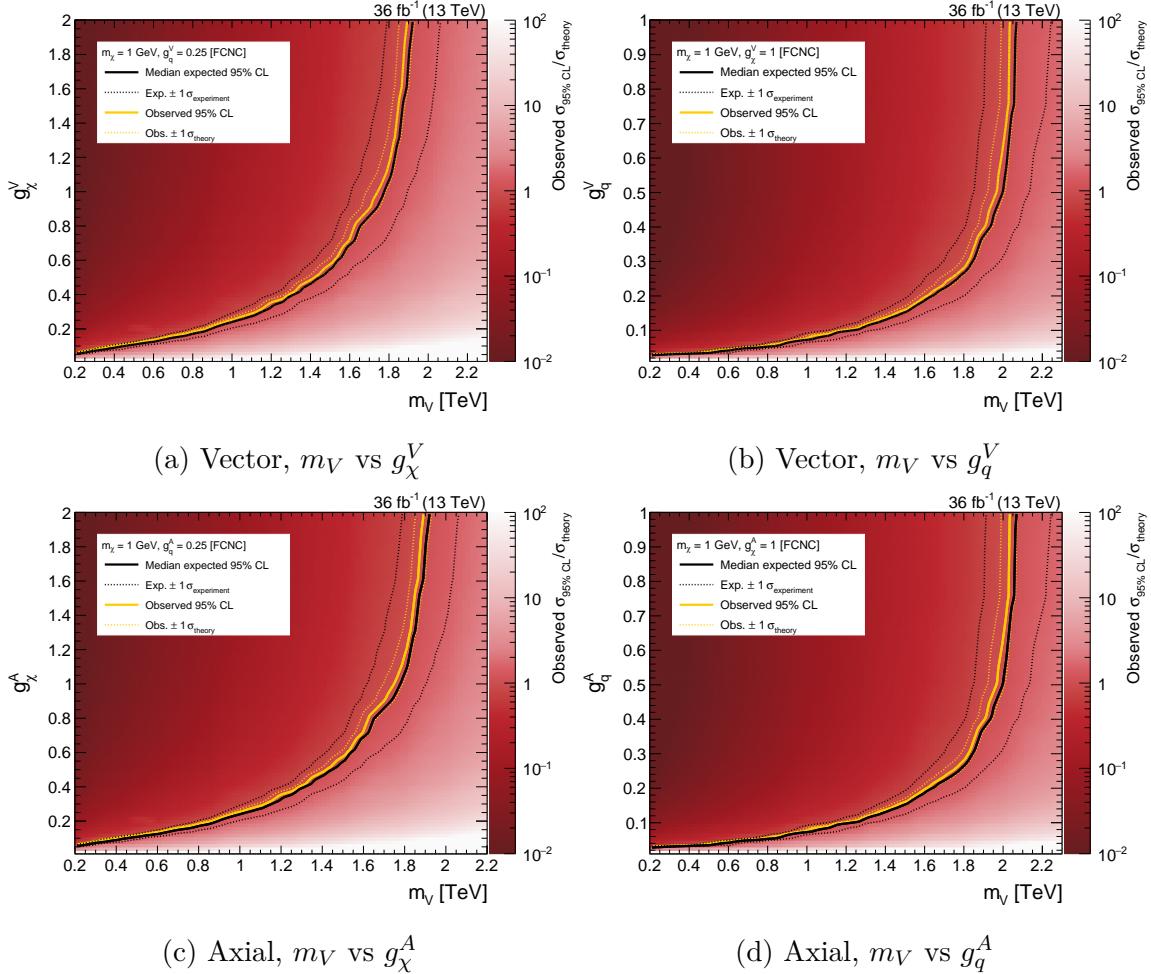


Figure 3.30: 95% CLs upper limits on the signal strength ( $\mu = \sigma/\sigma_{\text{theory}}$ ) of the flavor-violating process  $pp \rightarrow t\chi\bar{\chi}$  as a function of  $m_V$ ,  $g_q$ , and  $g_\chi$ . Two hypotheses are tested: assuming  $g_\chi^V, g_q^V \neq 0$  (vector-like) and assuming  $g_\chi^A, g_q^A \neq 0$  (axial vector-like). The DM mass  $m_\chi$  is fixed at 1 GeV.

sufficiently low masses (still corresponding to viable hypotheses), we exclude couplings as weak as  $g_q \cdot g_\chi \sim 0.01$ .

It is difficult to fully visualize more than two dimensions of the parameter space at a time. Figure 3.31 shows the largest mediator mass that is excluded as a function of  $g_q^V, g_\chi^V$  (assuming  $g_q^A = g_\chi^A = 0$ ). Given sufficiently large couplings, we exclude FCNCs as massive as 2.5 TeV, while only assuming that  $2m_\chi < m_V$ .

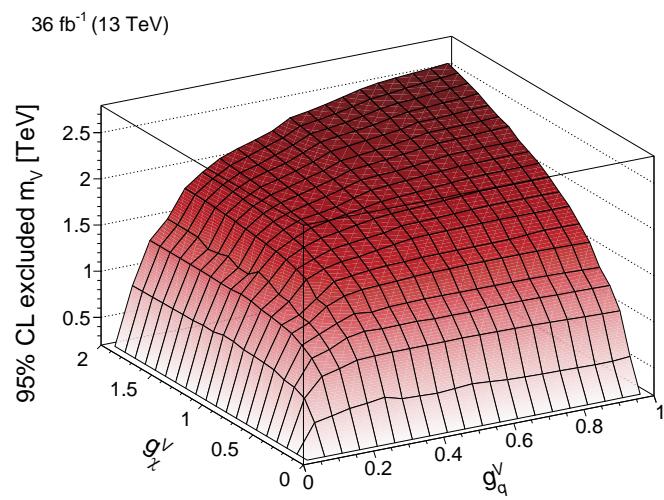


Figure 3.31: Maximum excluded value of  $m_V$  for each set of couplings, assuming vector-only couplings.



# Chapter 4

## The Search for Invisible Decays of the Higgs Boson

The discovery of the SM Higgs boson [?] involved multiple production modes. Gluon fusion has the largest cross section (49 pb) at the LHC because of the large gluon PDF, followed by vector boson fusion (VBF) (3.8 pb),  $WH$  (1.4 pb), and  $ZH$  (0.89 pb) [45]. While gluon fusion is the most frequent mode, the unique detector signatures of the other production modes can be combined the various Higgs decay signatures to define a signal topology with few backgrounds.

Many DM models [?] allow for DM fermions or scalars to acquire mass through the Higgs mechanism, coupling to the SM Higgs boson. If the DM candidate  $\chi$  satisfies  $2m_\chi < m_H$ , then we expect to observe  $H \rightarrow \chi\bar{\chi}$ . From measurements of the visible branching fractions, we can indirectly place an upper bound of  $\mathcal{B}(H \rightarrow \chi\bar{\chi}) < 0.2$  [?]. In this chapter, we describe a direct search for  $H \rightarrow \chi\bar{\chi}$  decays.

As with the case of the mono-top search, the  $H \rightarrow \chi\bar{\chi}$  process manifests as  $p_T^{\text{miss}}$ . Each of the aforementioned Higgs production modes translates into a  $p_T^{\text{miss}} + X$  signature, where  $X$  refers to one or more SM particles. Figure 4.1 shows each of the signatures; in this chapter, we will focus on the VBF production mode, as the unique final state topology provides the best sensitivity to  $H \rightarrow \chi\bar{\chi}$ .

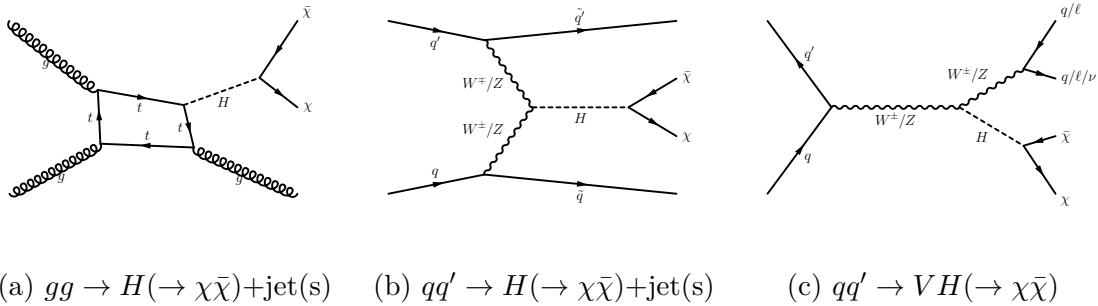


Figure 4.1: Diagrams that contribute to the production of the SM Higgs boson at the LHC, with the subsequent decay to DM candidates. The shown diagrams are all chosen to generate large  $p_T^{\text{miss}}$  through the presence of one or more SM particles in the final state.

## 4.1 Signal selection

VBF  $H \rightarrow \chi\bar{\chi}$  events are characterized by large  $p_T^{\text{miss}}$  and two jets. These jets are typically:

- Fairly forward in the detector
- Far apart from each other in  $\eta$
- Highly energetic (large  $E$ , moderate  $p_T$ )
- Close together in  $\phi$

A candidate VBF  $H \rightarrow \chi\bar{\chi}$  event displaying these properties is shown in a CMS event display in Figure 4.2.

### 4.1.1 Online trigger selection

The same trigger decisions (L1 and HLT) as described in Section 3.1.1 are used to select events in this analysis. However, the L1 seeds for the 2016 data run were designed with mono-top-like analyses in mind; i.e., searches where the momentum imbalance is created by central objects. To avoid noise and resolution issues in the forward calorimeters, the L1 seed only considers energy deposits in the region  $|\eta| < 3$ . Therefore, VBF events in which both jets are in the forward region are not selected.

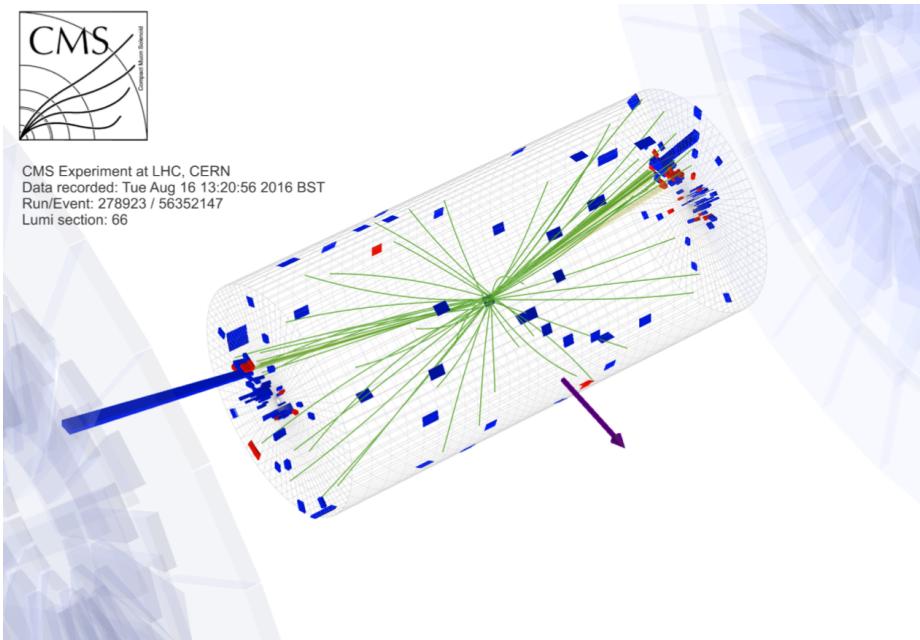


Figure 4.2: Candidate VBF  $H \rightarrow \chi\bar{\chi}$  event with two energetic forward jets ( $p_T = 180, 107$  GeV) and large  $p_T^{\text{miss}}$  (360 GeV). Red (blue) towers represent deposits in the hadronic (electromagnetic) calorimeter. Green lines are tracks reconstructed from hits of charged particles in the tracker. The blue arrow represents the direction and magnitude of the  $p_T^{\text{miss}}$ .

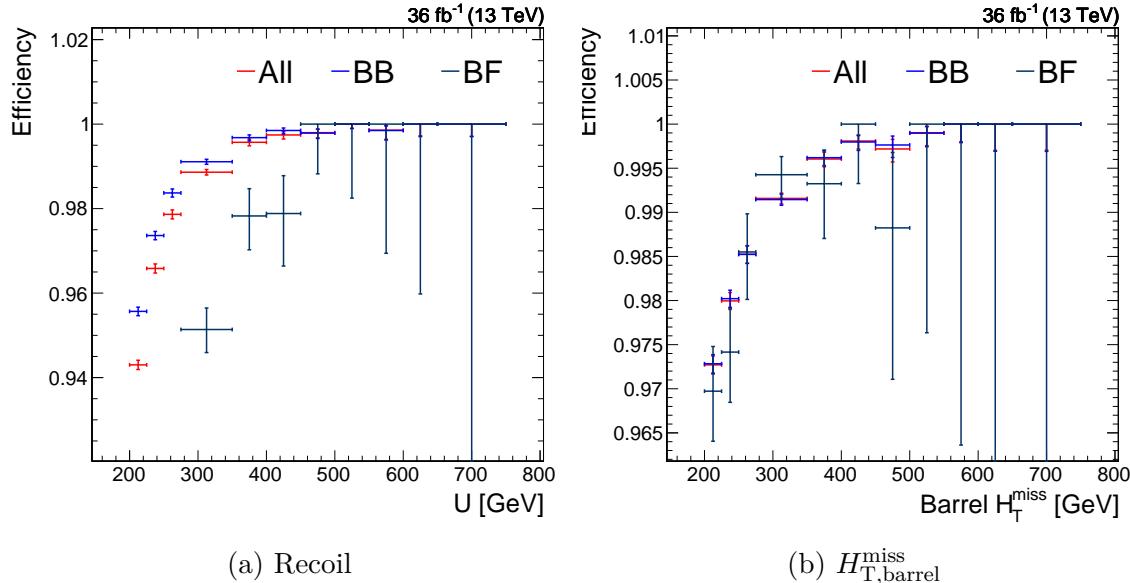


Figure 4.3: Trigger efficiency of events with a VBF-like topology (two jets with  $p_T > 80, 40$  GeV) as a function of two different observables. Events are split into two categories: those where both jets have  $|\eta| < 3$  (BB) and those where exactly one jet has  $|\eta| > 3$  (FF). "All" refers to the sum of these categories.

This is visible in Figure 4.3a, where events are classified based on the location of the two highest- $p_T$  jets. Events with both jets in the barrel (BB) have a higher efficiency than events with one jet in the forward detector (BF). Note that events with two forward jets (FF) are not considered at all, as the efficiency for such events is essentially zero.

The trigger efficiency is truly characterized by the energy deposited in the  $|\eta| < 3$  region of the detector, and will be dominated in VBF events by the energy of jets. Accordingly, we define the “missing barrel hadronic transverse momentum”:

$$H_{\text{T,barrel}}^{\text{miss}} = \left| \left( \sum_{j \in \text{barrel}} \vec{p}_j \right)_{\text{T}} \right|, \text{ where barrel refers to jets with } |\eta| < 3 \quad (4.1)$$

As shown in Figure 4.3b, the three categories (BB, BF, All) have similar behavior as a function of  $H_{T,\text{barrel}}^{\text{miss}}$ . Therefore, we use this parameterization of the efficiency to correct MC simulation to match data.

A second L1-related issue that plagues the 2016 dataset is caused by a “pre-firing”

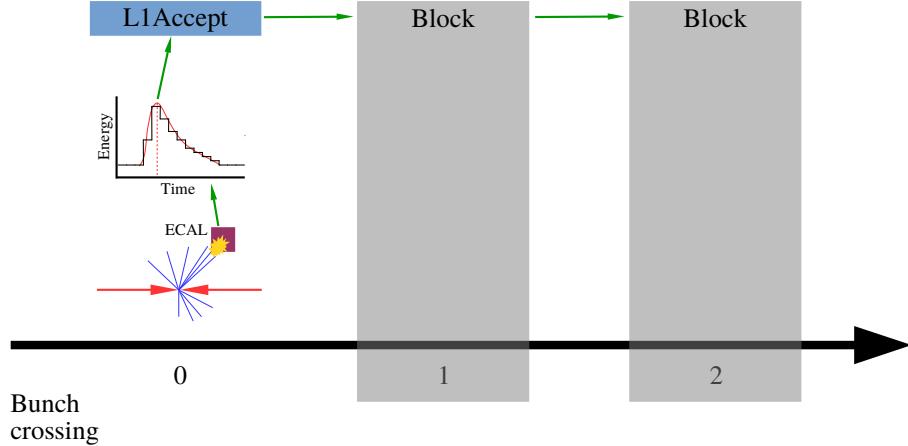


Figure 4.4: A normal event in which an ECAL seed triggers the L1A signal. The subsequent two bunch crossings are blocked. BX<sub>0</sub> refers to the event containing the physics object of interest. Green arrows indicate causality.

effect. When an L1 seed is triggered to accept an event (“L1Accept” or “L1A”), the following two bunch crossings (not necessarily corresponding to collisions) are blocked from firing L1As. At most, two in four consecutive events can fire L1A (i.e. the sequence L1A, blocked, blocked, L1A). Figure 4.4 is an example of a normal ECAL L1 seed accepting an event and blocking the subsequent bunch crossings. In what follows, we will refer to the bunch crossing with an interesting collision (i.e. the one we would like the trigger to select) as BX<sub>0</sub>.

A pre-fire refers to the case in which a malformed detector signal is mis-reconstructed, so that the peak of the pulse appears to have occurred in the previous bunch crossing (BX<sub>-1</sub>). In this particular case, a region of the ECAL ( $2.5 < |\eta| < 3$ ) suffered from a loss in transparency due to radiation damage and would produce pulse shapes that are poorly described by the model used to extract the pulse energy and time. When this happens, the L1 seeds for ECAL-based signatures (e.g. electron triggers) can fire an L1A for BX<sub>-1</sub>. This ECAL L1 seed in BX<sub>-1</sub> will zero out the corresponding ECAL clusters in BX<sub>0</sub> (known as zero suppression), further biasing the event description. So, we would have an L1A for an arbitrary event (BX<sub>-1</sub>), and the interesting event (BX<sub>0</sub>) would be blocked from passing the L1 altogether. This is depicted in Figure 4.5. Typically, BX<sub>-1</sub> contains uninteresting physics signatures, and so is not

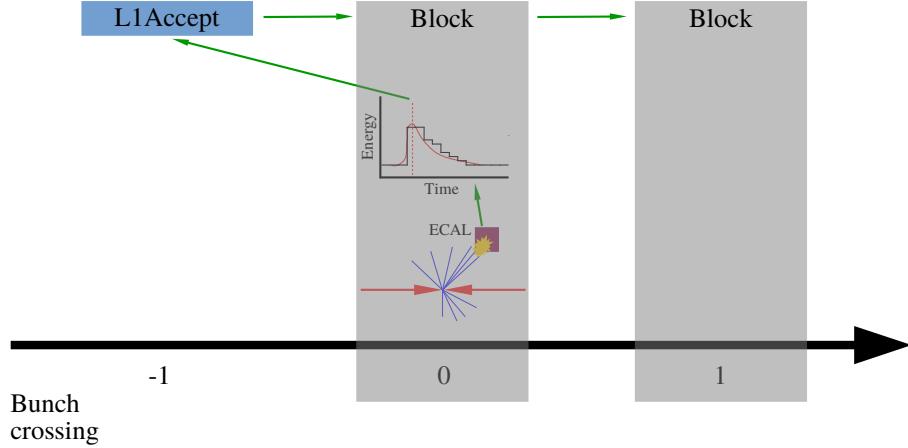


Figure 4.5: A pre-fired event in which an ECAL seed triggers the L1A signal for  $\text{BX}_{-1}$ . The subsequent two bunch crossings (including the one of interest) are blocked.  $\text{BX}_0$  refers to the event containing the physics object of interest. Green arrows indicate causality.

accepted by the HLT.

To measure how often an ECAL energy deposit (typically left by a jet) causes an event to be lost by pre-firing, we need to compute the following efficiency:

$$\epsilon_{\text{pre-fire}}(p_T, \eta, \phi) = \frac{N_{\text{pre-fire}}(p_T, \eta, \phi)}{N_{\text{all events}}(p_T, \eta, \phi)} \quad (4.2)$$

However, by definition, pre-fired events cannot be recorded, and therefore  $N_{\text{pre-fire}}(p_T, \eta, \phi)$  is difficult to measure. A very small subset of the recorded dataset (0.2%) consists of “un-pre-fireable” events. These are recorded events ( $\text{BX}_0$ ) in which an L1A fired 3 bunch crossings prior ( $\text{BX}_{-3}$ ). Due to the blocking rules, L1A cannot fire in  $\text{BX}_{-2}$  and  $\text{BX}_{-1}$ . Even if there is an ECAL seed in  $\text{BX}_0$  that pre-fires, it will be blocked from firing an L1A, and therefore  $\text{BX}_0$  is protected. If some other object in  $\text{BX}_0$  manages to pass L1 and HLT decisions, then  $\text{BX}_0$  will be recorded and can be studied. A schematic of such events is shown in Figure 4.6.

The L1 trigger system records trigger primitive (TP) information (4-vectors of physics objects considered in an L1 selection) for  $\text{BX}_{-1}$  if  $\text{BX}_0$  is triggered. This means we can identify the cases in which a physics object in  $\text{BX}_0$  coincides with a TP in  $\text{BX}_{-1}$ , indicating a pre-fire. Therefore (using the bunch crossing numbering in

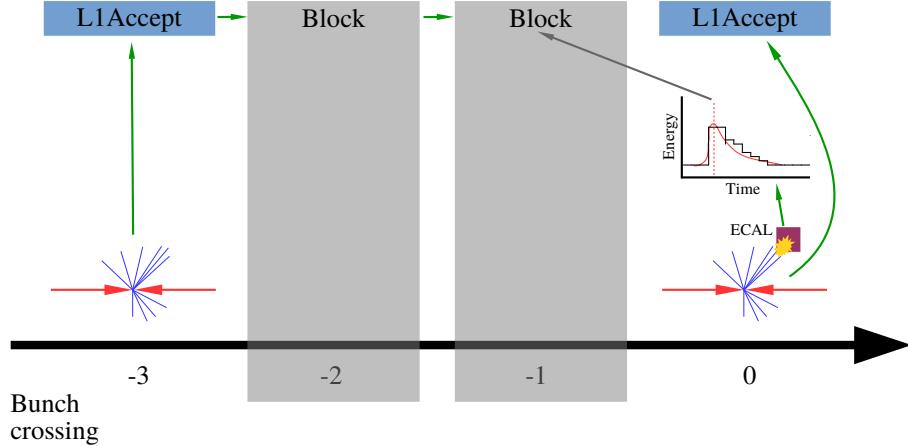


Figure 4.6: An un-pre-fireable event in which  $\text{BX}_{-3}$  protects  $\text{BX}_0$  from being pre-fired. Green arrows indicate causality.

Figure 4.6), we re-define the efficiency:

$$\epsilon_{\text{pre-fire}}(p_T, \eta, \phi) = \frac{N_{\text{pre-fire } \text{BX}_0 | \text{BX}_{-3}}(p_T, \eta, \phi)}{N_{\text{BX}_0 | \text{BX}_{-3}}(p_T, \eta, \phi)} \quad (4.3)$$

By definition, all events in this ratio will be recorded. Figure 4.7 shows this efficiency as a function of jet location. We observe there is a “hot” ECAL tower near the location  $\eta = -2.8$  and  $\phi = 2$ . Not only does this tower fire very frequently (leading to many particles, leading to many jets), but it almost always pre-fires. To first order, events with a jet in this crystal should be rejected. Beyond this, there is very little localization in the pre-fire probability (besides restriction to the ECAL endcap).

In Figure 4.8 we see  $\epsilon_{\text{pre-fire}}$  as a function of  $p_T$  in a restricted  $\eta$  range. Firstly, we observe that  $\epsilon_{\text{pre-fire}}$  increases as a function of  $p_T$ , and the turn-on is sharper as a function of EM  $p_T$ . This is explained by the mechanism of the pre-fire: the individual ECAL trigger seeds have a threshold of 30 GeV. The higher the jet  $p_T$ , the higher the probability of the jet depositing 30 GeV of EM energy in a localized area, setting off an L1 seed. Secondly, we observe a strong dependence on the reference triggers used to select  $\text{BX}_0$ . For example, jet-based triggers (JetHT) lead to a much higher efficiency than  $p_T^{\text{miss}}$ -based triggers (MET). This is a consequence of zero suppression biasing the  $\text{BX}_0$  triggers, as shown diagrammatically in Figure 4.9. Muon-based

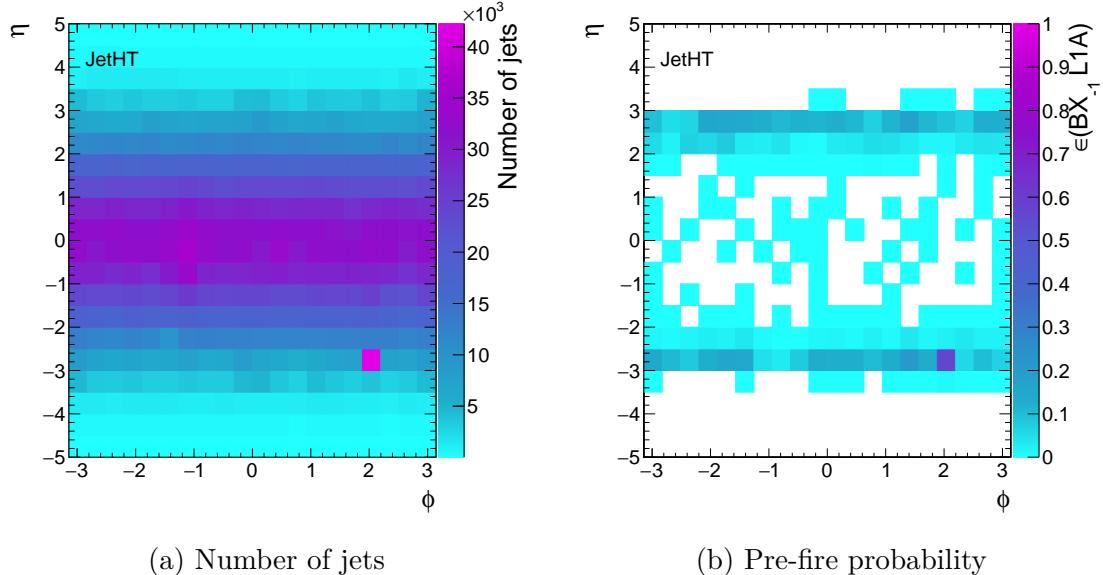


Figure 4.7: Distribution of jets and pre-fire events as a function of the jet location in the detector. Note the spike near  $(\eta, \phi) = (-2.8, 2)$ .

triggers (SingleMuon) are largely unaffected by the ECAL system, and therefore this measurement of  $\epsilon_{\text{pre-fire}}$  is the least biased.

The probability of at least one jet pre-firing in an event is:

$$\epsilon_{\text{pre-fire}}^{\text{event}} = 1 - \prod_{j \in \text{jets}} \left( 1 - \epsilon_{\text{pre-fire}}(p_T^j, \eta^j) \right) \quad (4.4)$$

The  $\phi$ -dependence has been dropped, since it is clear from Figure 4.7 that the effect can be averaged over  $\phi$  (once the spike is removed). Figure 4.10 shows  $\epsilon_{\text{pre-fire}}(p_T, \eta)$  using muon-triggered and jet-triggered events. In the former case, statistical fluctuations make the region with  $p_T > 250$  GeV unusable. Fortunately, this is the region in which the trigger bias is smallest, and so we switch to the jet triggered measurement above this threshold. A 20% uncertainty is assessed on the efficiency, which is derived from the difference between the SingleMuon and JetHT measurements.

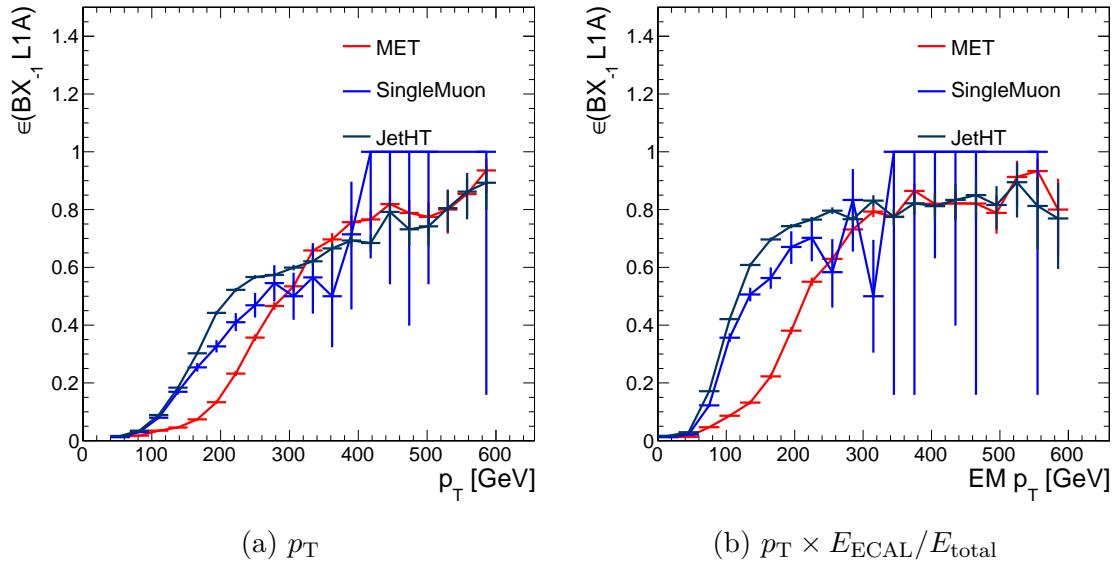


Figure 4.8: Probability that a given jet with  $2.25 < |\eta| < 3$  causes a pre-fire in the L1 trigger due to ECAL mistiming. Two parameterizations are used: jet  $p_T$  and EM  $p_T$ . The three curves refer to which set of triggers are used to select BX<sub>0</sub>.

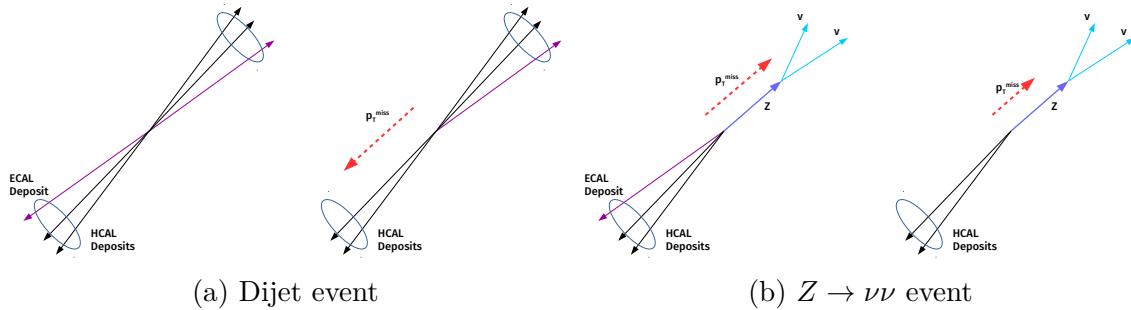


Figure 4.9: How zero suppression in the ECAL due to pre-firing can bias certain events. Subfigure (a) shows a dijet event, in which the loss of an ECAL deposit reduces the total  $H_T$  of the event, thereby lowering the probability of a jet-based trigger to fire. Subfigure (b) shows a  $Z \rightarrow \nu\nu$  event, in which the loss of an ECAL deposit reduces the total  $p_T^{\text{miss}}$  of the event.

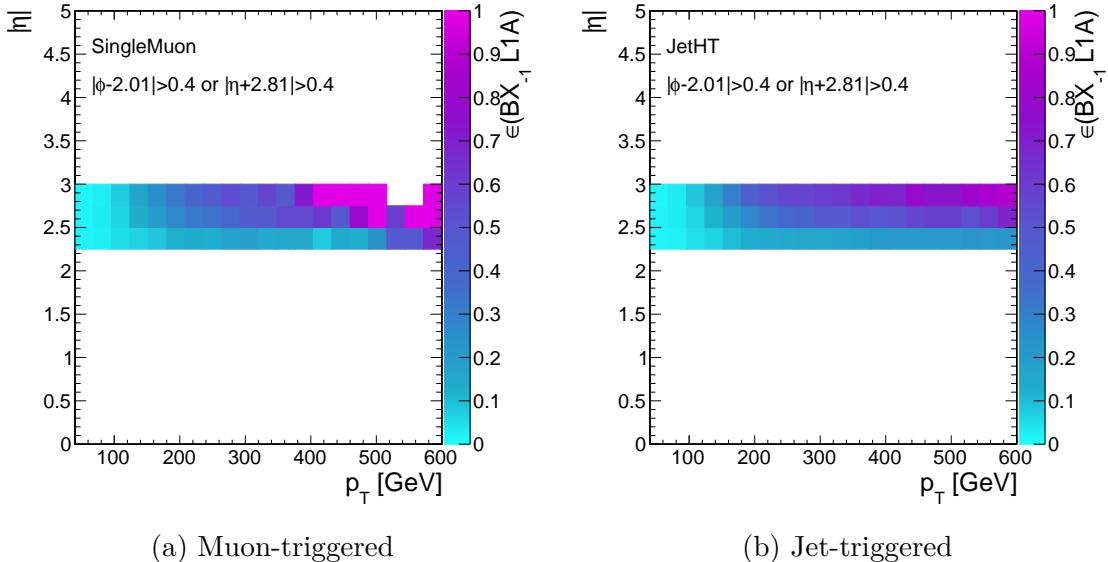


Figure 4.10:  $\epsilon_{\text{pre-fire}}(p_T, \eta)$  with two different sets of reference triggers used to select  $\text{BX}_0$ .

### 4.1.2 EW and QCD production of electroweak bosons

The primary backgrounds to the VBF production of invisibly-decaying Higgs bosons are  $Z(\rightarrow \nu\nu)+2 \text{ jet}$  and  $W(\rightarrow \ell\nu)+2 \text{ jet}$  production. At leading order, the relevant Feynman diagrams are either of the order  $\alpha_{\text{EW}}^2 \alpha_{\text{QCD}}^4$  or  $\alpha_{\text{EW}}^6$ . We refer to the former as the QCD production mode and the latter as the EW mode. Examples Feynman diagrams are shown in Figure 4.11. The EW mode is essentially vector boson fusion, and so the terms EW and VBF will be used interchangeably.

As the vector boson is not directly detectable, the only experimental signatures are the jets. The jet kinematics are sensitive to the production mode (vector boson fusion vs QCD), as well as the spin of the produced boson. Some conclusions can be drawn from the kinematic distributions (Figure 4.12):

1. The yield ( $\sigma \times A$ ) of the three VBF processes are relatively close in the relevant phase space (assuming  $\mathcal{B}(H \rightarrow \chi\bar{\chi}) = 1$ ), but the QCD processes are 1-2 orders of magnitude higher.
2. The jet  $p_T$  and  $p_T^{\text{miss}}$  distributions in the signal are comparable to or softer than

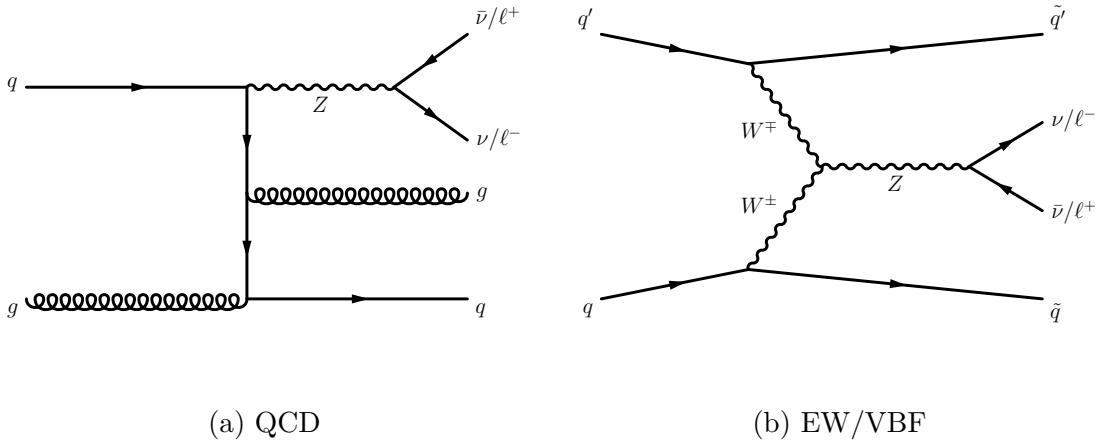


Figure 4.11: Examples of the two modes of producing  $Z$  bosons in association with 2 jets. Similar diagrams exist for  $W$  boson production.

the background processes. This is in contrast to other DM searches, in which the signal  $p_T^{\text{miss}}$  distribution is much harder than SM predictions.

3. VBF  $H \rightarrow \chi\bar{\chi}$  produces fewer jets than SM processes.
  4. VBF  $H \rightarrow \chi\bar{\chi}$  produces relatively forward jets. QCD  $V+jets$  produces mostly central jets. VBF  $V+jets$  produces jets that are somewhere between these distributions.

To fully exploit these kinematic distributions, we look at “VBF-tag” observables, which are functions of the two leading jets. These are defined as:

$m_{jj}$ : Invariant mass of the dijet system.

$\Delta\eta_{jj}$ : Absolute value of the difference in pseudorapidity of the two jets.

$\Delta\phi_{jj}$ : Absolute value of the difference in azimuthal angle of the two jets.

These distributions are shown in Figure 4.13. The first two distributions look different in QCD and VBF processes and are therefore useful to reduce QCD backgrounds. On the other hand,  $\Delta\phi_{jj}$  is sensitive to the spin of the boson produced in a VBF process, and therefore can distinguish between Higgs and electroweak boson production.

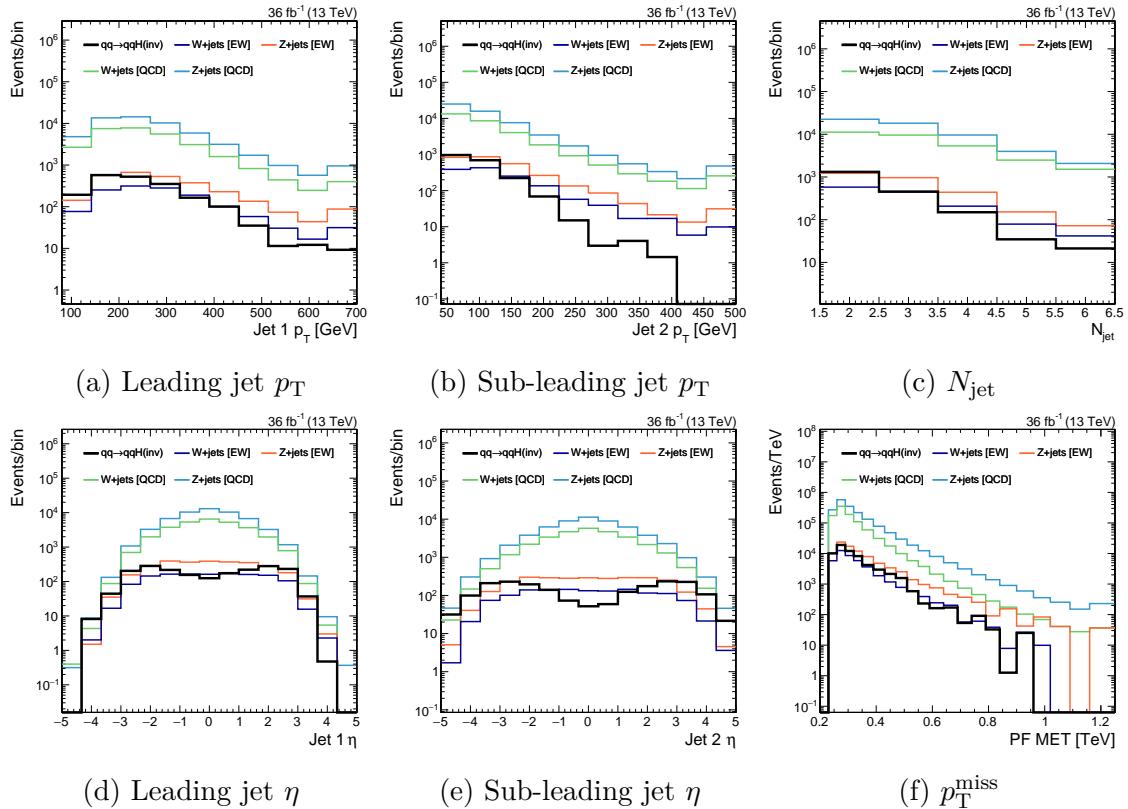


Figure 4.12: Event kinematic distributions, as compared between  $H$  vs  $Z$  vs  $W$  production, and VBF vs QCD modes.

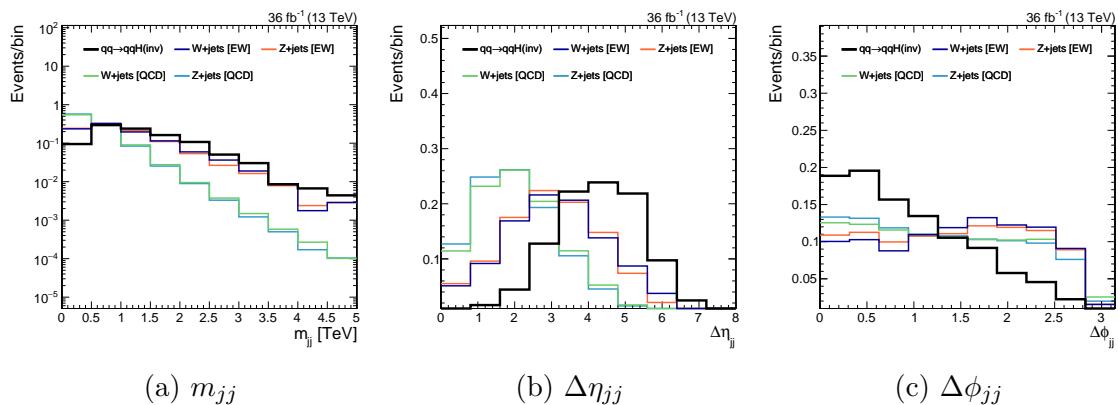


Figure 4.13: VBF tag observable distributions, as compared between  $H$  vs  $Z$  vs  $W$  production, and VBF vs QCD modes.

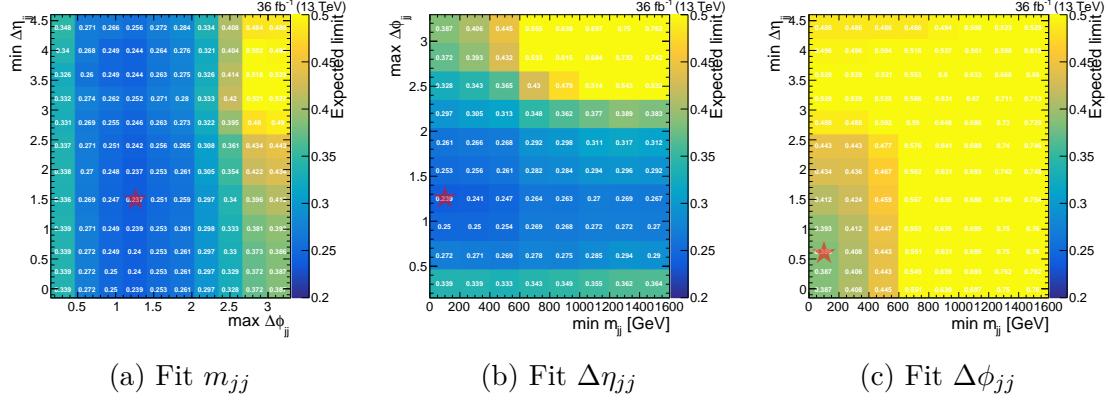


Figure 4.14: Optimization of the dijet kinematic selection, in three different fitting distribution scenarios.

### 4.1.3 Sensitivity optimization

A “baseline” selection is defined as:

- $p_T^{\text{miss}} > 250$  GeV: driven by trigger efficiency, as discussed in Section 4.1.1.
- $p_T^{\text{jet}} > 80, 40$  GeV: require two VBF jets, lower  $p_T$  thresholds set by trigger efficiency
- $N_{e,\mu,\tau,\gamma} = 0$ : veto leptonic decays of  $Z$  and  $W$ ,  $t\bar{t}$ , diboson production,  $\gamma$ +jet, etc.
- $\min \Delta\phi(\text{jet}, p_T^{\text{miss}}) > 0.4$ : remove QCD multijet events.
- $|p_{T,\text{calo}}^{\text{miss}} - p_T^{\text{miss}}| < p_T^{\text{miss}}/2$ : remove miscalibrated events.

As the tag variables each show some level of separation between signal and backgrounds, we can choose to either fit the distributions or use them to select events. To find the optimal choice, we fit each of the distributions in turn, and scan the other two observables. The details of this fit and the background estimation are described in Section 4.2. The metric is chosen to be the expected 95% CLs upper limit on  $\mathcal{B}(H \rightarrow \chi\bar{\chi})$ . Figure 4.14 shows the result of this optimization. The dijet mass is found to be the best distribution to fit, while requiring  $\Delta\eta_{jj} > 1$  and  $\Delta\phi_{jj} < 1.5$ .

## 4.2 Background estimation

To estimate the combined  $m_{jj}$  spectra of the EW and QCD  $V$ +jet backgrounds, we employ a similar visible-to-invisible strategy as described in Section 3.2. In this case, the transfer factors  $\mathbf{T}$  are a function of  $m_{jj}$ . Control regions are defined using dilepton (single-lepton) selections to estimate the  $Z$  ( $W$ ) contributions. Again,  $p_T^{\text{miss}}$  is replaced by  $U$  (Equation 3.7) to mimic the signal region selection. Because there are *two* components to estimate in each CR (QCD and EW), we slightly modify the likelihood. Adding only the  $\mu\mu$  CR to constrain the  $Z$ +jet component for now:

$$\begin{aligned} \mathcal{L}(\mathbf{d} | \mu, \boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}, \boldsymbol{\theta}) = \prod_{i \in \text{bins}} & \left[ \text{Pois} \left( d_i^{\text{SR}} \mid \mu S_i^{\text{SR}}(\boldsymbol{\theta}) + \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} + \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^{\text{QE}}(\boldsymbol{\theta})} + B_i^{\text{SR}}(\boldsymbol{\theta}) \right) \right. \\ & \times \text{Pois} \left( d_i^{\mu\mu} \mid \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^{\mu\mu}(\boldsymbol{\theta})} + \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^{\mu\mu}(\boldsymbol{\theta}) T_{Z,i}^{\text{QE}}(\boldsymbol{\theta})} + B_i^{\mu\mu}(\boldsymbol{\theta}) \right) \left. \right] \\ & \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \end{aligned} \quad (4.5)$$

While the notation largely follows that used in Equation 3.11, one additional term has been introduced. This is a “transfer factor” linking the QCD and EW components in the signal region, so that the only free parameter is  $\boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}$ :

$$T_{Z,i}^{\text{QE}} = \frac{N_i^{\text{SR}}(\text{QCD } Z \rightarrow \nu\nu)}{N_i^{\text{SR}}(\text{EW } Z \rightarrow \nu\nu)} \quad (4.6)$$

where as always, the yields  $N$  are predicted using MC. Kinematic distributions from the two dilepton CRs are shown in Figure 4.15.

In the region  $m_{jj} > 2.5$  TeV, the statistical power of the dilepton regions is extremely limited. For this reason, and to estimate the  $W$ +jets contribution in the SR, we add two single-lepton CRs in analogy to what is done in Section 3.2. Figure 4.16 shows the level of agreement between the data and MC in these CRs. The likelihood

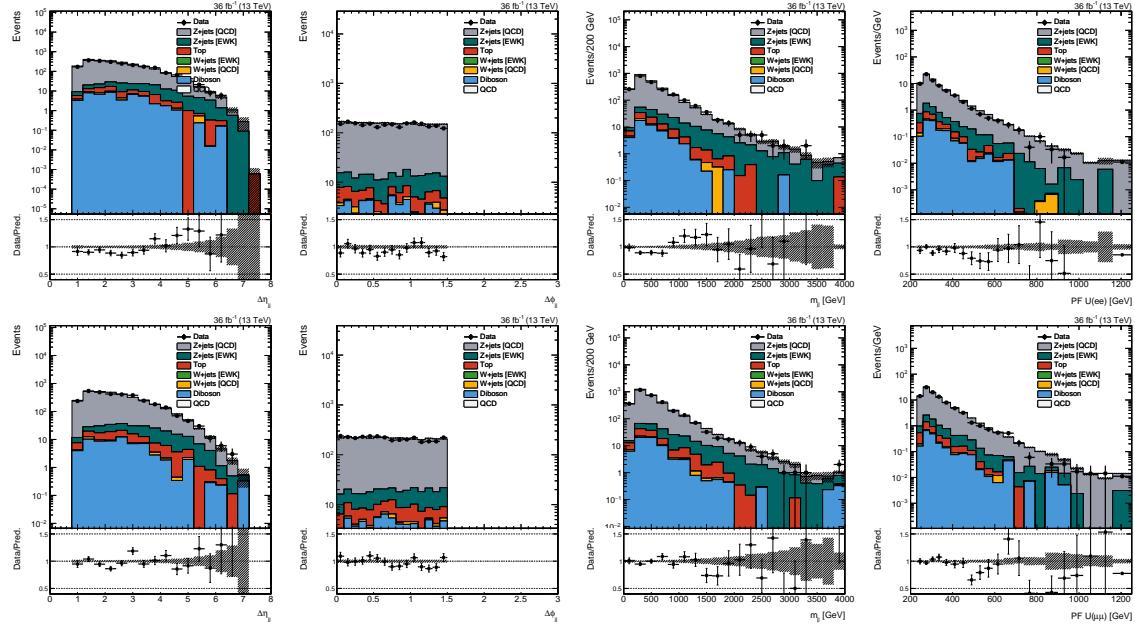


Figure 4.15: Dijet and recoil distributions in the dielectron (top) and dimuon (bottom) CRs.

is modified to include the constraints of the single-lepton CRs:

$$\begin{aligned} \mathcal{L}(\mathbf{d} | \mu, \boldsymbol{\mu}_{\text{SR}}^{Z \rightarrow \nu\nu}, \boldsymbol{\theta}) = & \\ \prod_{i \in \text{bins}} \left[ \text{Pois} \left\{ d_i^{\text{SR}} \mid \mu S_i^{\text{SR}}(\boldsymbol{\theta}) + \left( 1 + \frac{1}{T_{Z,i}^{\text{QE}}(\boldsymbol{\theta})} \right) \left( 1 + \frac{1}{T_{Z/W,i}^{\text{SR}}(\boldsymbol{\theta})} \right) \mu_{\text{SR},i}^{Z \rightarrow \nu\nu} + B_i^{\text{SR}}(\boldsymbol{\theta}) \right\} \right. & \\ \times \prod_{X=\mu,e} \text{Pois} \left\{ d_i^X \mid \left( 1 + \frac{1}{T_{Z,i}^{\text{QE}}(\boldsymbol{\theta})} \right) \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{W,i}^X(\boldsymbol{\theta}) T_{Z/W,i}^{\text{SR}}(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right\} & \\ \times \prod_{X=\mu\mu,ee} \text{Pois} \left\{ d_i^X \mid \left( 1 + \frac{1}{T_{Z,i}^{\text{QE}}(\boldsymbol{\theta})} \right) \frac{\mu_{\text{SR},i}^{Z \rightarrow \nu\nu}}{T_{Z,i}^X(\boldsymbol{\theta})} + B_i^X(\boldsymbol{\theta}) \right\} \right] & \times \prod_{j=0}^{n_\theta} p_j(\theta_j) \quad (4.7) \end{aligned}$$

To validate that the transfer factors are reasonably well-simulated (within the assigned uncertainties), Figure 4.17 uses the following ratios of CRs as proxies for transfer

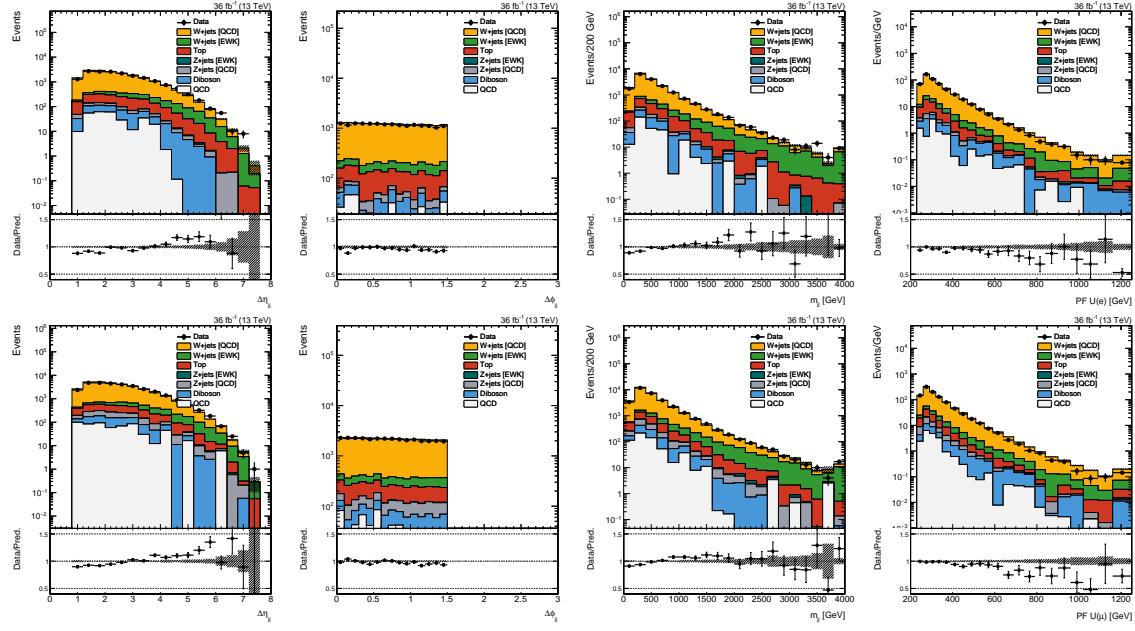


Figure 4.16: Dijet and recoil distributions in the single-electron (top) and single-muon (bottom) CRs.

factors:

$$\begin{aligned}
 \mathbf{T}_Z^{\mu\mu}, \mathbf{T}_Z^{ee} &\sim \frac{N_{\mu\mu}(Z \rightarrow \mu\mu)}{N_{ee}(Z \rightarrow ee)} \\
 \mathbf{T}_W^{\mu}, \mathbf{T}_W^e &\sim \frac{N_\mu(W \rightarrow \mu\nu)}{N_e(W \rightarrow e\nu)} \\
 \mathbf{T}_{Z/W}^{\text{SR}} &\sim \frac{N_\mu(W \rightarrow \mu\nu) + N_e(W \rightarrow e\nu)}{N_{\mu\mu}(Z \rightarrow \mu\mu) + N_{ee}(Z \rightarrow ee)}
 \end{aligned} \tag{4.8}$$

## 4.3 Results

The dijet mass distribution in data is fit in all signal and control regions, the results of which are shown in Figure 4.18. As no statistically significant excess over the Standard Model is observed, we translate the results into upper limits on the branching ratio of  $H \rightarrow \chi\bar{\chi}$ . As the signal hypothesis, both the VBF and gluon fusion (with 2 extra jets) Higgs production modes are considered; the latter contaminates the SR due to its relatively large cross section. After the signal region selection criteria, the

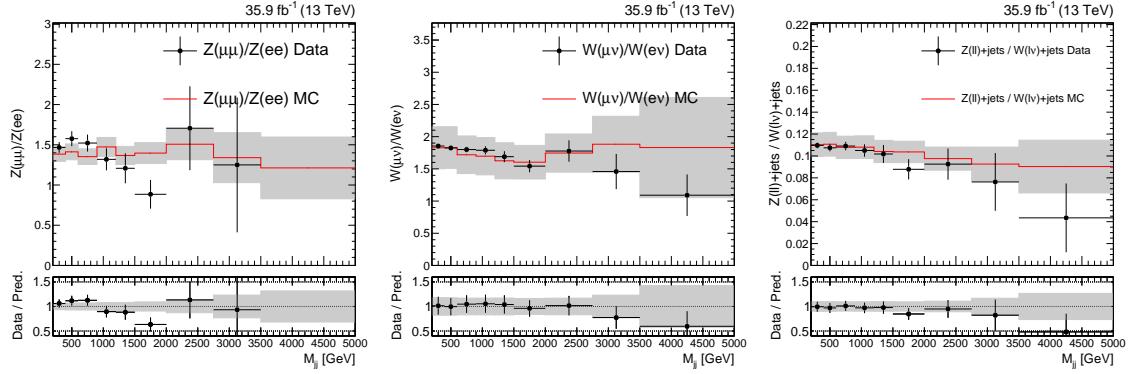


Figure 4.17: Validation of the VBF transfer factors using control region data. The transfer factor proxies are found to agree quite well with the data within the post-fit uncertainties

two modes contribute approximately equal yields. Assuming  $m_H = 125$  GeV, the observed 95% CL upper limit is 0.33. Assuming a background-only hypothesis, the expected distribution of upper limits has median 0.33, with the 1 standard deviation band covering [0.18, 0.35]; the observation therefore represents an upwards fluctuation slightly under  $1\sigma$ .

We further scan  $m_H$  and set upper limits on  $\sigma(qq \rightarrow qqH)\mathcal{B}(H \rightarrow \chi\bar{\chi})$ . The upper limits are shown in Figure 4.19, and the observed (expected) limits exclude  $m_H < 540$  GeV (635 GeV) assuming a branching ratio of 100%.

As described in the beginning of this chapter, each Higgs production mode corresponds to a potential invisible Higgs search channel. While VBF is the most sensitive, the total sensitivity can be improved by statistically combining all channels. Other results from CMS cover searches for associated production of a Higgs boson, either with a leptonically-decaying  $Z$  boson ([?]) or a hadronically-decaying weak boson ([?]); and for gluon fusion production, with at least one jet originating from the initial state or heavy quark loop ([?]). The details of these searches are left to the referenced literature, but a summary of their results is provided in Figure 4.20. When statistically combining the results, most experimental nuisances are treated as correlated between the searches, with the exception of the VBF jet energy scale dependence. This is because the VBF category selects forward jets, whereas other searches generally probe

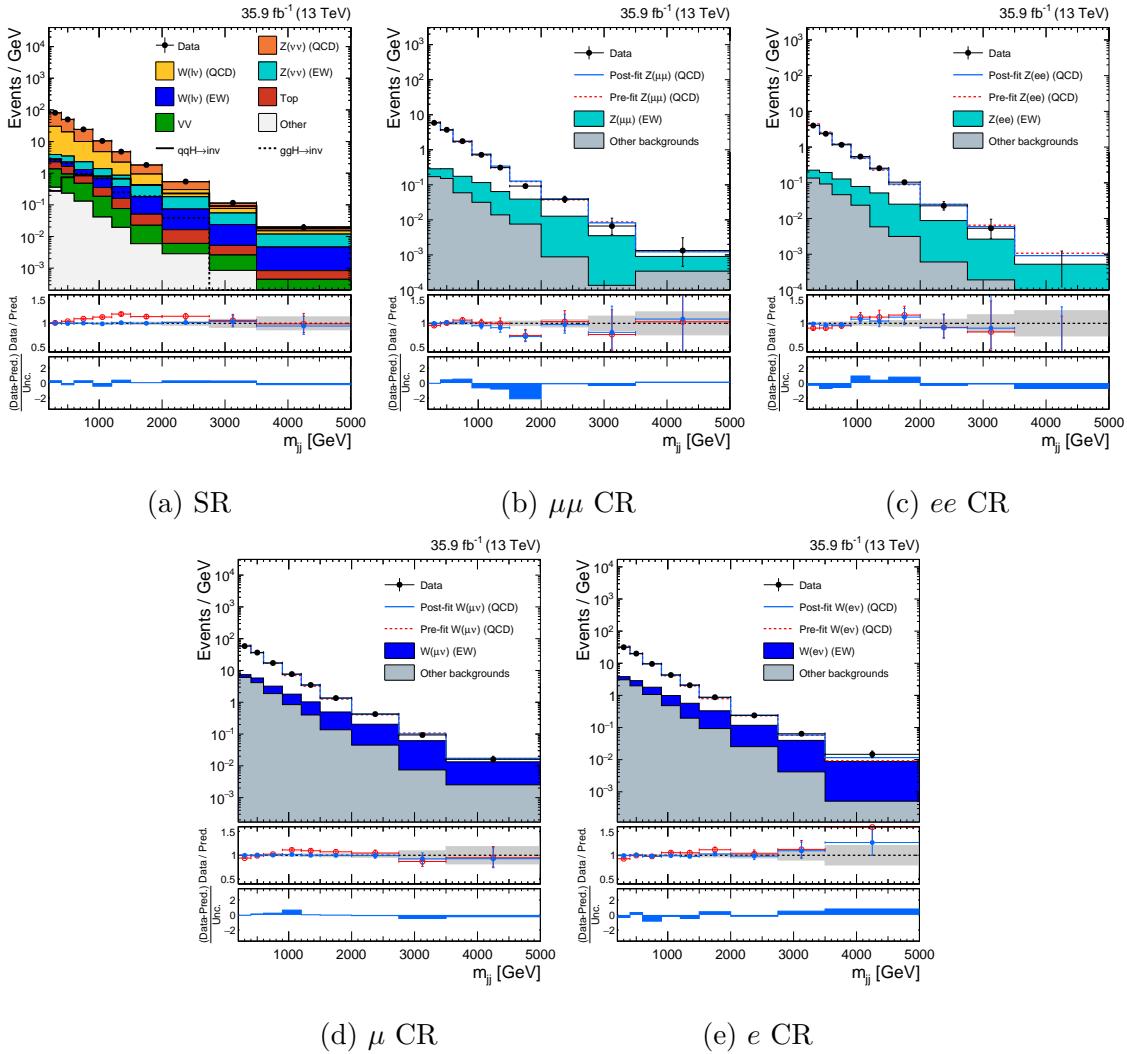


Figure 4.18: Post-fit  $m_{jj}$  distributions in the various signal and control regions. The uncertainties (gray bands) and bin pulls (blue bands) are defined by varying the nuisances by one standard deviation around the maximum likelihood estimate.

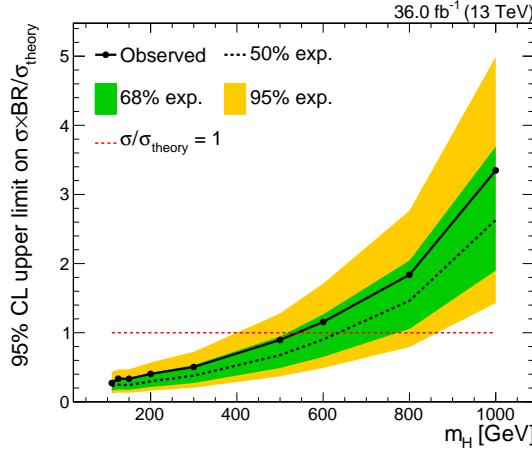


Figure 4.19: Upper limits on  $\sigma \times \mathcal{B}/\sigma_{\text{theory}}$  as a function of  $m_H$ , where  $\sigma$  refers to the total production cross section of the Higgs boson with mass  $m_H$ . If one assumes that  $\sigma = \sigma_{\text{theory}}$ , then the upper limits can be interpreted directly as constraints on  $\mathcal{B}(H \rightarrow \chi\bar{\chi})$ .

central jets. Theoretical nuisances (e.g. those affecting  $W/Z$  or  $ZZ/WZ$  ratios) are left uncorrelated between all searches. The combined result constrains  $\mathcal{B}(H \rightarrow \chi\bar{\chi})$  to be less than 0.26 at 95% CL, which approximately corresponds to a 1 standard deviation fluctuation upward relative to the median expected limit of 0.20.

Higgs-mediated DM can also be probed by direct detection (DD) experiments. We interpret the combined 90% CL upper limit (for consistency with DD conventions) on  $\mathcal{B}(H \rightarrow \chi\bar{\chi})$  as an upper limit on the spin-independent cross section of DM-nucleon scattering. First, we convert the branching ratio into a partial width:

$$\Gamma_{H \rightarrow \chi\bar{\chi}} = \frac{\mathcal{B}(H \rightarrow \chi\bar{\chi}) \cdot \Gamma_{\text{SM}}}{1 - \mathcal{B}(H \rightarrow \chi\bar{\chi})}, \text{ where } \Gamma_{\text{SM}} = 4 \text{ GeV} \quad (4.9)$$

Then, using the results described in Reference [?],  $\Gamma_{H \rightarrow \chi\bar{\chi}}$  is translated into  $\sigma_{\chi N}^{\text{SI}}(m_\chi)$ , where the only free parameter is the DM mass. Figure 4.21 compares the CMS exclusions with those from DD experiments [?]. Also shown for comparison is the “neutrino floor”, which is the cross section of coherent scattering of solar neutrinos, and a limiting factor for DD experiments. At low  $m_\chi$ , CMS is able to significantly extend the DD constraints, reaching well below the neutrino floor.

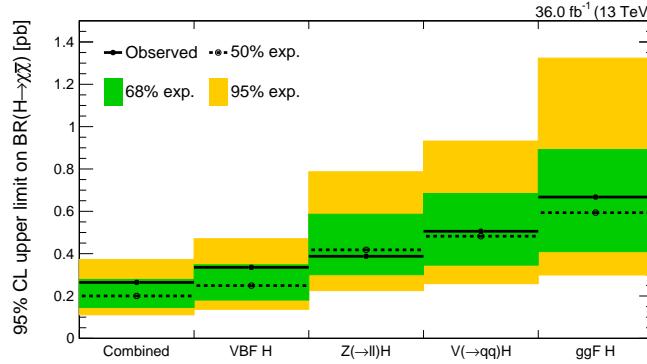


Figure 4.20: Upper limits on  $\mathcal{B}(H \rightarrow \chi\bar{\chi})$  after statistically combining all of the CMS searches for  $H \rightarrow \chi\bar{\chi}$  conducted on  $36 \text{ fb}^{-1}$  of data collected in 2016. For comparison, the upper limits of each of the individual categories are also shown.

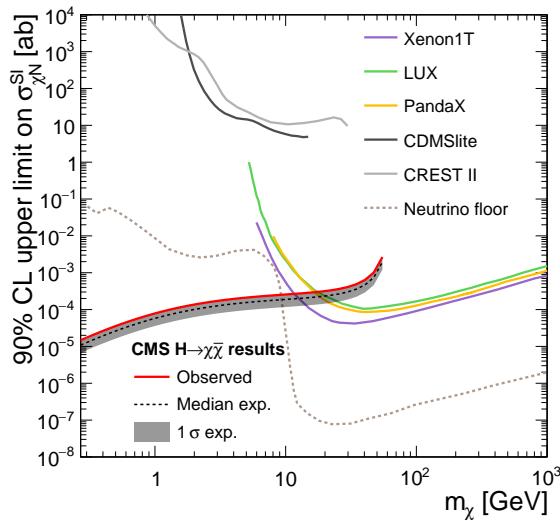


Figure 4.21: Upper limits on  $\sigma_{\chi N}^{\text{SI}}$  as a function of  $m_\chi$ . Shown are the combined results from the CMS invisible Higgs searches, as well as various direct detection experiments.

# Bibliography

- [1] L. Evans and P. Bryant, “LHC Machine,” *JINST*, vol. 3, p. S08001, 2008.
- [2] C. De Melis, “The CERN accelerator complex. Complexe des accélérateurs du CERN,” Jan 2016. General Photo.
- [3] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, “LHC Design Report. 3. The LHC injector chain,” 2004.
- [4] C. Collaboration. [https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGlobalHLT#Trigger\\_development\\_for\\_Run\\_2](https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGlobalHLT#Trigger_development_for_Run_2).
- [5] S. Chatrchyan *et al.*, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08004, 2008.
- [6] S. Chatrchyan *et al.*, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, no. 10, p. P10009, 2014.
- [7] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the kalman filtering method,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 294, no. 1, pp. 219 – 228, 1990.
- [8] R. Fruehwirth, “Application of Kalman filtering to track and vertex fitting,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444 – 450, 1987.
- [9] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,” *Proceedings of the IEEE*, vol. 86, pp. 2210–2239, Nov 1998.
- [10] R. Fruehwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting,” Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [11] A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST*, vol. 13, no. 05, p. P05011, 2018.
- [12] A. Benaglia *et al.*, “The CMS ECAL performance with examples,” *Journal of Instrumentation*, vol. 9, no. 02, p. C02008, 2014.

- [13] V. Khachatryan *et al.*, “Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV,” *JINST*, vol. 10, no. 08, p. P08010, 2015.
- [14] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections,” *Journal of Physics: Conference Series*, vol. 587, no. 1, p. 012001, 2015.
- [15] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS, Geneva: CERN, 1997.
- [16] A. M. Sirunyan *et al.*, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, no. 10, p. P10003, 2017.
- [17] C. Collaboration, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data,” *Journal of Instrumentation*, vol. 5, p. T03012, Mar. 2010.
- [18] A. M. Sirunyan *et al.*, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *JINST*, vol. 13, no. 06, p. P06015, 2018.
- [19] V. Khachatryan *et al.*, “The CMS trigger system,” *JINST*, vol. 12, no. 01, p. P01020, 2017.
- [20] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics,” *Rept. Prog. Phys.*, vol. 70, p. 89, 2007.
- [21] G. L. Bayatian *et al.*, “CMS Physics Technical Design Report, Volume I: Detector Performance and Software,” 2006.
- [22] W. Adam, R. FrÃijhwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc,” *Journal of Physics G: Nuclear and Particle Physics*, vol. 31, no. 9, p. N9, 2005.
- [23] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.
- [24] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions,” *Phys. Rev.*, vol. D48, pp. 3160–3166, 1993.
- [25] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms,” *JHEP*, vol. 08, p. 001, 1997.
- [26] “FastJet User Manual,” *Eur. Phys. J.*, vol. C72, p. 1896, 2012.
- [27] V. Khachatryan *et al.*, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST*, vol. 12, no. 02, p. P02014, 2017.

- [28] V. Khachatryan *et al.*, “Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  at CMS,” *JINST*, vol. 11, no. 01, p. P01019, 2016.
- [29] “Performance of missing transverse momentum in pp collisions at  $\sqrt{s}=13$  TeV using the CMS detector,” Tech. Rep. CMS-PAS-JME-17-001, CERN, Geneva, 2018.
- [30] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP*, vol. 10, p. 059, 2014.
- [31] “CMS Offline Software.” <https://github.com/cms-sw/cmssw>.
- [32] S. Agostinelli *et al.*, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.
- [33] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.*, vol. 53, p. 270, 2006.
- [34] S. Kallweit, J. M. Lindert, P. Maierhöfer, S. Pozzorini, and M. Schönher, “NLO electroweak automation and precise predictions for W+multijet production at the LHC,” *JHEP*, vol. 04, p. 012, 2015.
- [35] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schönher, and P. Maierhöfer, “NLO QCD+EW automation and precise predictions for V+multijet production,” in *50th Rencontres de Moriond on QCD and High Energy Interactions La Thuile, Italy, March 21-28, 2015*, 2015.
- [36] J. H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze, “Electroweak corrections to hadronic photon production at large transverse momenta,” *JHEP*, vol. 03, p. 059, 2006.
- [37] 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*, vol. 04, p. 040, 2015.
- [38] A. Denner, S. Dittmaier, T. Kasprzik, and A. Mück, “Electroweak corrections to W + jet hadroproduction including leptonic W-boson decays,” *JHEP*, vol. 08, p. 075, 2009.
- [39] A. Denner, S. Dittmaier, T. Kasprzik, and A. Mück, “Electroweak corrections to dilepton + jet production at hadron colliders,” *JHEP*, vol. 06, p. 069, 2011.
- [40] A. Denner, S. Dittmaier, T. Kasprzik, and A. Mück, “Electroweak corrections to monojet production at the LHC,” *Eur. Phys. J. C*, vol. 73, p. 2297, 2013.
- [41] J. H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze, “Logarithmic electroweak corrections to hadronic Z+1 jet production at large transverse momentum,” *Phys. Lett. B*, vol. 609, p. 277, 2005.

- [42] J. H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze, “One-loop weak corrections to hadronic production of Z bosons at large transverse momenta,” *Nucl. Phys. B*, vol. 727, p. 368, 2005.
- [43] J. H. Kuhn, A. Kulesza, S. Pozzorini, and M. Schulze, “Electroweak corrections to hadronic production of W bosons at large transverse momenta,” *Nucl. Phys. B*, vol. 797, p. 27, 2008.
- [44] D. A. *et al*, “Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum,” 2015.
- [45] D. de Florian *et al*, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,” Tech. Rep. FERMILAB-FN-1025-T, Oct 2016. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.