# Search for supersymmetry using boosted Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV

by

**Frank Jensen**

M.S., University of Colorado Boulder, 2014

B.A., University of California Berkeley, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

2018

This thesis entitled:
Search for supersymmetry using boosted Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV
written by Frank Jensen
has been approved for the Department of Physics

Professor Kevin Stenson

Professor William Ford

Professor Jamie Nagle

Professor John Cumalat

Professor Senarath de Alwis

Date ____________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Jensen, Frank (Ph.D.,experimental high-energy physics)

Search for supersymmetry using boosted Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV

Thesis directed by Professor Kevin Stenson

A search for physics beyond the Standard Model in events with one or more high-momentum Higgs bosons, H, decaying to pairs of b quarks in association with missing transverse momentum is presented. The data, corresponding to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$, were collected with the CMS detector at the LHC in proton-proton collisions at the center-of-mass energy $\sqrt{s} = 13\,\mathrm{TeV}$. The analysis utilizes a new b quark tagging technique based on jet substructure to identify jets from H $\rightarrow$ $b\bar{b}$. Events are categorized by the multiplicity of H-tagged jets, jet mass, and the missing transverse momentum. No significant deviation from standard model expectations is observed. In the context of supersymmetry (SUSY), limits on the cross sections of pair-produced gluinos are set, assuming that gluinos decay to quark pairs, H (or Z), and the lightest SUSY particle, LSP, through an intermediate next-to-lightest SUSY particle, NLSP. With large mass splitting between the NLSP and LSP, and 100% NLSP branching fraction to H, the lower limit on the gluino mass is found to be 2010 GeV.

## Dedications

*The taxpayers.*

# Contents

**Tables**

**Table**

## Figures

**Figure**

# Chapter 1

# Introduction

A description of the Standard Model is presented in Chapter 2. A description of the CMS detector is presented in Chapter 3. A description of the physics analysis is presented in Chapter 4. The conclusions are discussed in Chapter 5.

# Chapter 2

# The Standard Model of Particle Physics

The discovery of the Higgs (H) boson in 2012 was a monumental achievement solidifying the SM as it found the final fundamental particle within the theory.[1] A diagram of the current knowledge of fundamental particles of the SM is seen in Figure 2.1.

### 2.0.1 Electroweak Theory

### 2.0.2 Quantum Chromodynamics

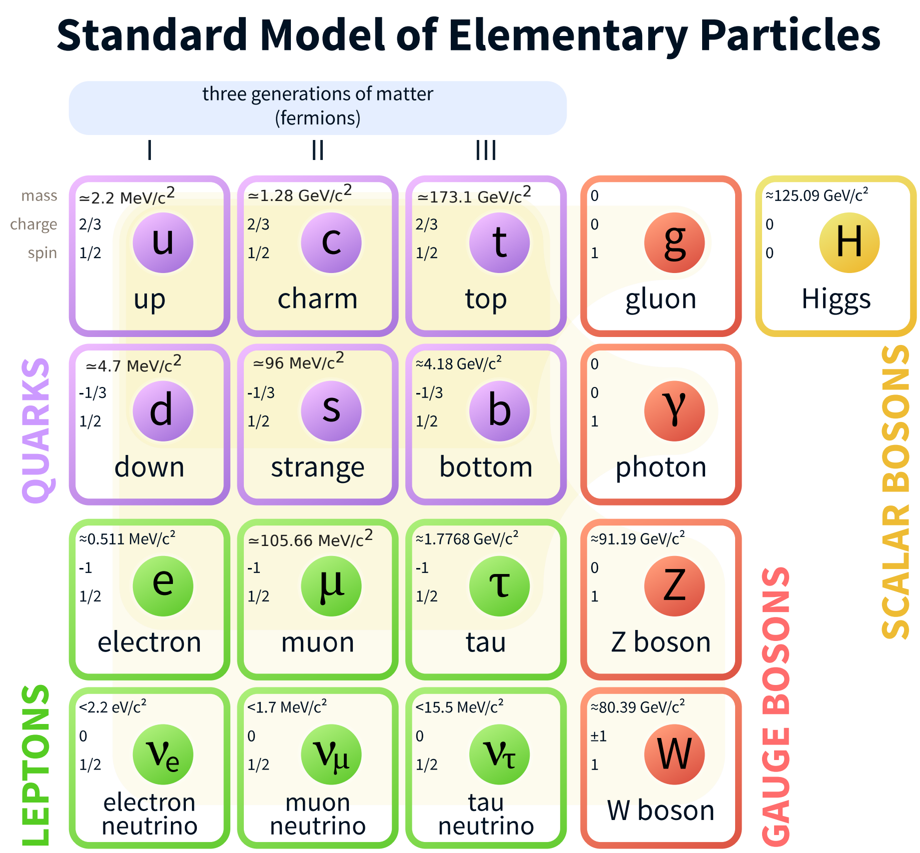



Figure 2.1: The particles of the Standard Model.

# Chapter 3

# The CMS Detector

The detector is cylindrically symmetric, with a modular design consisting of different subsystems specialized for detection or identification of specific particles. Many of the Standard Model particles are unstable. In fact only about 8 particles are directly detected in our experiment, the rest of those being reconstructed via their decay into these more stable elements. Of all the Standard Model particles,

Figure 3.1 is a depiction of the CMS detector. Figure 3.2 is a schematic view of the CMS detector and the particles it interacts with.

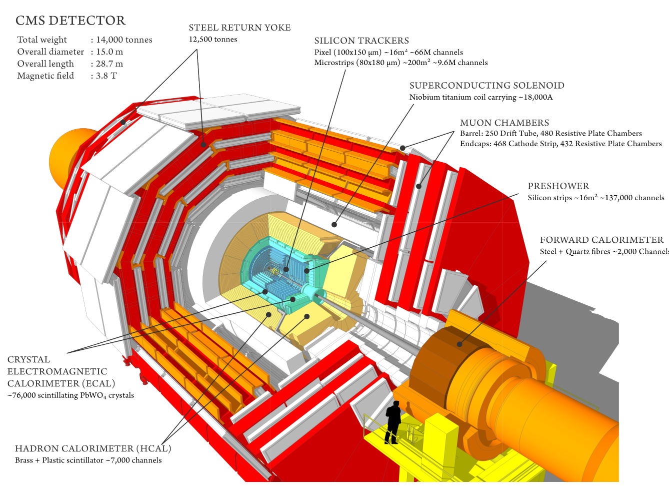



Figure 3.1: A depiction of the CMS detector.

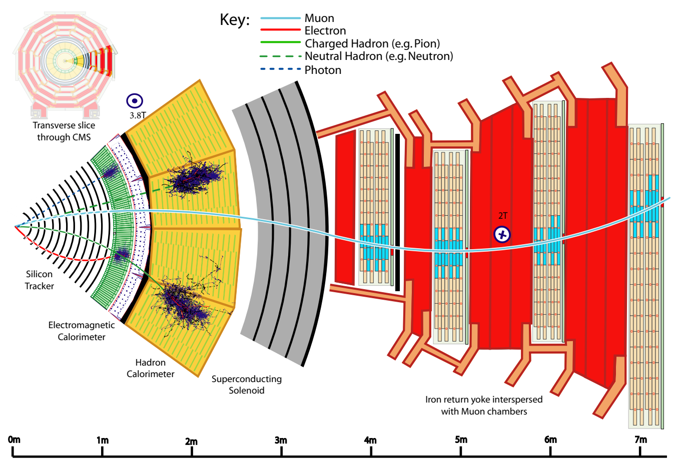



Figure 3.2: A schematic view of the CMS detector and the particles it interacts with.

## 3.1 Silicon Tracker

At the center of the detector lies the silicon tracker which is responsible for momentum reconstruction of charged particles and reconstruction of the primary vertices created from a proton-proton interaction. The tracker is composed of a number of concentric cylindrical layers each instrumented with silicon sensors. As a charged particle traverses an individual sensor a small amount of ionization energy is deposited in that element. A bias voltage is applied across the sensor to direct the charge to either end of the sensor. Bonded to the sensors are front-end electronics responsible for the amplification and shaping of the resultant signal. Neighboring elements are clustered together creating what is known as a "hit". These hits are then used to reconstruct the particle's trajectory through the tracker. To make a measurement of the momentum and charge of the particle, the tracker is immersed in a 3.8 T magnetic field which bends the trajectory of the particle.

The tracker is further divided into an inner and outer parts which make use of slightly different technologies. The inner part of the tracker is known as the pixel detector and consists of 4 layers which the sensitive elements are silicon pixels of size 100x150 $\mu$m.

## 3.2 Electromagnetic Calorimeter

## 3.3 Hadronic Calorimeter

## 3.4 Solenoidal Magnet

## 3.5 Muon System

## 3.6 Particle Flow Event Reconstruction

Particle Flow event reconstruction is pretty good.[2]

## 3.7 Trigger System

The trigger system is pretty good.[3]

# Chapter 4

# Search for supersymmetry using boosted Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV

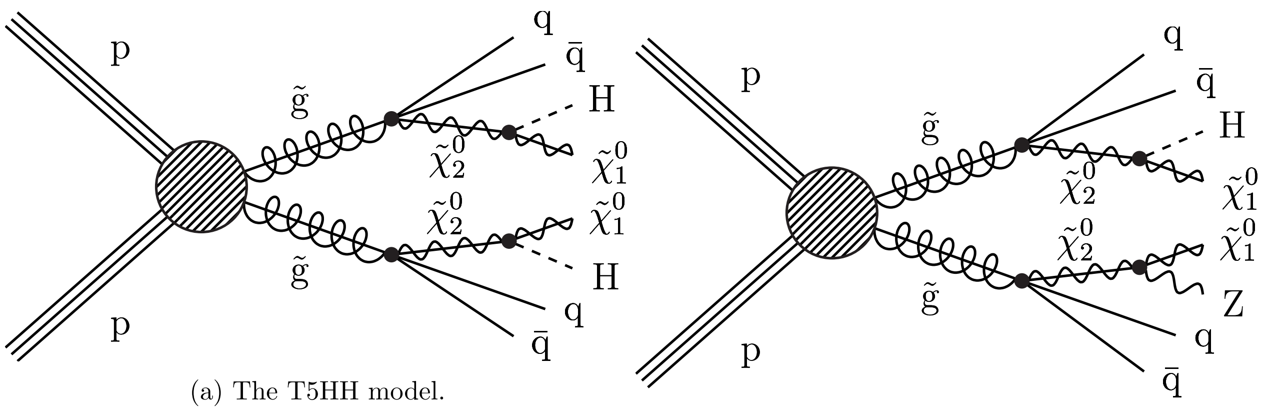
## 4.1 Motivation & Introduction

As the masses of potentially new particles are pushed to larger and larger scales [4] [5]. This results in any Standard Model (SM) particles radiated during their decay having correspondingly larger momentum (high boost). As a particle becomes more boosted its decay products are emitted at smaller angles, eventually collimating sufficiently to be reconstructed as a single jet. If new physics exists with masses achievable by the LHC, one could suspect that there exists non-zero coupling with the electroweak H, Z, or W bosons. Observation of events containing high-$p_T$ (>300 GeV) electroweak bosons are thus of considerable interest for potential SUSY signals.

The Minimal Supersymmetric Standard Model contains a $\mathbb{Z}_2$ symmetry in which all SM particles have charge -1 and all supersymmetric particles have charge +1, this is called R-parity[6]. One direct consequence of R-parity is that the decay of a massive supersymmetric particle must include at least one supersymmetric particle in the final state. Necessarily this is the lightest such particle in the theory, denoted the lightest supersymmetric particle (LSP). If the LSP is electrically neutral it may escape detection, creating an imbalance in the net momentum of the event (as would a neutrino). Events with a large momentum imbalance are also interesting as potential regions for SUSY signals.

With this as motivation, we designed an analysis searching for hints of new physics beyond the Standard Model in events with boosted H or Z bosons and a large transverse momentum imbalance of the event. We reconstruct the H and Z bosons in the $b\bar{b}$ channel, with 57% and 15% branching fractions respectively. Although our analysis is sensitive to any new physics with this final state, we have adopted two benchmark models (known as SMS models[7]), seen in Figure 4.1, to give motivation to the analysis. In this scenario, the proton-proton interaction produces a pair of gluinos $\tilde{g}$ which decay to a neutralino $\tilde{\chi}_2^0$ by the emission of Standard Model quarks. A small mass splitting between the gluino and neutralino $\tilde{\chi}_2^0$ will result in low-$p_T$ quarks and a high-$p_T$ neutralino $\tilde{\chi}_2^0$. This neutralino $\tilde{\chi}_2^0$ further decays into another neutralino $\tilde{\chi}_1^0$ with the emission of a H or Z boson. The neutralino $\tilde{\chi}_1^0$ escapes detection.

Past searches targeting similar final states (but different production scenario) have been performed in which the H bosons are produced with low-$p_T$, in this case the H bosons are reconstructed as a pair of b-tagged AK4 jets [8].




(a) The T5HH model.

(b) The T5ZH model.

Figure 4.1: Diagrams of the benchmark models used for motivation of the targeted signal.

## 4.2 Object Definition & Event Selection

We establish a baseline selection choosing events with all-hadronic final states and missing transverse momentum, as motivated by Figure 4.1 . Our jets are reconstructed using the "anti-kt" clustering algorithm with cone sizes of $\Delta R = 0.4$, 0.8 [9], denoted as AK4 and AK8 jets respectively. AK4 jets subtend less solid angle and are used to capture the hadronisation of single quarks and gluons. AK8 jets subtend a larger solid angle and are used for reconstruction of boosted objects (e.g. $t$, H, Z, W).

The baseline selection is as follows:

- $\geq 2$ AK8 jets; $p_T > 300\,\text{GeV}$ and $50 < \text{mass} < 250\,\text{GeV}$
- $\text{MET} > 300\,\text{GeV}$; $\text{MET} \equiv |-\sum_{\text{PFcandidates}} \vec{p_{T,i}}|$
- electron veto; $p_T > 10\,\text{GeV}$, isolated
- muon veto; $p_T > 10\,\text{GeV}$, isolated
- isolated track veto
- $\Delta\phi_{1,2,3,4} > 0.5, 0.5, 0.3, 0.3$; $\Delta\phi_i \equiv \Delta\phi(\vec{\text{MET}}, \text{AK4}\,\text{jet}_i)$

  The $\Delta\phi$ cut is designed to mitigate QCD events in which a jet is under-measured leading to an artificial imbalance in the event momentum. This cut requires that the difference in $\phi$ between the MET vector and each of the four leading AK4 jets is sufficiently large to remove events in which a jet has been under-measured giving rise to fake MET pointing in the same direction. If less than four AK4 jets are available the additional cuts are removed.

A dedicated heavy tagging algorithm is used to identify AK8 jets arising from the decay of two b quarks. The distribution of the output discriminator for the lead and subleading jets are seen in the left and right panels of Figure 4.2, respectively; signal-like events peak towards larger values. To $\mathrm{b\bar{b}}$ tag the AK8 jets we choose the loose working-point ($> 0.3$)corresponding to a signal efficiency of approximately 80% per AK8 jet. The stacked histogram and solid lines shows the distribution after baseline selection for simulation and two representative signal points, respectively.

Further H/Z tagging of an AK8 jet is accomplished by restricting the jet mass window to [85, 135 GeV] to be consistent with that of the H boson. The distributions of the jet mass are seen in Figure 4.3. The signal shown in the solid line is the T5HH model (i.e. Figure 4.1a). The same identification criteria are applied to tag an AK8 jet as either an H or Z boson, there is no distinction made.

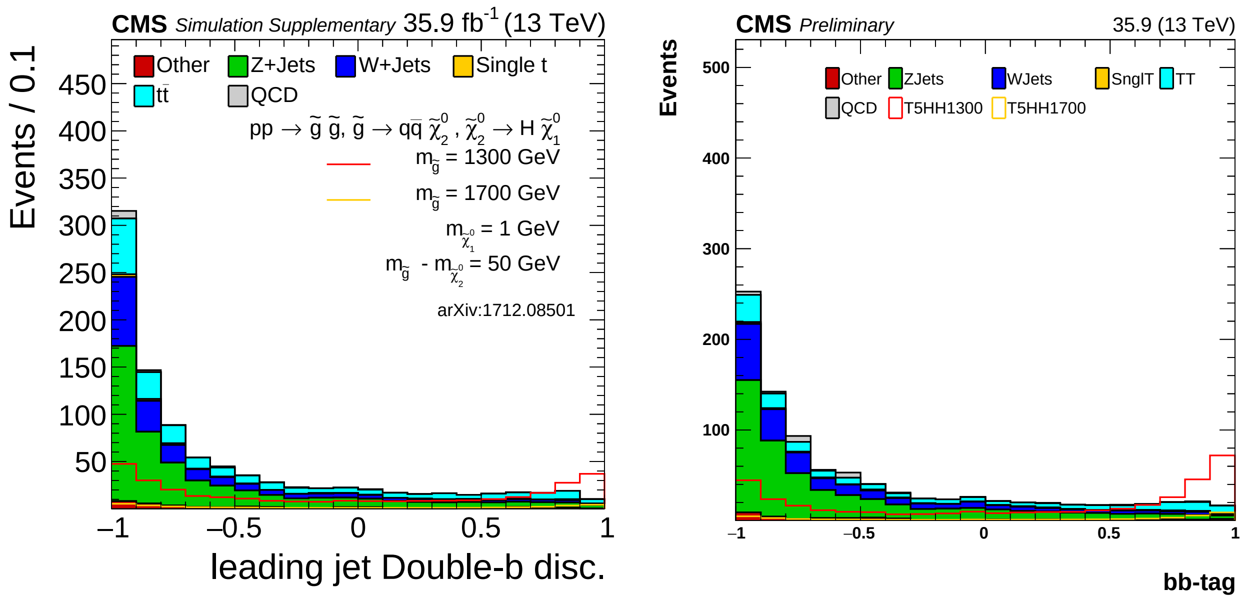



Figure 4.2: $b\bar{b}$ tagging neural network discriminator.

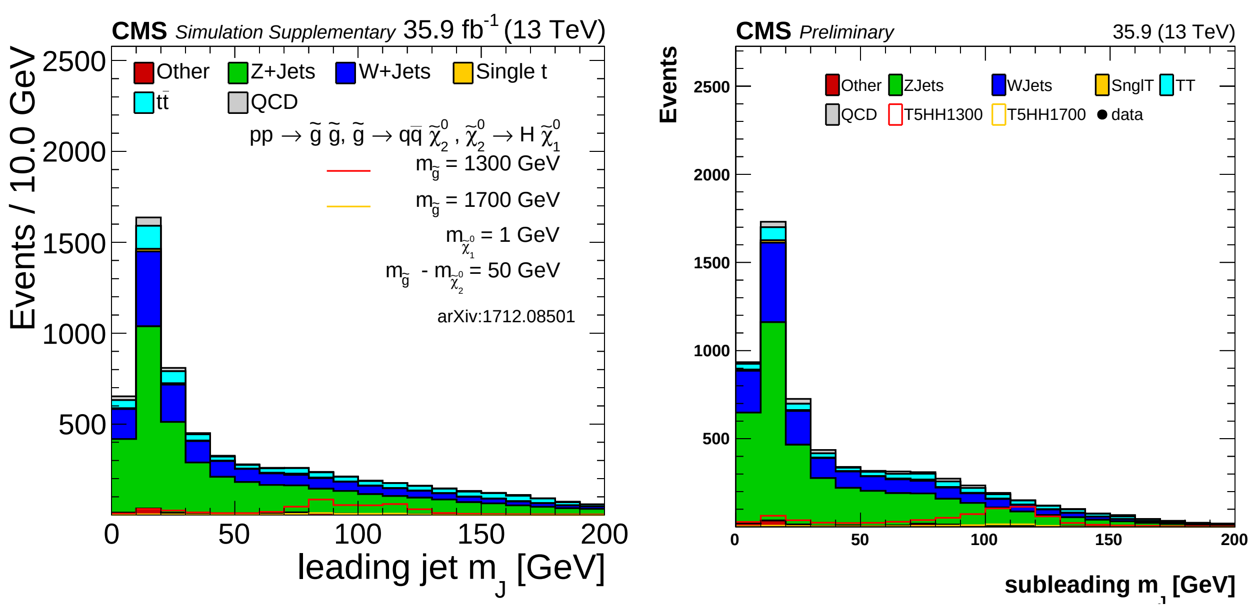



Figure 4.3: The AK8 jet mass.

## 4.3 Dataset and Trigger

We use a total of 35.9 $fb^1$ of data collected by the CMS experiment in 2016. Events are selected in data using a trigger which requires greater than 100 GeV of MET calculated at high-level trigger (HLT); additionally the logical OR of two other triggers with thresholds of 110 and 120 GeV are applied. The trigger efficiency is derived in data using a single-electron reference trigger requiring a tight-ID electron of $p_T > 27$ GeV. We further select events with at least three AK4 jets and exactly one reconstructed electron of $p_T > 25$ GeV. The signal region trigger is found to be greater than 98% for events with MET>250 GeV and HT>300 GeV [4].

## 4.4 Event samples

### 4.4.1 Standard model MC samples

Monte Carlo (MC) samples reconstructed with CMSSW release 8_0_X (Summer16) are used for all processes. The SM samples are listed in Tables 4.1, 4.2, 4.3, and 4.4. The cross sections listed correspond to next-to-next-to-leading-order (NNLO) calculations unless otherwise noted. All samples use the PU25bx25 pileup scenario, which simulates a pileup distribution with an average of 25 interactions per bunch crossing and a 25 ns interval between bunches. The Monte Carlo samples are used to validate the data-driven background estimation methods.

Table 4.1: SM $t\bar{t}$ MC samples used in the analysis. The cross sections are calculated to NNLO.

| Dataset | $\sigma$ (pb) | $\int \mathbb{L}$ ($fb^{-1}$) |
|---|---|---|
| TTJets_SingleLeptFromT_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 182.72 | 283.90 |
| TTJets_SingleLeptFromTbar_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 182.72 | 326.48 |
| TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 88.34 | 346.25 |
| TTJets_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 2.734 | 5231.81 |
| TTJets_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1.121 | 9416.61 |
| TTJets_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0.198 | 14819.34 |
| TTJets_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0.002 | 221088.29 |

Table 4.2: SM QCD MC samples used in the analysis. All cross sections are calculated to LO.

| Dataset | $\sigma$ (pb) | $\int \mathbb{L}$ (fb$^{-1}$) |
|---|---|---|
| QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1735000 | 0.03 |
| QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 366800 | 0.16 |
| QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 29370 | 1.95 |
| QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 6524 | 6.68 |
| QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1064 | 12.62 |
| QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 121.5 | 32.63 |
| QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 25.42 | 239.30 |

Table 4.3: SM $Z \to \nu\nu$+jets MC samples used in the analysis. The cross sections are calculated to NNLO.

| Dataset | $\sigma$ (pb) | $\int \mathbb{L}$ (fb$^{-1}$) |
|---|---|---|
| ZJetsToNuNu_HT-100To200_13TeV-madgraph | 344.8 | 54.13 |
| ZJetsToNuNu_HT-200To400_13TeV-madgraph | 95.53 | 208.46 |
| ZJetsToNuNu_HT-400To600_13TeV-madgraph | 13.20 | 77.30 |
| ZJetsToNuNu_HT-600To800_13TeV-madgraph | 3.148 | 1795.26 |
| ZJetsToNuNu_HT-800To1200_13TeV-madgraph | 1.451 | 1486.09 |
| ZJetsToNuNu_HT-1200To2500_13TeV-madgraph | 0.355 | 1029.81 |
| ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph | 0.0085 | 47498.87 |

Table 4.4: SM $W \to \ell\nu$+jets MC samples used in the analysis. The cross sections are calculated to NNLO.

| Dataset | $\sigma$ (pb) | $\int \mathbb{L}$ (fb$^{-1}$) |
|---|---|---|
| WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1627.45 | 18.16 |
| WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 435.24 | 45.88 |
| WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 59.18 | 123.64 |
| WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 14.58 | 221.32 |
| WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 6.66 | 1123.13 |
| WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1.608 | 153.44 |
| WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0.039 | 6497.28 |

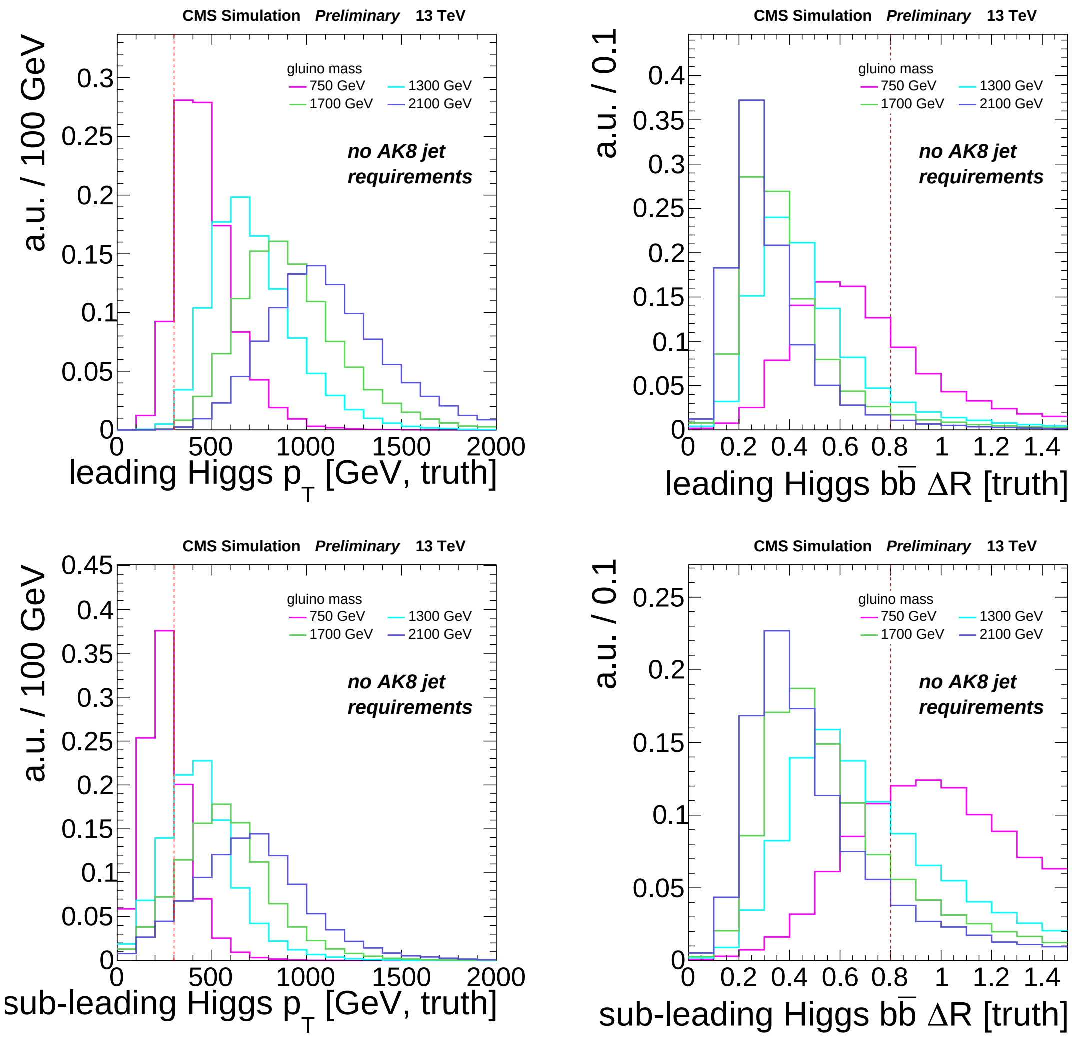



Figure 4.4: Generator level distributions for the leading (top row) and subleading (bottom row) H boson in the T5HH model. The plot on the left shows the $p_T$ of the H boson. The plot on right shows $\Delta R$ between the b-quark daughters - for large Higgs $p_T$ the daughters become collimated.

### 4.4.2 Signal models

Figure 4.1 shows the signal diagrams for the interpretations considered in this analysis. The mass splitting between the gluino $\tilde{g}$ and NLSP is fixed to 50 GeV, resulting in low $p_T$ quarks produced in gluino $\tilde{g}$ decays, and the mass of the LSP is fixed to 1 GeV so that the H $p_T$ is proportional to $m_{NLSP}/2 \sim m_{\tilde{g}}/2$. The signal regions for this analysis are designed to be sensitive to events with 2 H bosons in the final state with $H \to b\bar{b}$ or a mixed state of H and Z in the final state with an arbitrary relative branching fraction. Figure 4.4 shows the H $p_T$ for various $m_{\tilde{g}}$ model points. For most gluino masses, the b-quarks from $H \to b\bar{b}$ and $Z \to b\bar{b}$ are expected to be contained in a large-radius jet of $\Delta R = 0.8$. Events are generated with the Full Simulation using the reconstruction in CMSSW version `8_0_X`.

## 4.5 Event Binning & Background Estimation

The background estimation procedure makes use of what is known as an "ABCD" prediction in which the analysis phase space is divided into signal and sideband regions; scaling relations are applied to sideband yields to make predictions for the SM background (inclusive in all processes) in the signal regions. The events are categorized according to whether the AK8 jets are a) in the signal or sideband mass region and b) have or have not been $\mathrm{b\bar{b}}$ tagged. A diagram of this partitioning is seen in Figure 4.5. An additional dimension is added by binning in MET: [300, 500 GeV], [500, 700 GeV], 700+ GeV. This gives a total of 2x3=6 signal and 4x3=12 sideband bins. The two signal regions $\mathrm{A_1}$ and $\mathrm{A_2}$ contain events with one (and only one) or two jets being consistent with H/Z boson decay, respectively.

Assuming that there is no correlation between the jet mass and the $\mathrm{b\bar{b}}$ tagging, the ratio of events $\mathrm{A_{1,2}/B_{1,2}}$ should be the same as C/D. Rearranging this gives a prediction for the events in the signal regions, $(\mathrm{A_{1,2}})_{\text{prediction}} = (\mathrm{B_{1,2}} * \mathrm{C/D})_{\text{observed}}$.

The expected MET distribution from simulation is seen in the stacked histograms of Figure 4.6. The prediction using the ABCD method is seen in the red hash. The expected level of closure in

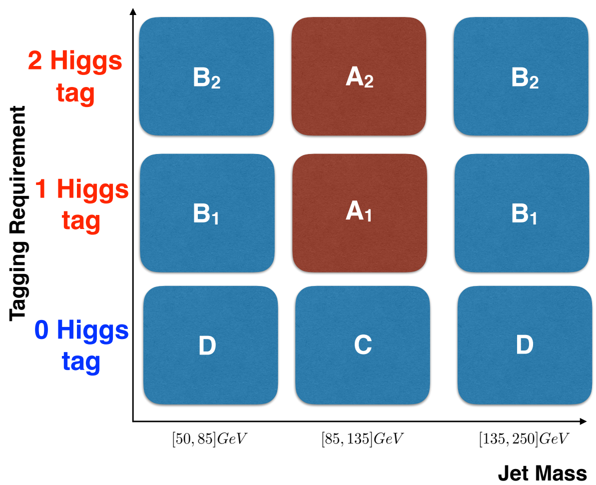
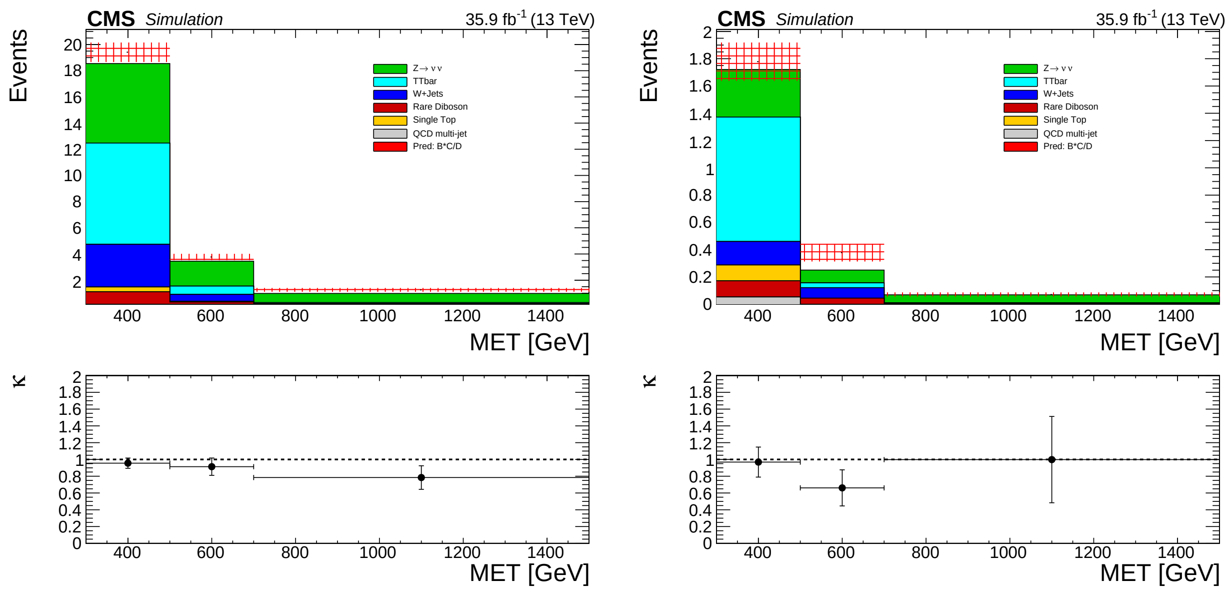



Figure 4.5: A diagram of the partitioned phase space. An additional binning in MET brings the number of analysis bins to 6x3=18.

simulation can be determined by dividing the prediction in the signal region with the true content. This ratio, denoted $\kappa$, is seen in the bottom panel of Figure 4.6. As will be discussed in Section 4.5.3, $\kappa$ is used as a correction in the background estimation procedure.




(a) The single Higgs tag region ($A_1$).

(b) The double Higgs tag region ($A_2$).

Figure 4.6: MET distributions and predictions in the signal regions using simulation only.

### 4.5.1 Study of the SM Background

We treat the SM backgrounds in the signal region as consisting of three main components:

- $Z \to \nu\bar{\nu}$ in which the invisible Z gives true MET ('Z-invisible').
- Semi-leptonic W or t production in which the lepton is not identified ('lost-lepton').
- fake MET arising from jet mis-measurement ('QCD'). Typically, the jet $p_T$ is underestimated (the thing dont work!).

To understand these backgrounds in data, three control regions were defined to serve as a proxy for each source.

- A single-photon control region, after artificially removing the photon from event reconstruction, closely mimics the Z-invisible background. For high-$p_T$, both photons and Z bosons become massless, neutral, gauge bosons whose kinematics are expected to be similar.

- The single-lepton control region mimics the lost-lepton background.

- The low-$\Delta\phi$ control region most closely mimics the QCD background.

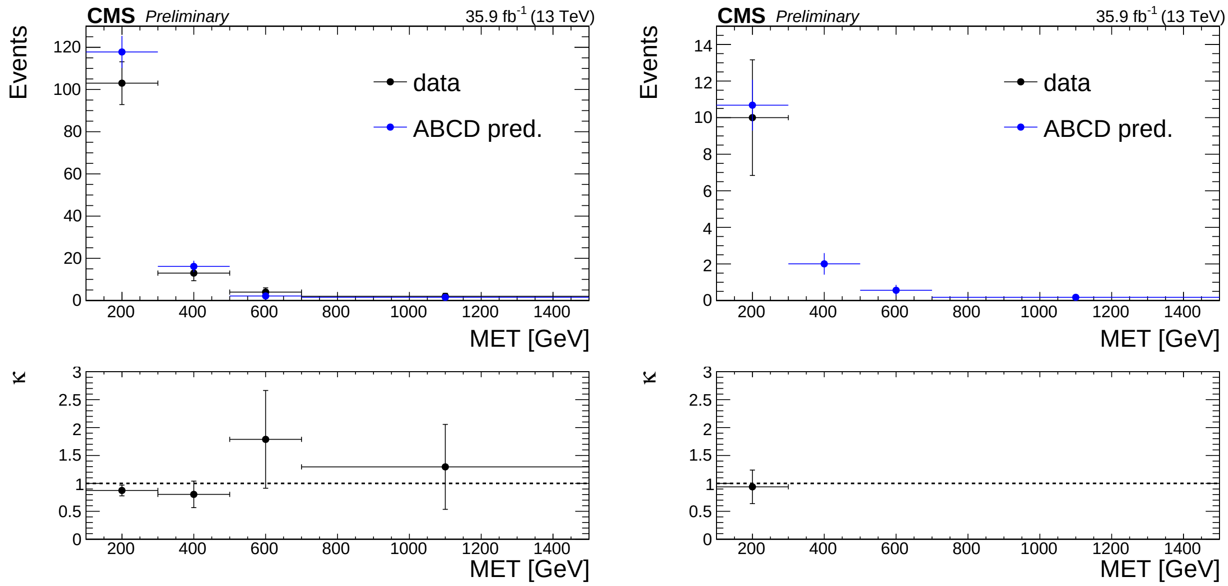
### 4.5.2 Closure of the Estimation Within the Data Control Regions

As they are orthogonal to our signal regions, we are able to test the closure of the background estimation technique independently in each of the three control regions. By comparing the prediction of the SM yields (using the ABCD method) with those observed in the signal-like bins, the validity of the technique can be verified for that particular background category. The comparisons for the single-photon, single-lepton and low-$\Delta\phi$ control regions can be seen in Figures 4.7, 4.8, 4.9, respectively. $\kappa$ in bottom panel is defined as the ratio of the prediction to the true event yield (as usual). These comparisons are used for validation of the background estimation technique only, they are not used in the estimation.

### 4.5.3 $\kappa$ as a Correction to the Estimation

To correct for the observed non-closure of the background estimation method a correction, $\kappa$ is applied to the SM background prediction. $\kappa$ is obtained by dividing the MC yields for the signal region by that predicted: $\kappa \equiv A_{mc} / \ (B_{mc} \cdot \ (\frac{C_{mc}}{D_{mc}}\ )\ )$. There are 2x3=6 values of $\kappa$, one for each signal bin. The corrections are applied as follows: $A_{pred} = B_{obs} \cdot \ (\frac{C_{obs}}{D_{obs}}\ ) \cdot \kappa$.

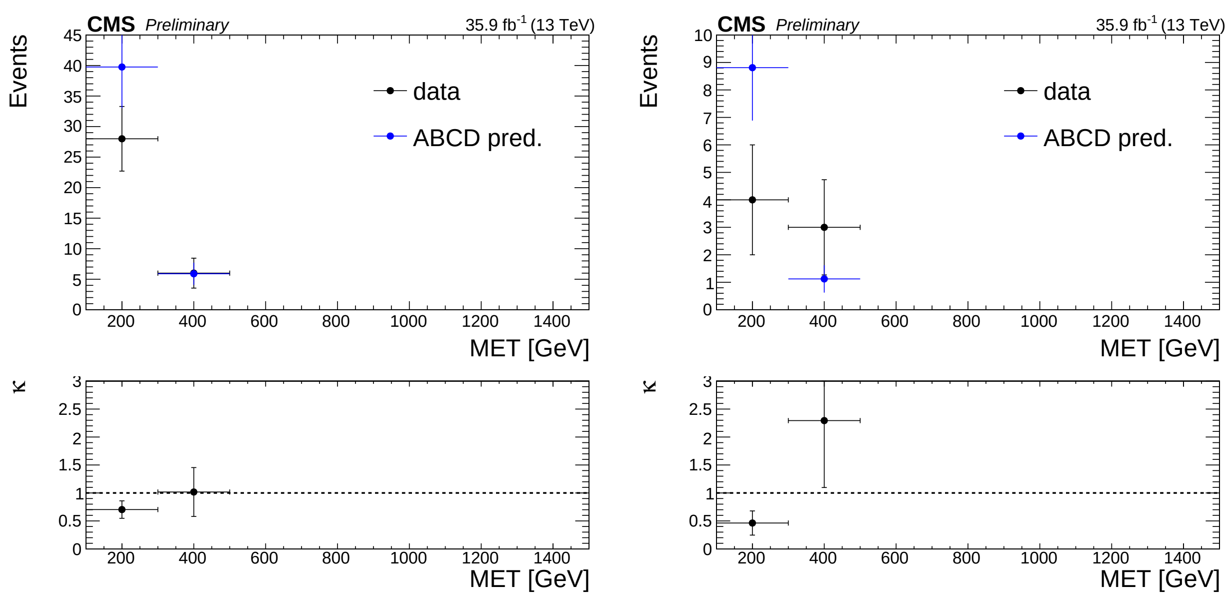
Taking into account the information from all the validation regions, we have obtained a set of scale-factors that can be applied to the signal region to test the closure in MC that mimics the data. The dependence of the scale-factors on MET is studied in a looser phase-space with AK8 jet $p_T > 170\,\text{GeV}$. The Low $\Delta\phi$ scale factors show no MET dependence and are determined integrated over MET$> 300\,\text{GeV}$. The Single-lepton region does show MET dependence. The scale-factors




(a) The single Higgs tag region ($A_1$).

(b) The double Higgs tag region ($A_2$).

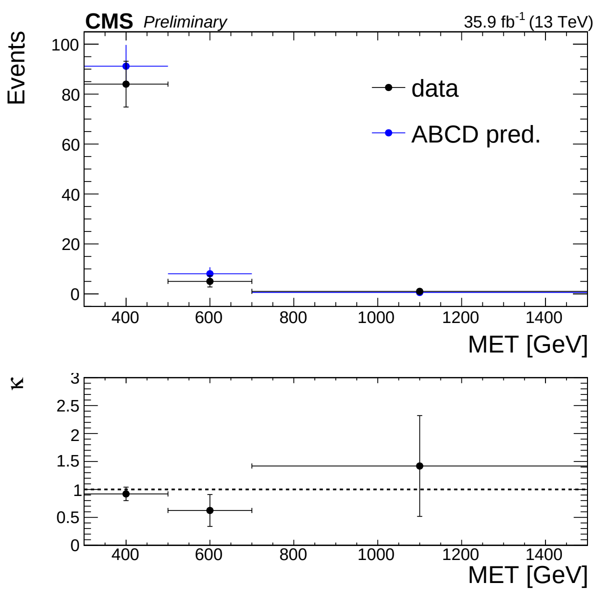
Figure 4.7: Closure within the single-photon data control region.



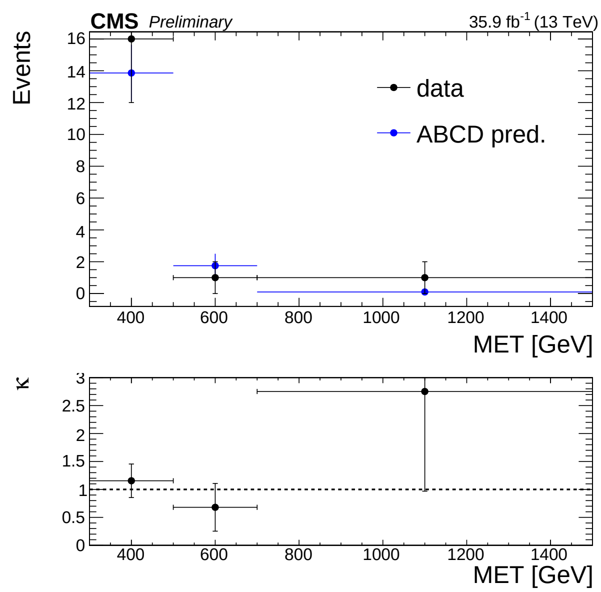

(a) The single Higgs tag region ($A_1$).

(b) The double Higgs tag region ($A_2$).

Figure 4.8: Closure within the single-lepton (e, $\mu$) data control region.




(a) The single Higgs tag region ($A_1$).




(b) The double Higgs tag region ($A_2$).

Figure 4.9: Closure within the low-$\Delta\phi$ data control region.

that are determined in MET bins are shown in Table 4.5. The scale factors derived from each of the validation regions are shown in Table 4.6.

These scale factors are then applied directly to the signal region MC to check $\kappa$ with a composition that better reflects the data. Since most of the scale factors are less than one the background decreases in Figure 4.6 relative to Figure 4.10 but still preserves closure so that $\kappa$ is statistically compatible with unity. A distribution of $\kappa$ is derived by throwing gaussian toys for each of the scale factors. We apply the scale factors from each toy to obtain a distribution of $\kappa$.

Table 4.5: Summary of the Single Lepton Validation region scale-factors for each of the MET bins in the anti-tag control regions.

| Single Lepton $C_{SF}$ | | |
|---|---|---|
| MET [300,500] | [500,700] | $>700$ GeV |
| $0.47 \pm 0.05$ | $0.54 \pm 0.15$ | $0.18 \pm 0.1$ |
| Single Lepton $D_{SF}$ | | |
| $0.49 \pm 0.02$ | $0.40 \pm 0.05$ | $0.35 \pm 0.08$ |

Figure 4.10 shows the modified MET distributions in the 6 signal regions after correcting simulation with the scale factors derived from the data control regions.

### 4.5.4 Sideband Yields & Predictions

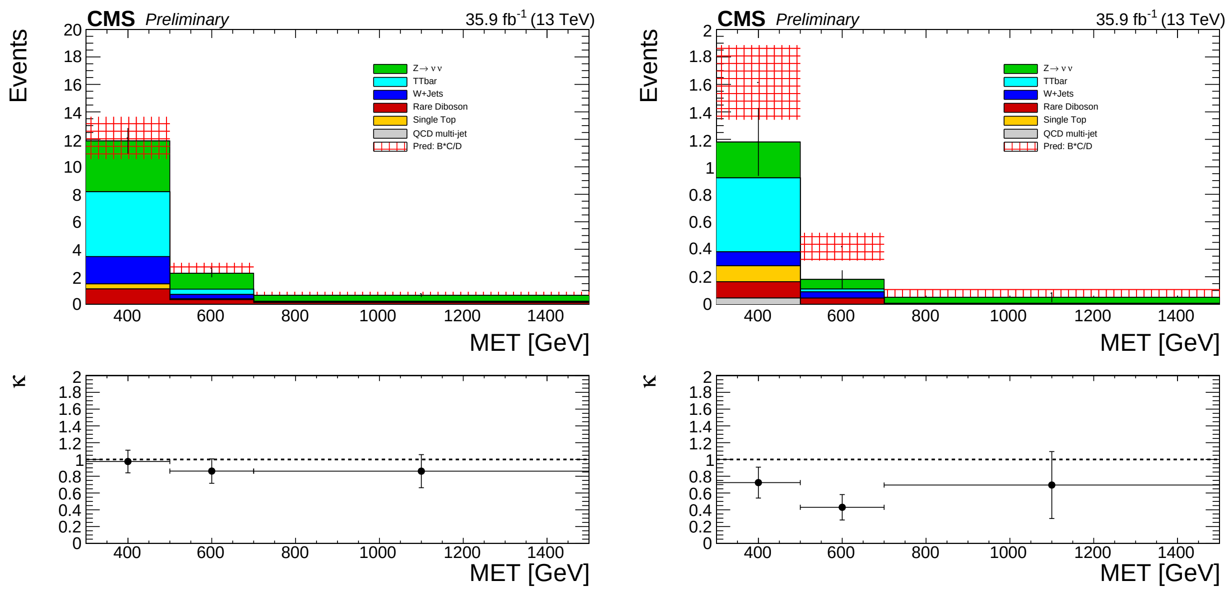
Observed yields in data for the 4x3=12 sideband regions (e.g. B, C, D in Figure 4.5) are seen in Table 4.7, alongside the prediction and $\kappa$ for each signal bin.

## 4.6 Signal Systematics

We consider a variety of systematic uncertainties on the signal efficiency and distribution. Some are common to more inclusive SUSY analyses [10] and there are additional systematics related to Double-b tagging efficiency and the effect of the pruned mass scale and resolution on the signal efficiency.

Table 4.6: Summary of the Validation region scale-factors derived integrated over the MET search bins. They are computed for each of the regions defined at the beginning of the Section.

| Low $\Delta\phi$ | | | | | |
|---|---|---|---|---|---|
| $A_{SF}^{1H}$ | $A_{SF}^{2H}$ | $C_{SF}$ | $B_{SF}^{1}$ | $B_{SF}^{2}$ | $D_{SF}$ |
| $1.1 \pm 0.33$ | $0.85 \pm 0.12$ | $0.93 \pm 0.1$ | $0.88 \pm 0.04$ | $1.2 \pm 0.16$ | $0.71 \pm 0.027$ |
| Single Lepton | | | | | |
| $A_{SF}^{1H}$ | $A_{SF}^{2H}$ | $C_{SF}$ | $B_{SF}^{1}$ | $B_{SF}^{2}$ | $D_{SF}$ |
| $0.61 \pm 0.04$ | $0.59 \pm 0.08$ | MET dependent | $0.59 \pm 0.016$ | $0.71 \pm 0.04$ | MET dependent |
| Photon | | | | | |
| $A_{SF}^{1H}$ | $A_{SF}^{2H}$ | $C_{SF}$ | $B_{SF}^{1}$ | $B_{SF}^{2}$ | $D_{SF}$ |
| $0.61 \pm 0.088$ | $0.75 \pm 0.29$ | $0.5 \pm 0.07$ | $0.98 \pm 0.094$ | $2.58 \pm 0.63$ | $0.71 \pm 0.035$ |




(a) The single Higgs tag region ($A_1$).

(b) The double Higgs tag region ($A_2$).

Figure 4.10: MET distributions and the predictions in the signal regions using simulation with scale factors from data.

Table 4.7: Yields in each of the 6 analysis regions.

| $N_H$ | $p_T^{miss}$ (GeV) | $\kappa$ | $A_{predicted}$ | A | B | C | D |
|---|---|---|---|---|---|---|---|
| 1 | 300 – 500 | $0.98 \pm 0.11$ | $17.7 \pm 3.8$ | 15 | 112 | 44 | 273 |
| 1 | 500 – 700 | $0.86 \pm 0.16$ | $3.4 \pm 1.5$ | 2 | 20 | 12 | 60 |
| 1 | >700 | $0.86 \pm 0.17$ | $0.61 \pm 0.45$ | 1 | 5 | 4 | 28 |
| 2 | 300 – 500 | $0.73 \pm 0.14$ | $1.52 \pm 0.57$ | 1 | 13 | 44 | 273 |
| 2 | 500 – 700 | $0.43 \pm 0.12$ | $0.09 \pm 0.08$ | 0 | 1 | 12 | 60 |
| 2 | >700 | $0.62 \pm 0.30$ | $0.09\ ^{+0.11}_{-0.09}$ | 0 | 1 | 4 | 28 |

- **Luminosity:** The recommendation for the 2016 dataset is currently a flat uncertainty of 2.5%.

- **Isolated track veto:** A flat uncertainty of 2% is assigned to the signal samples to account for any data/MC differences based on the study from the 2015 analysis [10].

- **MC statistics:** The signal MC sample statistical uncertainty is generally 2-4% .

- **Trigger efficiency:** The effect of the uncertainty on the signal yield is about 2%.

- **Pileup reweighting:** The sensitivity to the pileup distribution was studied for various benchmark signal models by comparing events with $n_{\text{vtx}} < 20$ (low PU) or $n_{\text{vtx}} \geq 20$ (high PU). Accordingly, no pileup reweighting is applied to the signal MC samples and no associated uncertainty is assessed.

- **ISR:** An ISR correction is derived from tt events, with a selection requiring two lep- tons (electrons or muons) and two b-tagged jets, implying that any other jets in the event arise from ISR. The correction factors are 1.000, 0.920, 0.821, 0.715, 0.662, 0.561, 0.511 for NISR = 0, 1, 2, 3, 4, 5, 6+. The corrections are applied to the simulated signal jet samples with an additional normalization factor, typically 1.15 (depending on the signal model), to ensure the overall cross section of the sample remains constant. The systematic uncertainty in these corrections is chosen to be half of the deviation from unity for each correction factor. The effect on the yield ranges from 0.01%, with the largest effect at high MET.

- **Scales:** The uncertainty is calculated using the envelope of the weights from varying the renormalization and factorization scales, $\mu_R$ and $\mu_F$,by a factor of 2 [11, 12]. The effect on the yield of is less than 0.1%.

- **Jet Energy Corrections:** The jet energy corrections (JECs) are varied using the $p_T$- and $\eta$-dependent jet energy scale uncertainties from the official database. These variations are propagated into the various jet-dependent variables, including: HT, MET, $\Delta\phi(\text{MET}, j_i)$. The overall effect is less than 1%.

- **Jet Energy Resolution:** The jet momenta in the MC samples are smeared to match the jet energy resolution in data. The smearing factors are varied according to the uncertainties on the jet energy resolution measurements. These variations are propagated into the various jet-dependent variables, including: HT, MET, $\Delta\phi(\mathrm{MET}, j_i)$. The overall effect ranges from 0.01%.

- **PDFs:** The LHC4PDF prescription for the uncertainty on the total cross section is included as $\pm1$ sigma bands in the results plots. No additional uncertainty is considered for the uncertainty in the acceptance due to PDFs, as per SUSY group recommendation.

The above signal systematics are applied as an uncertainty on the signal normalization. These uncertainties are in general small. The main signal systematics come from the AK8 Jet Double-b tagging efficiency data/MC scale factors and the uncertainty on the pruned mass resolution. The AK8 Jet Double-b tagging efficiency has an uncertainty which is propagated to the signal efficiency. This uncertainty is applied as a shape uncertainty across the Higgs tag regions and the anti-tag region. Also the pruned jet mass scale and resolution uncertainties are propogated to the final signal efficiency using POG recommendations. The pruned mass scale factor is derived using W-jets in semi-leptonic $\mathrm{t\bar{t}}$ and extrapolting to the Higgs mass. This uncertainty is assigned a shape uncertainty on the signal mass window and the sideband.

- A data/MC scale-factor is derived from double-muon tag data selected with HLT Trigger `HLT_BTagMu_AK8Jet300_Mu5_v` and muon enriched QCD Monte-Carlo. The scale factors have mainly a statistical error along with a smaller set of systematic errors due to shape systematics, Jet-Energy scale uncertainty, Pile-up corrections, uncertainty on the number of tracks, uncertainty of b-fragmentation and c-fragmentation, and the uncertainty on $K_s$ and $\Lambda$ fraction.

- The pruned mass scale-factor is derived by comparing the efficiency to select W-jets in data and MC within a mass window of $[65, 85]$ GeV. The fit for the gaussian resolution of the

W-mass peak is shown in Figure 4.11 and the fit results are shown in Table 4.8. The mass scale between MC and data is consistent though MC predicts a narrower mass resolution compared to data. The jet mass in each event is smeared to mimic the pruned jet mass resolution in data and an uncertainty is assigned based on the ratio of efficiencies between the smeared and un-smeared cases. [13]

The summary of the signal systematics and their effect on the signal yields is shown in Table 4.9. The dominant effect is from the mass resolution uncertainty.

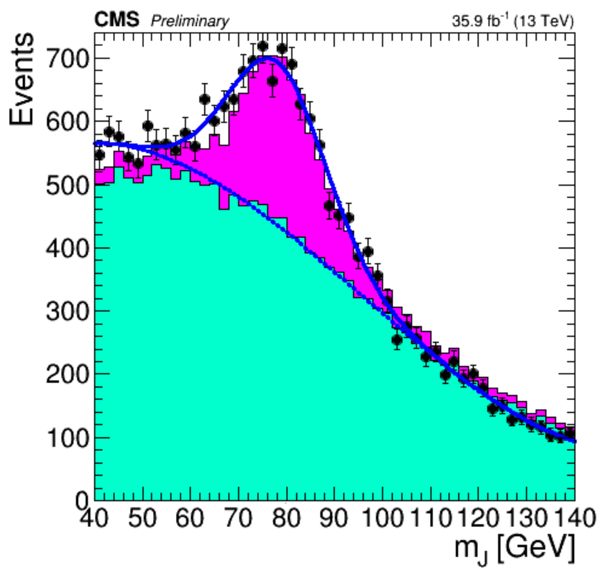



Figure 4.11: Pruned jet mass in in semi-leptonic $t\bar{t}$ events. The mass peak for the W-jets is used to derive the mass resolution uncertainty.

| Data | |
|---|---|
| Mean | $78.2 \pm 0.46$ |
| Sigma | $10.10 \pm 0.671$ |
| $t\bar{t}$ MC | |
| Mean | $78.4 \pm 0.35$ |
| Sigma | $7.23 \pm 0.48$ |

Table 4.8: Fit results for W-mass resolution in data and MC

| Unc. on Normalization | |
|---|---|
| Systematic | % Effect on yields |
| Luminosity | 2.6% |
| Trigger Eff. | 2.0% |
| Iso. Track Veto | 2% |
| ISR modeling | 0.01% |
| PDF Scale | 0.1% |
| JEC | 1% |
| JER | 0.01% |
| MC Stat | 1-4% |
| Shape Unc. | |
| Double-b SF | 6% |
| Mass Resolution | 1-15% |

Table 4.9: Summary of signal shape and normalization uncertainties.

## 4.7 Yields in the Signal Regions and Exclusion Curves

After unblinding the 4x3=12 sideband regions and performing the background estimation the 2x3=6 signal regions were unblinded. The observed yields, along with the SM background predictions, are seen in Table 4.10. Our signal region yields are consistent with the SM background expectation. Additionally, Table 4.10 shows the expected signal yields for two model points corresponding to gluino $\tilde{g}$ masses of 2000 or 1800 GeV; the mass of the neutralino $\tilde{\chi}_1^0$ is fixed at 1 GeV; the mass splitting between the gluino $\tilde{g}$ and neutralino $\tilde{\chi}_2^0$ is fixed at 50 GeV.

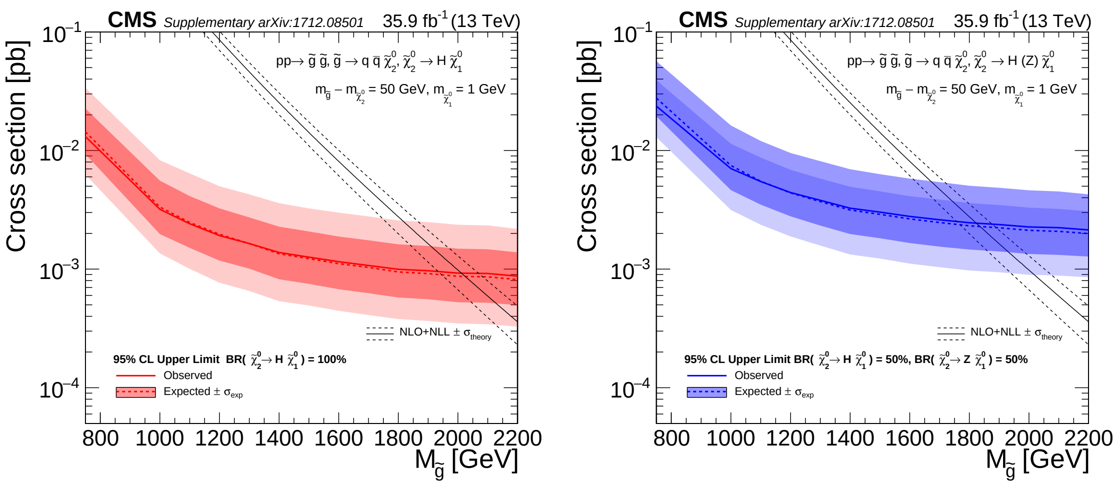
Table 4.10: Signal yields and SM background predictions

| MET bin | ABCD Pred | $\kappa$ | Total Pred. | Obs. | T5HH(2000) | T5HZ(1800) |
|---|---|---|---|---|---|---|
| 1-Higgs Tag | | | | | | |
| MET [300,500] | $18.05 \pm 3.39$ | $0.98 \pm 0.11$ | $17.68 \pm 3.85$ | 15 | 0.24 | 0.75 |
| MET [500,700] | $4 \pm 1.54$ | $0.86 \pm 0.16$ | $3.44 \pm 1.47$ | 2 | 0.32 | 0.98 |
| MET $> 700$ | $0.71 \pm 0.50$ | $0.86 \pm 0.17$ | $0.61 \pm 0.45$ | 1 | 2.13 | 4.34 |
| 2-Higgs Tag | | | | | | |
| MET [300,500] | $2.09 \pm 0.67$ | $0.73 \pm 0.14$ | $1.52 \pm 0.57$ | 1 | 0.17 | 0.35 |
| MET [500,700] | $0.2 \pm 0.20$ | $0.43 \pm 0.12$ | $0.09^{+0.08}_{-0.08}$ | 0 | 0.23 | 0.44 |
| MET $> 700$ | $0.14 \pm 0.16$ | $0.62 \pm 0.30$ | $0.09^{+0.11}_{-0.09}$ | 0 | 1.36 | 1.98 |

Interpreting our results in the context of the T5HH or T5ZH models, the absence of signal allows us to place lower limits on the mass of the gluino $\tilde{g}$. For the statistical treatment, we use the Higgs combination tool to encode the ABCD approach in the likelihood. In this approach, the data card for one search bin contains the observed number of events and the expected signal and background in each of the ABCD regions. A likelihood function is built that contains these ABCD regions and explicitly encodes the relation $A = \kappa B \frac{C}{D}$ and a Gaussian nuisance is assigned for the uncertainty on $\kappa$. The likelihood for each search bin can be described by:

$$\mathcal{L} = \prod_{i}^{ABCD} Poisson\left(n_i | bkg_i + r \cdot sig_i\right) \times \prod_{j}^{nuisances} Constraints\left(\theta_j, \hat{\theta}_j\right) \tag{4.1}$$

where the 4 regions are modeled by Poisson distribution and the term Constraints refers to either Gaussian distributions for the $\kappa$ uncertainties or log normal distributions that model the signal systematics. The expected and observed limits are then calculated based on the asymptotic approximation of the profile likelihood ratio using the CLs criterion to place limits at the 95% confidence level. These exclusion curves are seen in Figure 4.12. We are able to place lower limits at 95% confidence level on the gluino $\tilde{g}$ mass at 2010 and 1825 GeV for the T5HH and T5ZH models, respectively. The weaker limit for the T5ZH model is due to the smaller branching fraction of the Z boson to b-quarks and our choice of signal mass window not being optimal for Z reconstruction.




(a) T5HH (b) T5ZH

Figure 4.12: Observed and expected limits on the gluino cross section.

## 4.8 Reinterpretation

In Section 4.7 our results were presented in the context of limit-setting for the T5HH and T5ZH models. Many such SMS models exist within the MSSM which predict the production of high-$p_T$ bosons and it is therefore important to include information necessary to make predictions of yields for different final states. This is aided by providing the user with efficiencies for $b\bar{b}$ -tagging and mass-tagging efficiencies of the AK8 jets. Tagging efficiencies for the three largest decay channels for the H boson are seen in 4.13. Tagging efficiencies for the Z boson are seen in Figure 4.14, the much lower mass-tagging efficiency for the Z is due to our choice of signal mass window [85, 135 GeV] not being optimal for Z boson reconstruction. These are used to calculate the expected yields in the 6 analysis regions when performing a reinterpretation of the analysis using different final states.

For each event, the yield in each bin can be predicted by first forming the following weights using the tagging efficiencies for the leading two jets, as seen below.

- double mass tag weight = $\text{jet}_0$_signalmass * $\text{jet}_1$_signalmass
- anti mass tag weight = ($\text{jet}_0$_sidebandmass * $\text{jet}_1$_signalmass) + ($\text{jet}_0$_signalmass * $\text{jet}_1$_sidebandmass) + ($\text{jet}_0$_sidebandmass * $\text{jet}_1$_sidebandmass);
- double bb tag weight = $\text{jet}_0$_bbtag * $\text{jet}_1$_bbtag;
- single bb tag weight = ($\text{jet}_0$_bbtag * (1-$\text{jet}_1$_bbtag)) + ((1-$\text{jet}_0$_bbtag)*$\text{jet}_1$_bbtag);
- anti bb tag weight = (1-$\text{jet}_0$_bbtag) * (1-$\text{jet}_1$_bbtag)

These weights are then combined in the following manner to determine the yields across each of the 6 bins for a single event.

- A1 weight = (single bb tag weight) * (double mass tag weight)
- A2 weight = (double bb tag weight) * (double mass tag weight)
- B1 weight = (single bb tag weight) * (anti mass tag weight)
- B2 weight = (double bb tag weight) * (anti mass tag weight)
- C weight = (anti bb tag weight) * (double mass tag weight)
- D weight = (anti bb tag weight) * (anti mass tag weight)

Following this prescription, the authors found that the largest deficit was in the D region, with a difference of -36% difference from nominal, as seen in Table 4.11. The greatest over-prediction is found in the B2 region, with a surplus of +8.2% events relative to nominal. The closure in the other bins fall somewhere in this range. As a cross-check to your yield estimates, consult Table

4.12 which shows the true signal event efficiencies for the T5qqqqHH model with a gluino mass of 2200 GeV.

Table 4.11: Comparison of the true reco-level event yield with those obtained via the prediction following the prescription above. The columns with the RECO or GEN labels are the prediction using RECO or GEN event variables only, respectively. The prediction was made using the T5qqqqHH MC with a gluino mass of 2200 GeV.

| | RECO "truth" | RECO prediction | GEN prediction |
|---|---|---|---|
| Baseline | 4.08 | 3.46 (-15%) | 3.53 (-16%) |
| A1 | 1.21 | 1.18 (-2.3%) | 1.26 (+3.6%) |
| A2 | 0.777 | 0.748 (-3.7%) | 0.815 (+4.7%) |
| B1 | 0.802 | 0.703 (-12%) | 0.664 (-21%) |
| B2 | 0.322 | 0.338 (+5.0%) | 0.350 (+8.2%) |
| C | 0.498 | 0.473 (-4.9%) | 0.487 (-2.1%) |
| D | 0.478 | 0.353 (25%) | 0.308 (-36%) |

## 4.9 Conclusions

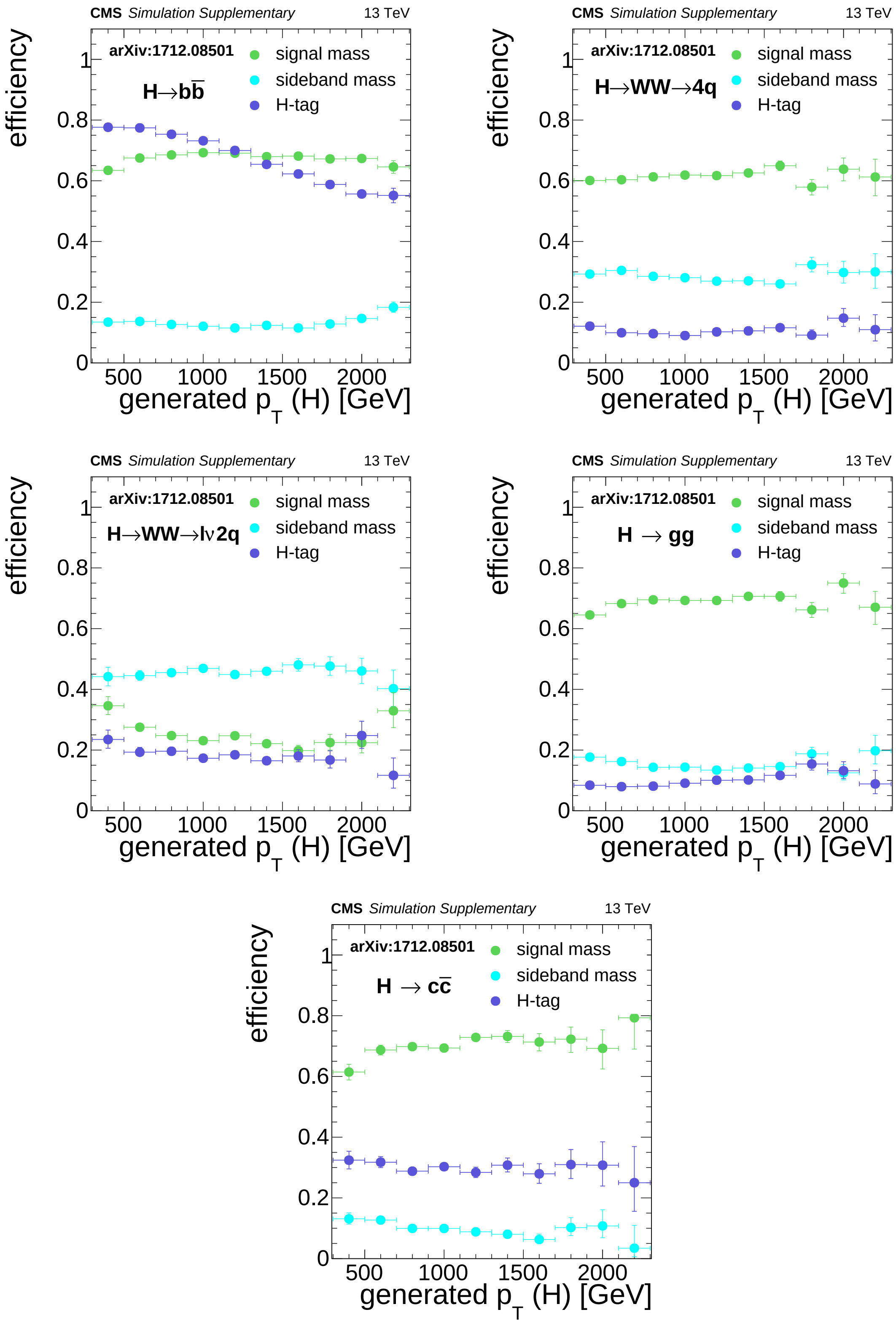



Figure 4.13: Efficiencies for an AK8 jet originating from Higgs boson decay, relative to baseline selection. The "signal mass" curve represents the probability the jet will fall within the mass region [85, 135 GeV]. The "sideband mass" curve represents the probability the jet will fall within the mass region [50, 85 GeV] or [135, 250 GeV]. The "H-tag" curve represents the probability the jet have a double-b discriminator value greater than 0.3, for jets within the mass region [50, 250 GeV]. The efficiencies were derived using the T5qqqqZH MC with a gluino mass of 2200 GeV.

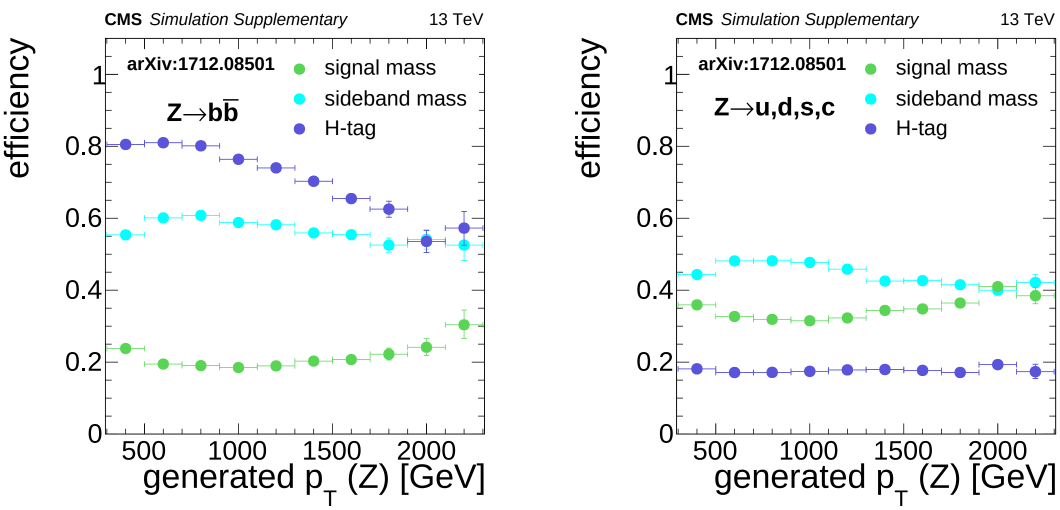



Figure 4.14: Efficiencies for an AK8 jet originating from Z boson decay, relative to baseline selection. The "signal mass" curve represents the probability the jet will fall within the mass region [85, 135 GeV]. The "sideband mass" curve represents the probability the jet will fall within the mass region [50, 85 GeV] or [135, 250 GeV]. The "H-tag" curve represents the probability the jet have a double-b discriminator value greater than 0.3, for jets within the mass region [50, 250 GeV]. The efficiencies were derived using the T5qqqqZH MC with a gluino mass of 2200 GeV.

Table 4.12: Signal efficiencies for an event to land in a given analysis bin. The efficiencies were derived using the T5qqqqHH MC with a gluino mass of 2200 GeV. Choosing a gluino mass of 1800 GeV decreases the efficiencies by a relative 5%.

| | Baseline | A1 | A2 | B1 | B2 | C | D |
|---|---|---|---|---|---|---|---|
| $p_{\mathrm{T}}^{\mathrm{miss}} < 300\,\mathrm{GeV}$ | 32% | 9.4% | 6.0% | 6.2% | 2.5% | 3.9% | 3.7% |
| $300 < p_{\mathrm{T}}^{\mathrm{miss}} < 500\,\mathrm{GeV}$ | 2.7% | 0.78% | 0.52% | 0.54% | 0.25% | 0.31% | 0.30% |
| $500 < p_{\mathrm{T}}^{\mathrm{miss}} < 700\,\mathrm{GeV}$ | 3.5% | 1.0% | 0.65% | 0.72% | 0.28% | 0.43% | 0.40% |
| $p_{\mathrm{T}}^{\mathrm{miss}} > 700\,\mathrm{GeV}$ | 26% | 7.6% | 4.9% | 5.0% | 2.0% | 3.1% | 3.0% |

# Chapter 5

# Conclusions

A chapter describing the conclusions. The analysis presented in this thesis is complete and has been published in a peer-reviewed journal[14].

# Bibliography


[1] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," Phys. Lett. B, vol. 716, no. 1, pp. 30 – 61, 2012.

[2] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," JINST, vol. 12, p. P10003, 2017.

[3] CMS Collaboration, "The CMS trigger system," JINST, vol. 12, p. P01020. 122 p, Sep 2016.

[4] CMS Collaboration, "Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV," Phys. Rev. D, vol. 96, p. 032003, 2017.

[5] CMS Collaboration, "Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV," 2016.

[6] S. P. Martin, "A Supersymmetry Primer." arXiv:hep-ph/9709356v7, Jan 2016.

[7] CMS Collaboration, "Interpretation of searches for supersymmetry with simplified models," Phys. Rev. D, vol. 88, p. 052017, 2013.

[8] CMS Collaboration, "Search for higgsino pair production in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ in final states with large missing transverse momentum and two Higgs bosons decaying via $\mathrm{H} \to b\bar{b}$," Phys. Rev. D, vol. 97, p. 032007, 2018.

[9] M. Cacciari, G. P. Salam, and G. Soyez, "The anti- k t jet clustering algorithm," Journal of High Energy Physics, vol. 2008, no. 04, p. 063, 2008.

[10] CMS Collaboration, "Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV," CMS Physics Analysis Note AN-2016/350, CERN, 2017.

[11] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, "The $\mathrm{t}\bar{\mathrm{t}}$ cross-section at 1.8 TeV and 1.96 TeV: A Study of the systematics due to parton densities and scale dependence," JHEP, vol. 04, p. 068, 2004.

[12] S. Catani, D. de Florian, M. Grazzini, and P. Nason, "Soft gluon resummation for Higgs boson production at hadron colliders," JHEP, vol. 07, p. 028, 2003.

[13] I. Belotelov, I. Golutvin, D. Bourilkov, A. Lanyov, E. Rogalev, M. Savina, and S. Shmatov, "Identification and calibration of boosted hadronic w/z bosons at 13 tev," CMS Note 2016/215, 2016.

[14] CMS Collaboration, "Search for physics beyond the standard model in events with high-momentum Higgs bosons and missing transverse momentum in proton-proton collisions at 13 TeV," Phys. Rev. Lett., vol. 120, p. 241801, 2018.