

**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

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 (SM) 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  $35.9 \text{ fb}^{-1}$ , were collected with the CMS detector at the Large Hadron Collider in proton-proton collisions at the center-of-mass energy  $\sqrt{s} = 13 \text{ TeV}$ . 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 SM 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  $Z$  bosons, and the lightest SUSY particle (LSP). The decay chain proceeds through an intermediate SUSY particle, assumed to be the second lightest in the theory (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

### Chapter

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Standard Model of Particle Physics</b>	<b>5</b>
2.1	Introduction . . . . .	5
2.2	Quantum Electrodynamics . . . . .	6
2.3	Quantum Chromodynamics . . . . .	7
2.3.1	Confinement, Hadrons, & the Quark Model . . . . .	9
2.4	Electroweak Theory . . . . .	10
2.4.1	Introduction . . . . .	10
2.4.2	The Higgs Mechanism . . . . .	12
2.4.3	Quark Mixing - the CKM Matrix . . . . .	15
2.4.4	Brief Electroweak Summary . . . . .	16
2.5	Parameters of the Standard Model . . . . .	16
2.6	Neutrino Mass . . . . .	17
<b>3</b>	<b>The Minimal Supersymmetric Model</b>	<b>19</b>
<b>4</b>	<b>The Large Hadron Collider</b>	<b>25</b>
<b>5</b>	<b>The CMS Detector</b>	<b>30</b>
5.1	Silicon Tracker . . . . .	31

5.1.1	Pixel Detector	34
5.1.2	Strips Detector	35
5.2	Electromagnetic Calorimeter	35
5.2.1	Preshower	37
5.3	Hadronic Calorimeter	37
5.4	Solenoidal Magnet	38
5.5	Muon System	40
5.5.1	Drift Tubes	41
5.5.2	Cathode Strip Chambers	41
5.5.3	Resistive Plate Chambers	43
5.6	Trigger System	43
<b>6</b>	<b>Event Reconstruction</b>	<b>46</b>
6.1	Basic Elements From the Detector	46
6.1.1	Tracks	46
6.1.2	ECAL & HCAL Clusters	48
6.1.3	Muon Tracks	48
6.2	Obtaining a Particle-level Description	48
6.3	Additional High-Level Objects	51
6.3.1	Jets	51
6.3.2	b-tagging of Jets	51
6.3.3	Invisible Particles $\rightarrow p_T^{\text{miss}}$	52
6.3.4	$\pi^0$ meson	52
<b>7</b>	<b>Search for new physics using boosted H bosons and missing energy</b>	<b>54</b>
7.1	Motivation & Strategy	54
7.2	Baseline Selection and Object Definition	55
7.3	Dataset & Trigger	58

7.4	Event Simulation . . . . .	60
7.4.1	Standard Model Processes . . . . .	60
7.4.2	Signal Models . . . . .	60
7.5	Event Binning & Background Estimation . . . . .	62
7.5.1	Control Regions within Data . . . . .	66
7.5.2	$\kappa$ as a Correction to the Estimation . . . . .	67
7.5.3	Sideband Yields & Predictions . . . . .	77
7.6	Signal Systematics . . . . .	77
7.7	Observed Yields in the Signal Regions . . . . .	81
7.8	Exclusion Curves & Mass Limits . . . . .	81
<b>8</b>	<b>Conclusions</b>	<b>86</b>

## Appendix

<b>A</b>	b $\bar{b}$ -tagging of AK8 Jets	<b>87</b>
<b>B</b>	Reinterpretation	<b>92</b>
<b>C</b>	Determining b $\bar{b}$ -tagging Scale Factors for W jets in t $\bar{t}$ Events	<b>97</b>
	<b>Bibliography</b>	<b>102</b>

## Tables

### Table

2.1	Summary of field content within the SM. . . . .	6
3.1	Summary of the additional particle content within the MSSM. . . . .	20
3.2	Mixing of the supersymmetric electroweak fields. . . . .	20
4.1	The stages of proton beam acceleration for the LHC at CERN. . . . .	26
6.1	Seeding requirements for each step in the iterative track reconstruction. . . . .	47
7.1	SM samples used in the analysis. . . . .	61
7.2	Summary of the control region scale-factors integrated over $p_T^{\text{miss}}$ . . . . .	73
7.3	Single lepton control region scale-factors in the anti-tag sideband region. . . . .	73
7.4	The $\kappa$ factor computed by throwing Gaussian toys for the scale factors. . . . .	74
7.5	Corrected yields in the signal regions. $\kappa = \text{AD} / \text{BC}$ . . . . .	76
7.6	Corrected yields in the sideband regions. $\kappa = \text{AD} / \text{BC}$ . . . . .	76
7.7	Sideband region yields, $\kappa$ , and background predictions for the 6 signal bins. . . . .	77
7.8	Fit results for W-mass resolution in data and simulation. . . . .	81
7.9	Summary of signal shape and normalization uncertainties. . . . .	82
7.10	Yields and predicted background in the signal regions. . . . .	83
A.1	Data/MC scale factors for AK8 jet $b\bar{b}$ -tagging with the loose working point $\mathcal{D}_{bb} > 0.3$ . .	91

B.1	Effective event weights for $b\bar{b}$ and mass tagging of AK8 jets. . . . .	93
B.2	Predicted yields in the 6 signal bins using reinterpretation prescription. . . . .	94
C.1	W jet $b\bar{b}$ mistag rates in data and simulation, inclusive in $p_T$ . . . . .	100
C.2	Summary of scale factors for $b\bar{b}$ -tagging W jets in $t\bar{t}$ events. . . . .	101

## Figures

### Figure

1.1	The Standard Model of Particle Physics.	2
2.1	SU(3) flavor symmetry multiplets predicted by the quark model.	10
3.1	Interactions between SM and MSSM particles.	22
3.2	Tree-level gluino pair-production mechanisms.	23
3.3	Previous results for searches of gluino-mediated supersymmetry.	24
4.1	The CERN accelerator complex.	26
4.2	The number of pileup interactions in a 2016 event recorded by CMS.	28
4.3	The integrated luminosity of CMS over its lifetime.	29
5.1	A diagram of the CMS detector.	32
5.2	A diagram of the CMS detector in the $r\text{-}\phi$ plane; particle signatures are shown.	32
5.3	A diagram of the CMS detector in the $r\text{-}z$ plane.	33
5.4	The CMS silicon tracker	34
5.5	The CMS electromagnetic calorimeter.	36
5.6	The CMS hadron calorimeter.	39
5.7	A simulation of the 4T CMS magnetic field.	40
5.8	The CMS muon drift tube detector.	42
5.9	The CMS muon cathode strip chambers.	44

5.10	The CMS muon resistive plate chamber.	44
7.1	Diagrams of the benchmark models used for motivation of the targeted signal.	56
7.2	Distributions of the $b\bar{b}$ -tagging discriminator and jet mass for AK8 jets.	59
7.3	Generator level distributions for the H bosons in the T5HH model.	63
7.4	Partitioning of the signal and sideband regions for event binning.	64
7.5	MC distributions of $p_T^{\text{miss}}$ after baseline selection.	65
7.6	$p_T^{\text{miss}}$ distributions and predictions in the signal regions using simulation only.	66
7.7	Comparisons of the predicted and observed yields within the data control regions.	68
7.8	Signal and sideband yields in the single photon control region.	70
7.9	Signal and sideband yields in the single lepton control region.	71
7.10	Signal and sideband yields in the low- $\Delta\phi$ control region.	72
7.11	Signal region $p_T^{\text{miss}}$ distributions using scale-factor corrected simulation.	75
7.12	Sideband region $p_T^{\text{miss}}$ distributions of data and scale-factor corrected simulation.	78
7.13	Pruned jet mass in semi-leptonic $t\bar{t}$ events.	82
7.14	Observed yields in the signal regions.	83
7.15	The single event in the $A_2$ region.	84
7.16	Observed and expected limits on the gluino cross section.	85
A.1	ROC curve of the signal efficiency and mistag rate for the $b\bar{b}$ tagger.	88
A.2	$H \rightarrow b\bar{b}$ tagging efficiencies in signal events.	89
A.3	Signal efficiency to be in the single or double Higgs tag event category.	90
B.1	Diagram for compiling the weights into each of the 6 analysis bins.	94
B.2	Efficiencies for an AK8 jet originating from H boson decay, relative to baseline selection.	95
B.3	Efficiencies for an AK8 jet originating from Z boson decay, relative to baseline selection.	96
C.1	Diagram of a semileptonic $t\bar{t}$ event.	98
C.2	AK8 jet mass and $b\bar{b}$ -tagging discriminator distributions in data and simulation.	99

## Chapter 1

### Introduction

The emerging picture of the last century is that the elements described by the Periodic Table are themselves not the most fundamental forms found in Nature. We have learned that these elements are made of atoms consisting of bound states of protons, neutrons, and electrons. We have discovered that these protons and neutrons themselves are made of more fundamental components called quarks and gluons. And so as the Periodic Table before it, the Standard Model of particle physics (SM) seen in Figure 1.1, has been developed to provide an organizing principle for what we currently understand to be the correct description of the fundamental particles and forces which form the Universe. The discovery of the Higgs boson in 2014 [2] completed the search for all the fundamental particles addressed within the theory, a truly monumental achievement. As time goes on, its experimental verification only grows stronger.

But, there are many both experimental and theoretical indications that the SM is not the final story. Cosmological observations hint at the presence of a “dark matter” in the Universe, a ubiquitous substance thought to account for 85% of the total matter budget in the Universe. Although its influence has been observed in gravitational lensing phenomena and galaxy rotation curves, we do not yet have a particle description of its nature. One theoretical motivation which points toward the SM being a part of some grander theory is known as a “fine-tuning” problem. Because the Higgs boson is a scalar particle (the only such fundamental in the SM), its mass receives quantum mechanical corrections which are strongly dependent on the ultraviolet cutoff  $\Lambda_{\text{UV}}$  and other high mass scales in the theory. As the most natural cutoff is likely at the GUT or Planck

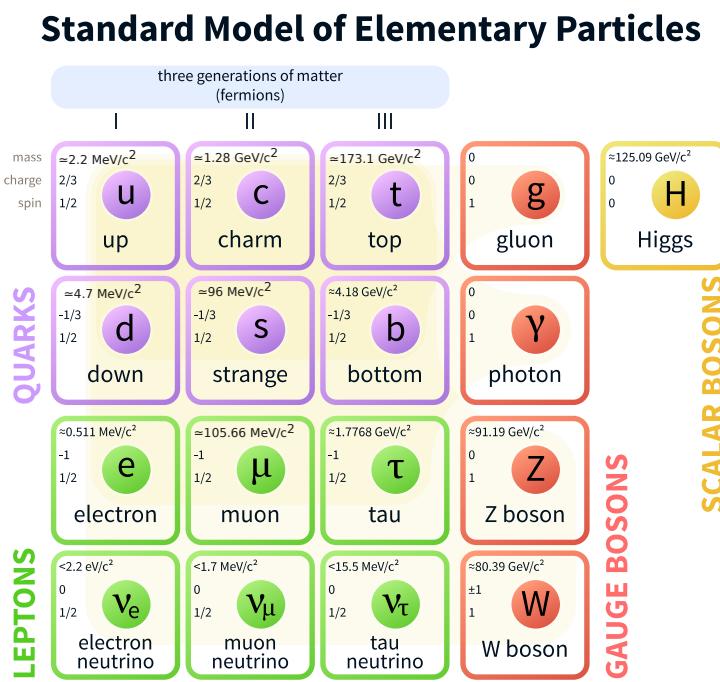


Figure 1.1: The Standard Model of Particle Physics.

scale, we would expect the Higgs boson mass to be extremely large, many orders of magnitude than it has experimentally been determined to be. This implies that there is some sort of “un-natural” collusion between the terms, which is somehow able to bring the mass to the electroweak scale.

Supersymmetry (SUSY) is an elegant extension to the SM which is able to address many of these issues by the introduction of a new symmetry of the Lagrangian which relates fermionic and bosonic states. The particle content of SUSY is over twice that of the SM as the theory requires a doubling of the particle content to give each of the SM particles a “partner” of spin differing by 1/2 unit. Among its phenomenology, the lightest particles of the theory, expected to remain stable, are able to serve as candidates for dark matter. The addition of these superpartners to the theory yield quantum mechanical corrections to the Higgs mass which enter with the opposite sign naturally are able to cancel the terms of high-order in  $\Lambda_{\text{UV}}$ . An exact supersymmetry requires that these superpartners have identical mass to their SM counterpart. This is obviously not the case and therefore, if SUSY is correct, these new particles must have acquired mass through some other means. These masses must be sufficiently large such that their production at our collider experiments has not allowed for their unambiguous detection. A major motivating force for the construction and experimentation at the LHC is that of finding evidence of SUSY. The analysis presented in this thesis presents one such search, focusing in one small parameter space of the vast possibilities.

The thesis is outlined as follows: A description of the Standard Model (SM) of particle physics, the theory describing the known fundamental forces and particles, is presented in Chapter 2. A description of the Minimal Supersymmetric Standard Model (MSSM), one extension to the SM able to provide answers to many of our questions, is presented in Chapter 3. A description of the Large Hadron Collider (LHC), the facility which acts as our source of high-energy proton-proton collisions, is presented in Chapter 4. A description of the CMS particle detector, responsible for the detection of particles and reconstruction of the proton interactions, is presented in Chapter 5. A description of how the data from the detector are reconstructed and identified as physical particles is presented in Chapter 6. The focus of the thesis, a description of the how the physics data can

be used to search for evidence of new particles such as those predicted by SUSY, is presented in Chapter 7. The conclusions are presented in Chapter 8. Appendix A presents a more detailed discussion of the  $b\bar{b}$ -tagging algorithm used in this analysis. Appendix B provides resources for additional physics interpretations of our results presented in Section 7. Appendix C presents the calculation of data-mc scale factors for  $b\bar{b}$ -tagging W boson jets, a partly independent topic from the rest of the thesis.

## Chapter 2

### The Standard Model of Particle Physics

#### 2.1 Introduction

The Standard Model (SM) is a mathematical framework for the quantum mechanical description of matter and physical forces. It provides a description of the weak, strong, and electromagnetic interactions as mediated via “messenger” particles (gravity has so far eluded a quantum mechanical description). The particles constituting matter can be classified as *quarks* or *leptons* - quarks participate in the weak and strong force, leptons participate in the weak force; any charged particle experiences the electromagnetic force. *Gauge bosons* are the particles responsible for the force interactions. The Higgs field is important for the generation of particle mass. This particle content has been seen in Figure 1.1, we now turn to the field theory approach for these objects.

Matter particles are massive spin-1/2 fermions which are represented by solutions to the free-particle Dirac equation generated by the following Lagrangian, with equation of motion:

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \\ i\gamma^\mu\partial_\mu\psi - m\psi &= 0\end{aligned}\tag{2.1}$$

respectively.

Particle interactions are generated by requiring the free-particle Lagrangian to be invariant under the action of different symmetry groups. Demanding local (gauge) invariance requires one to introduce spin-1 vector fields to the Lagrangian which couple with the fermions. The vector fields are to be identified with the generators of the symmetry group and act as the mediator of

Table 2.1: Summary of field content within the SM.

Gauge Sector	Matter Fields	Gauge Fields
<b>SU(3)</b>	u, d, c, s, t, b	$G_\mu^{1\dots 8} \rightarrow g_\mu^{1\dots 8}$
<b>SU(2) <math>\times</math> U(1)</b>	$Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} s \\ c \end{pmatrix}_L, \begin{pmatrix} t \\ b_L \end{pmatrix}_L$ $q = u, d, c, s, t, b$ $L_L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$ $\ell = (e, \mu, \tau)$	$W_\mu^{012}, B_\mu^0$ $W_\mu^0, B_\mu^0 \rightarrow Z_\mu^0, A_\mu^0$ $W_\mu^1, W_\mu^2 \rightarrow W_\mu^+, W_\mu^-$
<b>Higgs Sector</b>	$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix} \rightarrow h^0$	

the force via particle exchange. U(1) generates electromagnetism via interactions with photons. A combination of U(1) and SU(2) generates the electroweak theory, simultaneously describing the electromagnetic and weak nuclear force via interactions with  $W^\pm$  bosons, Z bosons and photons. SU(3) generates quantum chromodynamics, the theory of the strong nuclear force via interactions with gluons. These massive spin-1 gauge bosons are represented by solutions to the free particle Proca equations generated by the following Lagrangian, with equation motion:

$$\begin{aligned} \mathcal{L} &= -\frac{1}{16\pi} B^{\mu\nu} B_{\mu\nu} + \frac{1}{8\pi} m^2 B_\nu B^\nu \\ \partial^\mu \partial_\mu \psi - m^2 \psi &= 0 \end{aligned} \tag{2.2}$$

where  $B_{\mu\nu} \equiv \partial_\mu B_\nu + \partial_\nu B_\mu$  is known as the *energy-momentum tensor* representing the kinetic energy of the field.

The mathematical fields which are contained with the SM are seen in Table 2.1. This table will serve as a reference for the remainder of the chapter.

## 2.2 Quantum Electrodynamics

Quantum electrodynamics describes the interactions of particles with electric charge. Beginning with a free-particle Dirac fermion we see the Lagrangian is invariant under the following U(1)

transformation:

$$\psi(x) \xrightarrow{U(1)} e^{iq\alpha} \psi(x) \quad (2.3)$$

where  $q$  is the electric charge, and  $\alpha$  is an arbitrary constant.

In light of this symmetry, Noether's theorem implies the existence of a conserved (electromagnetic) current  $j^\mu = -e\bar{\psi}\gamma^\mu\psi$  ( $\partial_\mu j^\mu = 0$ ). If we then allow the  $U(1)$  transformation to be space-time dependent, that is  $\alpha = \alpha(x)$ , we must introduce a new spin-1 vector field  $A^\mu$  in order for the derivative to transform such that the Lagrangian remain invariant. This new field is introduced by making a redefinition of the partial derivative, called the *covariant derivative*, and the following transformation property for the new field:

$$\begin{aligned} \partial_\mu &\rightarrow \partial_\mu - ieA_\mu \\ A_\mu &\xrightarrow{U(1)} A_\mu + \frac{1}{e}\partial_\mu\alpha \end{aligned} \quad (2.4)$$

We then make the substitution into the Lagrangian:

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - e\bar{\psi}\gamma^\mu\psi A_\mu - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (2.5)$$

where  $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$  is the field strength tensor.

We see that in order for the Lagrangian to remain invariant under this  $U(1)$  transformation, we were forced to introduce an additional term which links the conserved electromagnetic current with the spin-1 field:  $e\bar{\psi}\gamma^\mu\psi A_\mu = j^\mu A_\mu$ . This new field is to be identified with the photon, it acts as a mediator of the force between two particles with electric charge. A photon mass term of the form  $\frac{1}{2}m^2 A^\mu A_\mu$  is forbidden as it is not invariant under the transformation rule, the photon remains massless.

## 2.3 Quantum Chromodynamics

Quantum chromodynamics describes the interactions of quarks due to the strong nuclear force. The theory is generated by demanding local invariance of the Lagrangian under an  $SU(3)$

symmetry operating on *color triplets*. Eight new gauge fields must be introduced to give the proper transformation rule of the covariant derivative. These new gauge fields become the gluons and act as the mediator of the strong force, mixing the color states within a quark.

Consider the following SU(3) transformation:

$$q_c \xrightarrow{\text{SU}(3)} e^{\frac{1}{2}ig_s\boldsymbol{\xi}(x)\cdot\boldsymbol{\lambda}} q_c \quad (2.6)$$

where  $g_s$  is the strong coupling constant,  $\boldsymbol{\xi}$  is an 8-dimensional vector of arbitrary functions of spacetime,  $\boldsymbol{\lambda}$  are the 8 3x3 Gell-Mann matrices, and  $\bar{q}_c = (\bar{q}_{red}, \bar{q}_{green}, \bar{q}_{blue})$  is a color-triplet such that  $q(x) = q_c q_D(x)$  and  $q_D(x)$  is a Dirac fermion (any of the 6 SM quarks).

SU(3) gauge invariance requires us to modify the definition of the partial derivative to include 8 spin-1 vector fields  $\mathbf{G}_\mu$ :

$$\begin{aligned} \partial_\mu &\rightarrow \partial_\mu - ig_s \boldsymbol{\lambda} \cdot \mathbf{G}_\mu \\ G_\mu^k &\xrightarrow{\text{SU}(3)} G_\mu^k - \frac{1}{g_s} \partial_\mu \xi_k - f_{ijk} \xi_i G_\mu^j \end{aligned} \quad (2.7)$$

where  $f_{ijk}$  are known as the *structure constants* of SU(3) and arise from its non-abelian nature, they satisfy  $[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k$ .

The complete Lagrangian becomes:

$$\mathcal{L}_{\text{QCD}} = i\bar{q}\gamma^\mu \partial_\mu q - \frac{1}{2}(g_s \bar{q}\gamma^\mu \boldsymbol{\lambda} q) \cdot \mathbf{G}_\mu - m\bar{q}q - \frac{1}{4}\mathbf{G}^{\mu\nu} \cdot \mathbf{G}_{\mu\nu} \quad (2.8)$$

where  $G_i^{\mu\nu} \equiv \partial^\mu G_i^\nu - \partial^\nu G_i^\mu - g_s f_{ijk} G_j^\mu G_k^\nu$  is the field strength tensor for the gluon field  $i$ .

We see the non-abelian nature of SU(3) manifests itself as self-couplings within the gluon field, giving rise to interaction vertices with 3 or 4 gluons. 8 conserved color currents, analogous to the electromagnetic current, are seen as interaction terms between two quarks and a gluon ( $\frac{1}{2}g_s \bar{q}\gamma^\mu \boldsymbol{\lambda} q \cdot \mathbf{G}_\mu = j^\mu \cdot \mathbf{G}_\mu$ ).

$$\mathbf{G}_\mu = \mathbf{j}^\mu \cdot \mathbf{G}_\mu.$$

To conserve color charge at the QCD vertices, gluons themselves must carry both color and anti-color. The color charge allows for self interactions of the gluons (contrast with the to the elec. The 8 physical gluons are members of a color/anticolor octet. 6 of them are expressed as ladder operators within SU(3) and the other two are diagonal matrices.

A gluon mass term of the form  $\frac{1}{2}m^2\mathbf{G}^\mu \cdot \mathbf{G}_\mu$  is forbidden as it is not invariant under the transformation rule, the gluons remain massless.

### 2.3.1 Confinement, Hadrons, & the Quark Model

Experimentally we have never detected a free quark. The physical manifestation of quark production results in a final state consisting of many lower energy particles produced collinear with the original quark. This “spray” of particles is known as a *jet*. This behavior can be explained if one accepts the hypothesis of *color confinement* stating that all objects with color charge are confined to colorless singlets. The origins of color confinement may be traced to the self interactions of the gluon fields resulting in the strong force between two quarks to **increase** as they separate. It eventually becomes energetically favorable to create quark-antiquark pairs from the vacuum which can subsequently combine with other quarks to form the colorless objects (constituting the jets).

*Hadrons* are bound states of quarks which interact through the strong force. Color confinement limits the possible combinations of quarks and antiquarks which can exist together in a bound state (antiquarks carry anticolor). Among states consisting of two quarks+antiquarks, the only possible color singlet wave function is a  $q\bar{q}$  pair:  $\psi = \frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b})$  - these states are called *mesons*. For systems of three quarks+antiquarks, the only color singlet consists of three quarks  $qqq$ :  $\psi = \frac{1}{\sqrt{6}}(rgb - rbg + gbr - grb + brg - bgr)$  - these states are called *baryons*.

The strength of the strong force between two quarks is independent of their flavor and therefore QCD possesses a symmetry under rotations in flavor-space. This would be an exact symmetry if it were not for the physical quark masses (Figure 1.1). To a crude approximation the *u*, *d*, and *s* quarks have the same mass: 2.2, 4.7, and 96 MeV, (compare to the *c* with a mass of 1.28 GeV). The *uds* quarks can therefore be thought of as different states within a triplet. This is known as the *SU(3) flavor symmetry*. Within the symmetry, the  $q\bar{q}$  meson states can be combined to form two nonets with angular momentum  $l = 0$ , and spins of  $s = 0$  or  $s = 1$ . The  $s = 0$  states are the lowest energy and provide a quark description of the kaons and pions. The  $qqq$  baryons can form flavor states consisting of a  $j = 1/2$  octet and  $j = 3/2$  decuplet. The  $j = 1/2$  states are the lowest

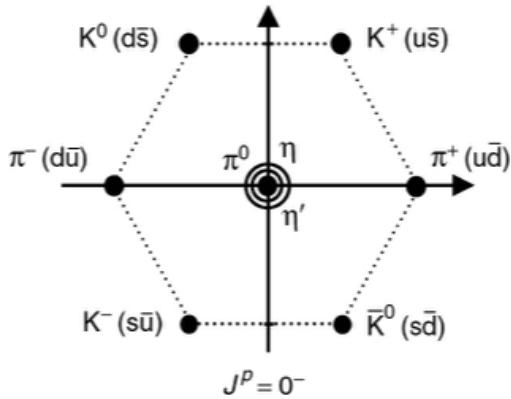
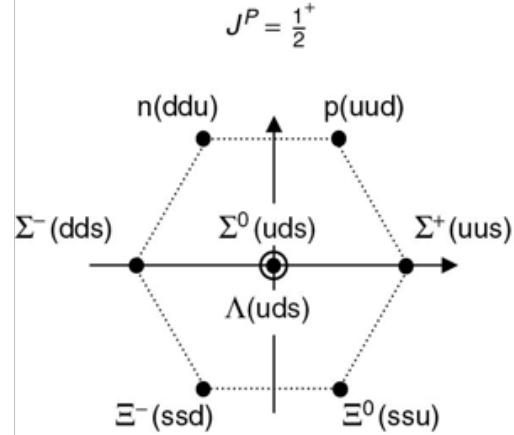
(a) Two quark bound states ( $l = s = 0$ ).(b) Three quark bound states ( $l = 0, s = 1/2$ ).

Figure 2.1: SU(3) flavor symmetry multiplets predicted by the quark model.

energy and describe the proton and neutron. These multiplet structures can be seen in Figure 2.1.

## 2.4 Electroweak Theory

### 2.4.1 Introduction

The electroweak theory provides a unified and self-consistent description of both the electromagnetic and weak forces. The complete theory is generated by demanding local invariance of the Lagrangian under a combined  $SU(2) \times U(1)$  symmetry. The  $SU(2)$  invariance requires the addition of three new vector bosons, two of which are used to construct the physical  $W^\pm$  bosons responsible for the weak *charged current* interactions. An additional gauge boson is required for the  $U(1)$  symmetry. A mixing between the remaining (neutral)  $SU(2)$  gauge field and the  $U(1)$  gauge field yield the  $Z$  boson and photon, responsible for weak *neutral current* and electromagnetic interactions, respectively.

Consider the following  $U(1)$  transformation on a fermion  $\psi$ , and  $SU(2)$  transformation on an

*isospin doublet*  $\Psi$ :

$$\begin{aligned}\psi(x) &\xrightarrow{\text{U}(1)} e^{ig'\frac{Y}{2}\alpha(x)} \psi(x) \\ \Psi(x) &\xrightarrow{\text{SU}(2)} e^{ig_W \xi(x) \cdot \frac{1}{2}\boldsymbol{\sigma}} \Psi(x)\end{aligned}\tag{2.9}$$

where  $g'$  is the hypercharge coupling constant,  $Y$  is the hypercharge operator,  $g_W$  is the weak coupling constant,  $\alpha$  and  $\xi$  are arbitrary functions of spacetime, and  $\boldsymbol{\sigma}$  represents the 3 2x2 Pauli spin matrices.

As usual, SU(2) and U(1) gauge-invariance requires us to modify the definition of the partial derivative to include three spin-1 vector fields  $\mathbf{W}^\mu$  and a single spin-1 vector field  $B^\mu$ :

$$\begin{aligned}\partial_\mu &\rightarrow \partial_\mu - ig'\frac{Y}{2}B_\mu - ig_W\frac{1}{2}\boldsymbol{\sigma} \cdot \mathbf{W}_\mu \\ B_\mu &\xrightarrow{\text{U}(1)} B_\mu - ig'\partial_\mu\alpha \\ \mathbf{W}_\mu^k &\xrightarrow{\text{SU}(2)} \mathbf{W}_\mu^k - g_W\partial_\mu\xi^k - g_W\epsilon_{ijk}\xi^i\mathbf{W}_\mu^j\end{aligned}\tag{2.10}$$

where  $\epsilon_{ijk}$  is the totally antisymmetric Levi-Civita tensor (the structure constants of SU(2)).

The charged current interaction connects two elements within an isospin doublet  $\Psi$ ; by convention, the upper element has electric charge +1 relative to the lower element. There are doublets which connect the leptons with a corresponding neutrino (massless spin-1/2 particles), and there are doublets which connect an 'up-type' quark (top entry of the doublet) to a 'down-type' quark (bottom entry):

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix}$$

The  $W^1$  and  $W^2$  gauge fields correspond to the first two Pauli matrices; appropriate linear combinations of these two fields therefore define raising and lowering operators which transform elements within a doublet. The physical  $W^\pm$  bosons are the following linear combinations of the two gauge fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)\tag{2.11}$$

As Nature has it, the weak force is a *chiral* theory which does not treat the left and right-chiral components of a Dirac fermion on equal footings. The projection operator  $P_{R/L} = \frac{1}{2}(1 \pm \gamma^5)$  is used to define these left and right chiral states, all Dirac fermions can be decomposed as  $\psi = \psi_L + \psi_R$  using these operators. In the Standard Model, only left-handed particle and right-handed antiparticle states enter into the isospin doublets participating in the electrically-charged weak interaction. This is summarized in Table 2.1.

Because of the SU(2) symmetry and doublet nature, we must introduce **two** fermions to the theory, where the left and right chiral components may transform differently under gauge interactions. Consider fields  $\chi$  and  $\tau$ ; the left handed components are members of an isospin doublet  $\overline{\psi_L} = (\overline{\chi_L}, \overline{\tau_L})$ , all components participate in the U(1) transformation. The complete Lagrangian becomes:

$$\begin{aligned}\mathcal{L}_{\text{EWK}} = & i\bar{\chi}\gamma^\mu\partial_\mu\chi - m\bar{\chi}\chi + i\bar{\tau}\gamma^\mu\partial_\mu\tau - m\bar{\tau}\tau - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\mathbf{W}_{\mu\nu}\cdot\mathbf{W}^{\mu\nu} \\ & - g'\bar{\chi}\gamma^\mu\frac{Y}{2}\chi B_\mu + g'\bar{\tau}\gamma^\mu\frac{Y}{2}\tau B_\mu - g_W\overline{\psi_L}\gamma^\mu\boldsymbol{\sigma}\psi_L\cdot\mathbf{W}_\mu\end{aligned}\quad (2.12)$$

Mass terms of the form  $\frac{1}{2}m^2A^\mu A_\mu$  are forbidden as they are not invariant under the transformation rule, the  $B$  and  $\mathbf{W}$  bosons remain massless.

#### 2.4.2 The Higgs Mechanism

The gauge bosons responsible for the electroweak force have observationally been determined to have mass, which is a problem as the gauge symmetry requires them to be massless. Surely there must be some mechanism which can be introduced to achieve this meanwhile (at least initially) preserving the  $SU(2)\times U(1)$  symmetry. The *Higgs field* is introduced to generate mass terms for the electroweak gauge bosons. The discovery of the Higgs boson was the final particle contained within the SM to be observed, its discovery in 2012 was monumental.[2]

This Higgs mechanism proceeds by introducing a massive spin-0 complex scalar field with the following Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi^\dagger)(\partial^\mu\phi) - \frac{1}{2}\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2\quad (2.13)$$

where  $\mu$  and  $\lambda$  are the strengths of the self-coupling terms.

Within the SM, the field is implemented as an isospin doublet which consists of electrically neutral and charged components:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (2.14)$$

Solving for the minimum of the potential, it is found that the ground state of  $\phi$  is non-zero and satisfies  $\phi^\dagger \phi = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = \frac{1}{2}v^2 = -\mu^2/2\lambda$ . This is called *spontaneous symmetry breaking* - the Higgs acquiring a *vacuum expectation value*. Perturbation theory of interactions represent particles as fluctuations above the vacuum - we must express the fields in the same manner. Electric charge conservation requires that this vacuum expectation value lie entirely inside the neutral  $\phi^0$ . This ground state is then expressed as:

$$\phi = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (2.15)$$

where  $h(x)$  is to be identified as the Higgs boson.

If we substitute the ground-state expansion of  $\phi$  to the Lagrangian of Equation 2.13 we obtain the following expression:

$$\mathcal{L} = \frac{1}{2}(\partial^\mu h)(\partial_\mu h) - \lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 + \lambda v^4 \quad (2.16)$$

where we see have generated a mass term  $m_h = \sqrt{2\lambda}v$  for the Higgs boson, additionally there are now 3 and 4-point Higgs self-couplings.

#### 2.4.2.1 Masses of the $W^\pm$ and $Z$ bosons

The kinetic energy term  $\frac{1}{2}(\partial_\mu \phi^\dagger)(\partial^\mu \phi)$  for the Higgs field introduces a coupling with the  $\mathbf{W}^\mu$  and  $\mathbf{B}^\mu$  bosons when they are added to the covariant derivative:

$$\partial_\mu \phi = \left( \frac{1}{2}\partial_\mu + \frac{1}{2}ig_W \boldsymbol{\sigma} \cdot \mathbf{W} + ig' \frac{Y}{2} \mathbf{B}^\mu \right) \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.17)$$

After performing the matrix calculations, and lots of algebra, there are terms quadratic in the gauge fields:

$$\frac{1}{8}v^2g_W^2(W_\mu^1W_1^\mu + W_\mu^2W_2^\mu) + \frac{1}{8}v^2(g_WW_\mu^3 - g'B_\mu)(g_WW_3^\mu - g'B^\mu) \quad (2.18)$$

Where we see we have generated mass terms for the  $W_1^\mu$  and  $W_2^\mu$  fields:  $m_W = \frac{1}{2}vg_W$ . The last term in the expansion introduces mixed couplings between the electrically neutral and massless  $W_3^\mu$  and  $B^\mu$  fields. The mixing can be represented via a non-diagonal mass matrix. Physical particles propagate as independent eigenstates of the free particle Hamiltonian and therefore we must find the basis in which this matrix is diagonal. Upon diagonalization, we find the states corresponding to these eigenvalues:

$$\begin{aligned} A_\mu &= \frac{1}{\sqrt{g_W^2 + g'^2}} (g'W_\mu^3 + g_WB_\mu); & \text{with mass } 0 \\ Z_\mu &= \frac{1}{\sqrt{g_W^2 + g'^2}} (g_WW_\mu^3 - g'B_\mu); & \text{with mass } \frac{1}{2}v\sqrt{g_W^2 + g'^2} \end{aligned} \quad (2.19)$$

where  $A_\mu$  corresponds to the photon of electromagnetism, and  $Z_\mu$  the neutral gauge boson responsible for the weak neutral currents.

We have seen how the Higgs mechanism is able to generate mass terms for the gauge bosons in the electroweak theory.

#### 2.4.2.2 Masses of the Fermions

It has not been mentioned that the fermion mass terms  $-m\bar{\psi}\psi = -m(\overline{\psi_R}\psi_L + \overline{\psi_L}\psi_R)$  are in fact forbidden within the SM - the chiral nature of SU(2) treats the two chiral states differently and therefore this term is not invariant under the transformation rules. Masses for the fermions are created by introducing *Yukawa interactions* with the Higgs field. Consider the following terms which are invariant under the  $U(1) \times SU(2)$  transformation:

$$\begin{aligned} \mathcal{L} &= -y[\overline{\psi_L}\phi\psi_R + \overline{\psi_R}\phi^\dagger\psi_L] \\ &= -y \left[ \begin{pmatrix} \bar{\nu}, \bar{\ell} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \ell_R + \overline{\ell_R} \begin{pmatrix} \phi^{+*}, \phi^{0*} \end{pmatrix} \begin{pmatrix} \nu \\ \ell \end{pmatrix}_L \right] \end{aligned} \quad (2.20)$$

where  $y$  is the *Yukawa coupling*,  $\overline{\psi}_L$  is an isospin doublet of left-chiral fermions, and  $\psi_R$  is a right-chiral fermion.

After spontaneous symmetry breaking, this reduces to:

$$\begin{aligned}\mathcal{L} &= -\frac{1}{\sqrt{2}}yv(\overline{\ell}_L\ell_R + \overline{\ell}_R\ell_L) - \frac{1}{\sqrt{2}}yh(\overline{\ell}_L\ell_R + \overline{\ell}_R\ell_L) \\ &= -\frac{1}{\sqrt{2}}yv\bar{\ell}\ell - \frac{1}{\sqrt{2}}yh\bar{\ell}\ell\end{aligned}\tag{2.21}$$

and we see we have obtained a mass term for the fermion  $m_\ell = \frac{1}{\sqrt{2}}yv$  and an interaction term  $\frac{1}{\sqrt{2}}yh\bar{\ell}\ell$  between the fermion in the lower member of the isospin doublet and a single Higgs boson.

To generate a mass term for the upper component of the isospin doublet we need to follow the same prescription but with the *conjugate* Higgs field:

$$\phi_c = -i\sigma_2\phi^* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -\phi_3 + i\phi_4 \\ \phi_1 - i\phi_2 \end{pmatrix},\tag{2.22}$$

where the same story plays out.

#### 2.4.3 Quark Mixing - the CKM Matrix

The charged weak current acts on isospin doublets connecting quarks of different flavors - the flavor eigenstates: u, d, s, c, b, t. The quarks propagate as mass eigenstates of the free-particle Hamiltonian. This introduces mixing between these two bases and is represented by the *CKM matrix*. The probability of a transition between two states is proportional to a matrix element  $|V|^2$  in the matrix. It is unitary 3x3 matrix with 4 degrees of freedom:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}\tag{2.23}$$

The best estimate of these parameters are:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \quad (2.24)$$

#### 2.4.4 Brief Electroweak Summary

We started with a theory of two fields each governed by the Dirac equation. We demanded that the theory be gauge-invariant under combined  $U(1) \times SU(2)$  symmetry operations. The gauge bosons are required to be massless as their transformation rules do not allow their kinetic energy term to be invariant. The Higgs field was introduced, the action of the covariant derivative on the Higgs field generates an interaction term between the gauge bosons and the Higgs field. The Higgs field obtained a vacuum expectation value, re-expressing the field about this ground state led us to mass terms for the gauge bosons. The physical  $W^\pm$ ,  $Z$ , and  $A$  bosons become mixtures of these states. Fermion mass terms are not initially allowed as the chiral  $SU(2)$  symmetry treats the left and right-chiral components differently and therefore can not remain invariant. Fermion mass terms are generated by introducing Yukawa interactions between the Higgs field and fermions which generate appropriate mass terms after the Higgs field expansion. In addition, under the charged-current interaction, quark flavor and mass eigenstates are not the same, this introduces the CKM matrix parametrizing the mixing.

### 2.5 Parameters of the Standard Model

There are 18 parameters that must be specified as an input to the SM - these must all be determined experimentally:

- 9 quark and lepton masses:  $m_u, m_d, m_s, m_c, m_b, m_t, m_e, m_\mu, m_\tau$ ; the values are listed in Figure 1.1. (Alternatively these can be expressed as the 9 appropriate Yukawa couplings  $y_f = \sqrt{2}m_f/v$ : 1, 1, 1, 1, 1, 1, 1, 1, 1 respectively)

- 4 parameters describing the mixing between quark mass and flavor eigenstates (CKM matrix, see Equations 2.23 and 2.24): often parametrized as  $\lambda, A, \rho, \eta$
- 2 parameters for the Higgs field: mass and vacuum expectation value  $m_H = 125$  GeV,  $v = 246$  GeV
- 3 coupling constants for the relative strengths of the gauge group:

$$\begin{aligned}\alpha &\equiv e^2/4\pi, & \alpha(q^2 \approx 0) &= 1/137.0 \\ && \alpha(q^2 = (193 \text{ GeV})^2) &= 1/(127.4 \pm 2.1) \\ \alpha_s &\equiv g_s^2/4\pi, & \alpha_s(q^2 = m_Z^2 = (91 \text{ GeV})^2) &= 0.1184 \pm 0.0007 \\ G_F &\equiv \sqrt{2}g_W^2/8m_W^2, & G_F(q^2 \approx 0) &= 1.1663787 \times 10^{-5}\end{aligned}$$

## 2.6 Neutrino Mass

Within the 'canonical' SM the neutrinos are assumed to be massless spin-1/2 fermions. In the last decade, experiments have shown that neutrinos go through *flavor oscillations* wherein they change flavor as they propagate through space. This flavor oscillation is dependent on **massive** neutrinos and can be described by a mixing between these mass and flavor eigenstates. The mixing is described by a 3x3 unitary matrix known as the *PMNS matrix* (analogous to the CKM matrix). One could consider giving the neutrinos a Yukawa coupling with the Higgs and give them a Dirac mass  $-m\bar{\nu}\nu = -m(\bar{\nu}_R\nu_L + \bar{m}\bar{\nu}_L\nu_R)$ , but the non-observation of right-handed neutrinos does not allow that approach. The correct implementation of the neutrino mass is still an open question in the field. The currently known estimates for their masses are seen in Figure 1.1.

The neutrino mass sector possibly adds an additional 7 parameters to the SM:

- 3 neutrino masses:  $m_{\nu_1}, m_{\nu_2}, m_{\nu_3}$
- 4 parameters describing the mixing between the neutrino mass and flavor eigenstates (PMNS matrix): often parameterized as  $\theta_{12}, \theta_{13}, \theta_{23}, \delta$ .

The best fit values of the PMNS matrix are:

$$\begin{pmatrix} |U_{e1}| & |U_{e2}| & |U_{e3}| \\ |U_{\mu 1}| & |U_{\mu 2}| & |U_{\mu 3}| \\ |U_{\tau 1}| & |U_{\tau 2}| & |U_{\tau 3}| \end{pmatrix} \approx \begin{pmatrix} 0.85 & 0.50 & 0.17 \\ 0.35 & 0.60 & 0.70 \\ 0.009 & 0.35 & 0.70 \end{pmatrix} \quad (2.25)$$

## Chapter 3

### The Minimal Supersymmetric Model

In addition to the  $SU(3) \times SU(2) \times U(1)$  symmetry imposed on the Lagrangian within the Standard Model, one can introduce additional *supersymmetries* to the theory. Supersymmetry transformations act on fields containing both fermionic and bosonic degrees of freedom, the Lagrangian is required to remain invariant under rotations between these two states. Many such extensions to the Standard Model exist, but only the simplest of these, requiring a single global transformation  $Q$  is phenomenologically viable: the Minimal Supersymmetric Standard Model (MSSM).

Within the MSSM, every SM particle has a *superpartner* which differs by  $1/2$  unit of spin but otherwise has identical properties. For example, the electron is paired with an electrically charged, massive, spin-0 scalar field, called a “selectron”. A striking prediction of the MSSM is more than a doubling of the known particle content of the Universe. In addition to the single SM Higgs doublet, the MSSM requires another spin-0 Higgs isospin doublet, with opposite hypercharge. These two spin-0 doublets each have their own chiral superpartner. Table 3.1 lists the fields required within the MSSM (the sort of complement of Table 2.1). As the nature of the newly supersymmetric electroweak sector is unknown, the various components of the fields may mix. The charged components from the new  $SU(2)$  fields and the Higgs summarized in Table 3.2 (color charge prevents gluinos from mixing outside QCD). This chapter is adapted from [3].

Matter particles (both the left and right-chiral components separately) are placed in *chiral supermultiplets* consisting of a spin- $1/2$  Majorana fermion  $\psi$  and complex scalar field  $\phi$ . In the massless and non-interacting case (i.e. just the two kinetic terms in the Lagrangian, known as

Table 3.1: Summary of the additional particle content within the MSSM.

<b>SM Gauge Sector</b>	sfermions	gauginos
SU(3)	$\tilde{u}, \tilde{d}, \tilde{c}, \tilde{s}, \tilde{t}, \tilde{b}$	$\tilde{g}_{1\dots 8}$
SU(2) $\times$ U(1)	$\tilde{Q}_L = \begin{pmatrix} \tilde{u} \\ \tilde{d} \end{pmatrix}_L, \begin{pmatrix} \tilde{s} \\ \tilde{c} \end{pmatrix}_L, \begin{pmatrix} \tilde{t} \\ \tilde{b} \end{pmatrix}_L$ $\tilde{q} = \tilde{u}, \tilde{d}, \tilde{c}, \tilde{s}, \tilde{t}, \tilde{b}$ $\tilde{L}_L = \begin{pmatrix} \tilde{\nu}_e \\ \tilde{e}^- \end{pmatrix}_L, \begin{pmatrix} \tilde{\nu}_\mu \\ \tilde{\mu}^- \end{pmatrix}_L, \begin{pmatrix} \tilde{\nu}_\tau \\ \tilde{\tau}^- \end{pmatrix}_L$ $\tilde{\ell} = (\tilde{e}, \tilde{\mu}, \tilde{\tau})$	$\widetilde{W}_{012}, \widetilde{B}_0$
<b>Higgs Sector</b>	spin-0 $H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix}, H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix}$	spin-1/2 $\tilde{H}_u = \begin{pmatrix} \tilde{H}_u^+ \\ \tilde{H}_u^0 \end{pmatrix}, \tilde{H}_d = \begin{pmatrix} \tilde{H}_d^0 \\ \tilde{H}_d^- \end{pmatrix}$

Table 3.2: Mixing of the supersymmetric electroweak fields.

name	gauge states	mass eigenstates
neutralinos	$\tilde{H}_u^0 \leftrightarrow \tilde{H}_d^0 \leftrightarrow \widetilde{W}^0 \leftrightarrow \widetilde{B}^0$	$\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$
charginos	$\tilde{H}_u^+ \leftrightarrow \tilde{H}_d^- \leftrightarrow \widetilde{W}^+ \leftrightarrow \widetilde{W}^-$	$\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$
neutral Higgs	$H_u^0 \leftrightarrow H_d^0$	$h^0, H^0, A^0$
charged Higgs	$H_u^+ \leftrightarrow H_d^-$	$H^\pm$

the *Wess-Zumino model* [4]), the transformation laws of the fields can be deduced by demanding invariance under the simple Lagrangian, they are found to be:

$$\begin{aligned}\phi &\xrightarrow{\text{Q}} \phi + \epsilon\psi \\ \psi &\xrightarrow{\text{Q}} \psi - i(\sigma^\mu\epsilon^\dagger)\partial_\mu\phi\end{aligned}\tag{3.1}$$

where  $\epsilon$  is a 2-component Weyl **spinor** parametrizing the transformation. For the duration of this chapter, all references to *auxiliary* fields will be omitted. Auxiliary fields are internal to the theory and must be introduced to allow the fields to satisfy their classical wave equations.

The requirement of renormalizability restricts the numbers of fields in any interaction involving  $\psi$  and  $\phi$ , the most generic Lagrangian for a chiral supermultiplet is of the form:

$$\mathcal{L}_{chiral} = -D^\mu\phi^*D\mu\phi - V(\phi, \phi^*) + i\psi^\dagger\bar{\sigma}^\mu D\mu\psi - \frac{1}{2}(M\psi\psi + h.c.) - \frac{1}{2}(y\phi\psi\psi + h.c.)\tag{3.2}$$

where  $D^\mu$  is the covariant derivative,  $V(\phi, \phi^*)$  is a scalar potential for the theory,  $\bar{\sigma}^0$  is the 2x2 identity matrix and  $\bar{\sigma}^{123} \equiv -\sigma^{123}$ ,  $M$  is a (Majorana) mass term,  $\psi\psi \equiv \epsilon^{ab}\psi_a\psi_b$ , and  $y$  is a Yukawa coupling. The Yukawa coupling connects two SM fermions with the corresponding supersymmetric scalar field - a vertex diagram for this process is seen in Figure 3.1(a). The covariant derivative  $\partial_\mu\phi \rightarrow \partial_\mu\phi - igA_\mu^aT^a\phi$ , when introduced in the kinetic term, creates two interactions between the SM gauge bosons and the new supersymmetric scalar field:  $-ig[(\partial_\mu\phi)A_\mu^aT^a\phi + h.c.]$  and  $g^2A_\mu^{a\mu}\phi^*t^aA_\mu^aT^a\phi$ , seen in Figures 3.1d,e, respectively. Throughout this chapter I will only discuss interactions involving at least one SM particle and at least one supersymmetric particle.

Gauge bosons (before spontaneous symmetry breaking) are placed in *gauge supermultiplets* consisting of gauge bosons  $A_\mu^a$  and spin-1/2 gauginos  $\lambda^a$ ;  $a$  is a label which runs over the SM gauge fields within the theory. Under the supersymmetry, fields can be found to transform as:

$$\begin{aligned}A_\mu^a &\xrightarrow{\text{Q}} A_\mu^a - \frac{1}{\sqrt{2}}(\epsilon^\dagger\bar{\sigma}_\mu\lambda + h.c.) \\ \lambda_\alpha^a &\xrightarrow{\text{Q}} \lambda_\alpha^a + \frac{i}{2\sqrt{2}}(\sigma^\mu\bar{\sigma}^\nu\epsilon)F_{\mu\nu}^a\end{aligned}\tag{3.3}$$

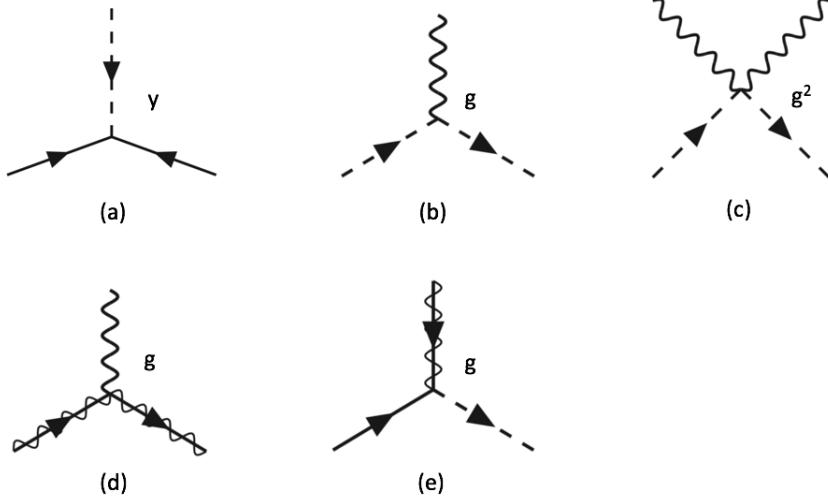


Figure 3.1: Interactions between SM and MSSM particles.  $y$  is a Yukawa coupling to be determined,  $g$  is the SM gauge coupling.

where  $F_{\mu\nu}^a$  is the regular field strength tensor for the gauge field  $A_\mu^a$ . The SM symmetries (i.e. SU(3), SU(2), U(1)) transform the gauge supermultiplet in the following way:

$$\begin{aligned} A_\mu^a &\xrightarrow{\text{SM}} A_\mu^a + \partial_\mu \Lambda^a + g f^{ijk} A_\mu^b \Lambda^c \\ \lambda^a &\xrightarrow{\text{SM}} \lambda^a + g f^{abc} \lambda^b \Lambda^c \end{aligned} \quad (3.4)$$

where  $\Lambda^a$  is a parameter describing the transformation. (The transformation law for  $A^\mu$  is the same we have seen in Chapter 2.)

The Lagrangian for a free gauge multiplet consists simply of the kinetic terms for each field  $A_\mu^a$  and  $\lambda^a$ :

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \lambda^\dagger \bar{\sigma}^\mu \nabla^\lambda \lambda^a \quad (3.5)$$

where  $f^{abc}$  are the structure constants of the gauge group.  $\nabla \lambda^a = \partial_\mu \lambda^a + g f^{abc} A_\mu^b \lambda^c$  represents the covariant derivative acting on  $\lambda^a$  - creating an interaction term between a gauge boson and two gauginos:  $i g \lambda^\dagger \bar{\sigma}^\mu f^{abc} A_\mu^b \lambda^c$ , as seen in Figure 3.1c.

The requirement of renormalization restricts the interactions between the gauge and chiral supermultiplets to be only of the form  $-\sqrt{2}g(\phi^* T^a \psi \lambda^a + h.c.)$ , involving a single spin-0, spin-1/2, and spin-1 particle - a vertex diagram is seen in Figure 3.1e.

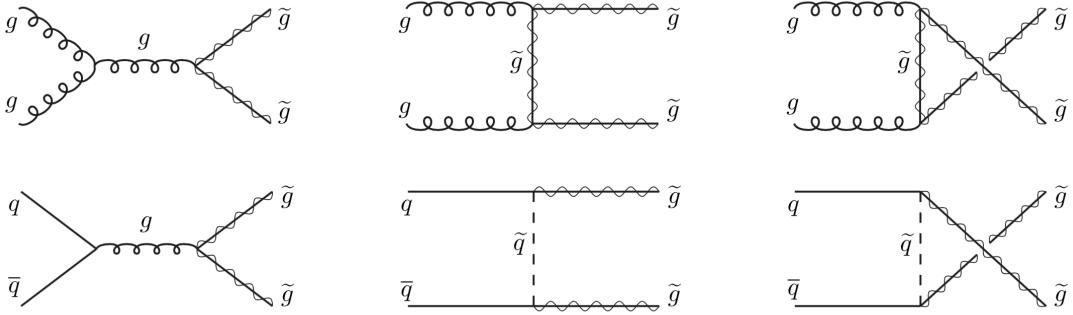


Figure 3.2: Tree-level gluino pair-production mechanisms.

Within a supersymmetric theory, the superpartners are required to have the same mass as their corresponding SM field. If SUSY were exact, we would have expected to see evidence of sparticles over the years. There must be some mechanism which generates large mass for the sparticles such that their production is highly suppressed at our colliders. The remainder of this thesis presents a search for evidence of physics beyond the SM, such as the MSSM. Our motivation is taken from final state topologies arising from gluino pair production - possible mechanisms seen in Figure 3.2. The gluino is the spin-1/2 fermion which is the superpartner to the gluon, the mediator of the strong force.

Within the context of QCD - the chiral supermultiplets consist of spin-1/2 quarks and spin-0 squarks. Searches for supersymmetric particles of QCD are partly motivated by the fact that most of their production mechanisms proceed through diagrams proportional to the strong coupling constant  $g_s$ , which is largest among the three in the SM. The supersymmetric partners of QCD necessarily carry color charge, and therefore the QCD squarks and gluinos do not directly interact with other MSSM particles. There have been many searches for SUSY which provide lower limits on the mass of the gluino, such as [5, 6]. An example is seen in Figure 3.3, setting a lower bound at about 1.8 TeV, given assumptions about the decay (seen in the diagram below the plot). The blob in the diagrams below the limit plots indicates we are not interested in the particular production mechanism, but in the decay chain.

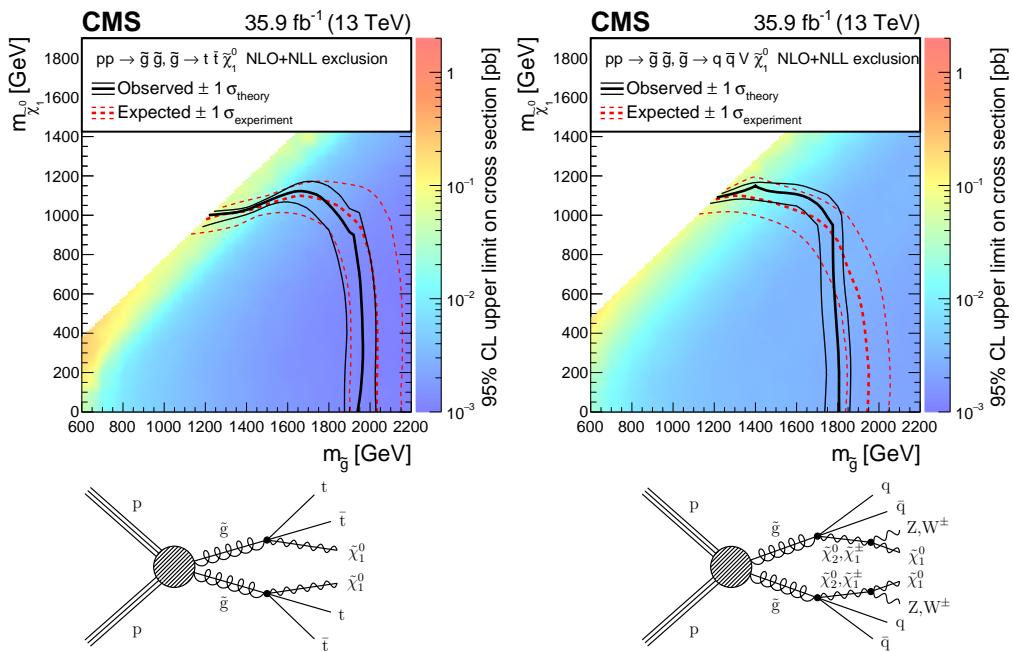


Figure 3.3: Previous results for searches of gluino-mediated supersymmetry for two the T1tttt (left) and T1qqqqVV models [5]. The excluded regions are towards the lower left.

## Chapter 4

### The Large Hadron Collider

The Large Hadron Collider (LHC) is a particle accelerator capable of creating the most energetic (man-made) collisions of particles to date. The LHC is housed within a tunnel 27 km in circumference and 100 m underground, at the European Organization for Nuclear Research (CERN) site near Geneva, Switzerland. Two beampipes, 5.6 cm in diameter, contain protons circulating in opposite directions around the LHC tunnel. Over 1200 superconducting dipole magnets, 15 m in length and providing a field strength of 8.3 T, are placed along the beamline to guide the protons within the circular trajectory. Radio-frequency electric fields are used to accelerate the particles to nearly the speed of light.

Before the protons are stored in the LHC ring, they must first make their way through a number of stages which comprise the CERN accelerator complex. The protons are sourced from a simple bottle of hydrogen gas. A large electric field is used to ionize the gas and the protons are fed into a linear accelerator (Linac 2) which increases their energy to 50 MeV. These protons subsequently are fed through three synchrotrons: Proton Synchrotron Booster, Proton Synchrotron, Super Proton Synchrotron, where the proton beam energy is increased to 1.5, 25, and 450 GeV, respectively. After the Super Proton Synchrotron, the beamlines of the LHC are filled resulting in two counter-propagating beams of 6.5 TeV each. Figure 4.1 is a diagram of the entire CERN complex. As can be seen, there accelerator complex is rich with activity. Table 4.1 is a summary of the successive stages relevant in the beam development.

At four points around the ring, magnets are used to further confine and direct each of the

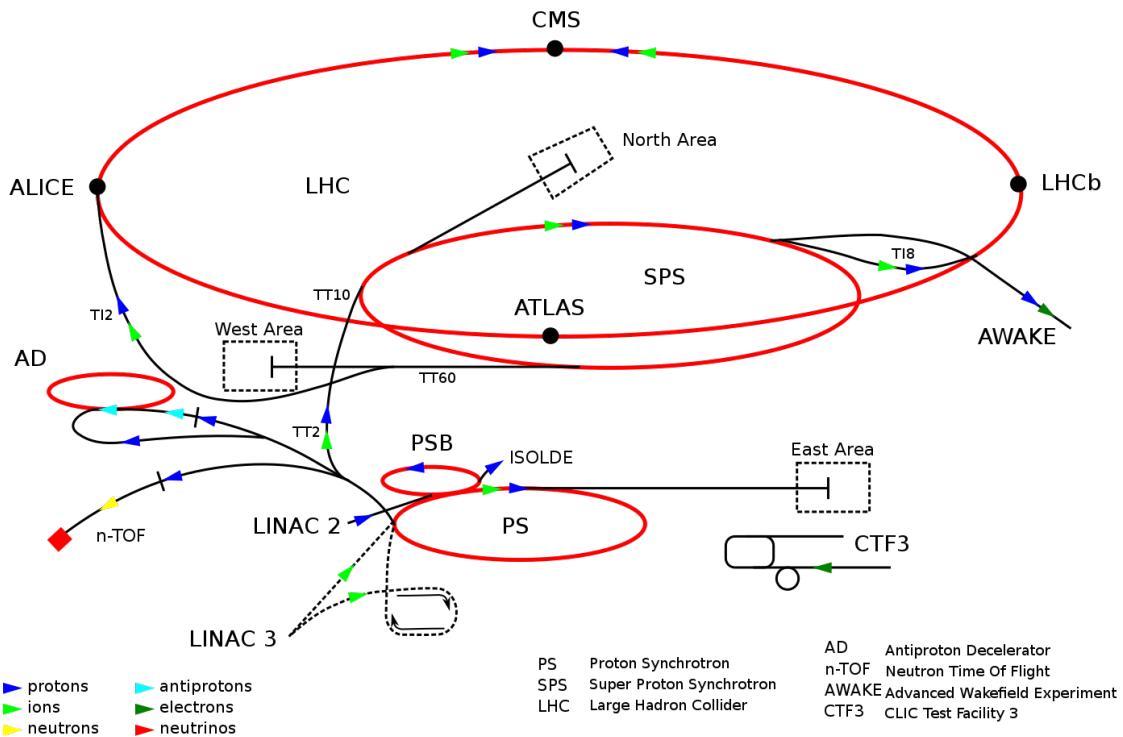


Figure 4.1: The CERN accelerator complex.

Table 4.1: The stages of proton beam acceleration for the LHC at CERN.

stage	final energy
bottle of hydrogen gas	...
Linac 2	50 MeV
Proton Synchrotron Booster	1.5 GeV
Proton Synchrotron	25 GeV
Super Proton Synchrotron	450 GeV
LHC	6.5 TeV

two (counter-rotating) beams at one another. At each of these interaction points, there is a large detector placed to capture the remnants of the collisions. These detectors are named CMS, LHCb, ATLAS, and ALICE, and labeled as such in Figure 4.1. CMS and ATLAS are considered “general-purpose” detectors striving to surround the interaction region as much as possible, allowing to reconstruct a wide variety of particles and the full event. The LHCb detector instruments the forward region that is specialized in reconstruction events containing b hadrons. The ALICE detector is specifically designed to record events from heavy-ion collisions.

A measure of the rate of particle collisions produced at a collider is given by the *instantaneous luminosity*. It relates the cross section probability ( $\sigma$ ) for some interaction to occur with the expected numbers of events  $N$  of those type over some time period:  $N = \sigma \int \mathcal{L}(t)$ . It is an important parameter of an accelerator as it dictates how many proton interactions can be made within a given amount of time, limiting the total amount of data. In general, the luminosity is dependent on the size and shape of each of the beams, the number of protons in each beam  $n_{1,2}$ , and how frequently they can be made to interact at the LHC. The instantaneous luminosity can be defined as:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \quad (4.1)$$

where  $f$  is the collision frequency, and  $\sigma_x, \sigma_y$  are the effective transverse widths of the beams.

In the LHC, the protons within each beam are arranged in 2808 “bunches” about 30cm long containing  $10^{11}$  protons each ( $n_1, n_2$ ) (arranged in this manner to the radiofrequency chambers). At each of the interaction points, the bunches are steered into one another every 25 ns ( $f = 40$  MHz). Although there are many protons within each bunch, on average there are only about 25 appreciable proton interactions per crossing with a large energy transfer between two partons (constituents of a proton, i.e. quarks and gluons). Each of these interactions creates its own *primary vertex* where many tracks are found to emanate from, indicating a single interaction between two particular protons. The number of additional interactions per bunch crossing is known as *pile up* and poses a formidable challenge in the reconstruction. One of the long term goals for the LHC is to increase

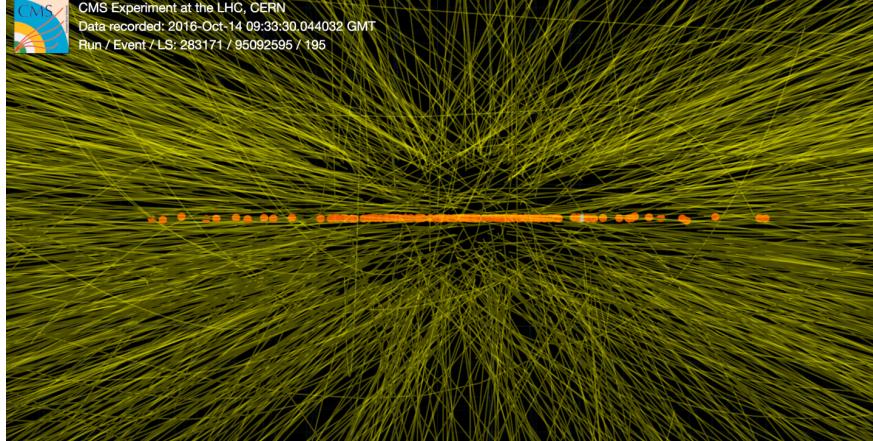


Figure 4.2: The number of pileup interactions in a 2016 event recorded by CMS [7]. The green lines are tracks reconstructed in the silicon tracker. The orange dots are identified interaction vertices.

its instantaneous luminosity leading to a drastic increase in the number of these vertices which must be reconstructed. In Figure 4.2 we see the challenge we already face - an image of an event recorded in 2016, the green lines are tracks, the orange dots are interaction vertices.

The LHC began taking data in 2009 with a total center-of-mass energy of 900 GeV. The center-of-mass energy of the beam collisions has been increasing over the years, with runs at 7, 8, and now 13 TeV. In coming years the machine is planned to deliver 14 TeV collisions, the design energy. The high performance of the LHC machine has allowed us to take over  $150 \text{ fb}^{-1}$  of data to date, the record for hadronic machines to date. Figure 4.3 shows the integrated luminosity collected by the CMS experiment over this time.

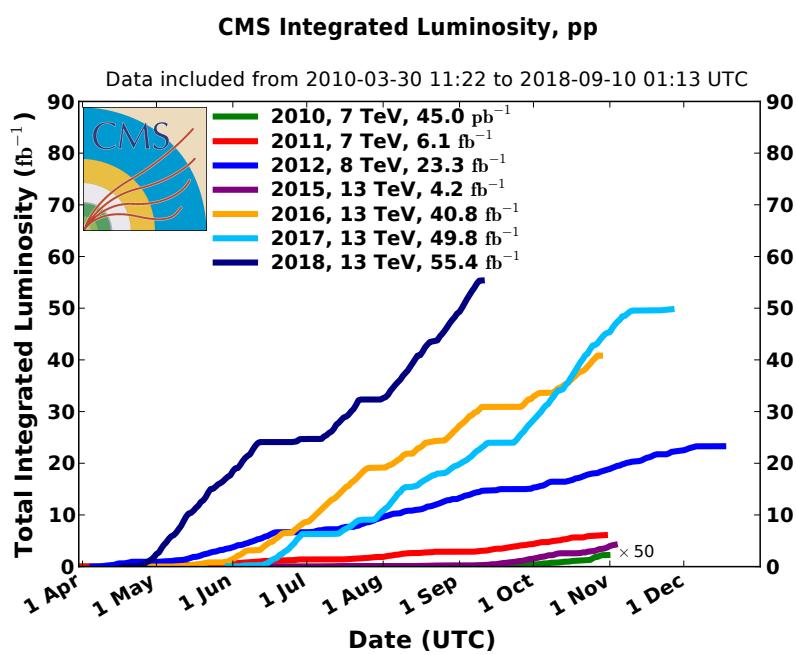


Figure 4.3: The integrated luminosity of CMS over its lifetime.

## Chapter 5

### The CMS Detector

The interactions of the protons more often than not are just glancing blows off one another, there is not much energy transfer and the protons generally remain intact. The CMS experiment is predominantly interested in the dynamics of collisions with a very large energy transfer, unveiling quarks and gluons that are within the proton. Large energy transfers are necessary for the creation of massive particles such as the electroweak bosons, or potentially a new particle not within the SM. These high-energy collisions result in particles with a large amount of momentum in the direction perpendicular to the beam line ( $p_T$ ), where the components of the CMS detector are carefully arranged. By detecting the outgoing flux of particles from these high-energy events, we are able to reconstruct the dynamics and quantum mechanical processes involved in the proton interactions.

The detector is composed of a modular design of subsystems allowing for measurements of a wide spectra of particles. As seen in Figure 5.1, these detectors are placed around the interaction point to collect as much of the collision remnants as possible. In this figure, the beam line is seen as a small grey tube extending from the bottom right towards the top left. The interaction point is within the silicon trackers.

Each different system is capable of detecting specific types of particles. A silicon tracker allows for the reconstruction of charged particles traveling at least 50 cm (e.g. electrons  $e^\pm$ , charged pions  $\pi^\pm$ ); calorimeters measure the energy of hadrons (e.g. protons,  $K_L^0$ ) and photons  $\gamma$ ; muon  $\mu^\pm$  identification is made with gas detectors. The particle identification exploits the unique signature each of these particles leave in our detector. Figure 5.2 illustrates these signatures for some common

SM particles.

Many of the particles produced in the interaction are unstable and decay before traveling any appreciable distance in the detector. These particles, such as top quarks, electroweak bosons, and many hadrons, must be reconstructed by the identification of their decay products. For instance, a Z boson may be reconstructed as a pair of oppositely charged muons  $\mu^\pm$ .

This chapter will discuss the main elements of the CMS detector, beginning with the innermost (closest to the beam pipe) silicon tracker and concluding with the muon system. The detector can generically be divided into central *barrel* and forward *endcap* regions. The geometry either takes the form of concentric cylinders (in the barrel) or flat planes of detectors (in the endcaps). This is most apparent in Figure 5.3, where the beamline is seen as the thin cyan line at the bottom of the image. The origin is defined as the center of the detector. The radial coordinate  $r$  is the distance from the beam line, in the transverse plane. The z axis is the direction parallel to the beamline, counterclockwise when viewed from above. The coordinate  $\eta \equiv -\tan \ln \theta/2$  is used to represent the polar angle above the beamline. The  $\phi$  coordinate is the azimuth angle about the beamline, with  $\phi = 0$  aligning with the x-axis pointing towards the center of the LHC ring. The y-axis points up.

## 5.1 Silicon Tracker

The silicon tracker is responsible for the reconstruction of charged particles, which are mainly electrons, muons, kaons and pions. The particle trajectory is reconstructed using ionization deposits left in layers of thin silicon. The particle momentum is measured by the curvature of the trajectory when immersed in the magnetic field. The silicon tracker is divided into two major components. The pixel detector is at a closer proximity to the beam line and has finer spatial segmentation. The strips detector covers a much larger spatial volume and is responsible for the majority of the hits along a particle trajectory. [8, 9]. A diagram of the geometry and layers of the tracker is seen in Figure 5.4.

The tracker is built of modules consisting of a layer of sensitive silicon bonded on top readout

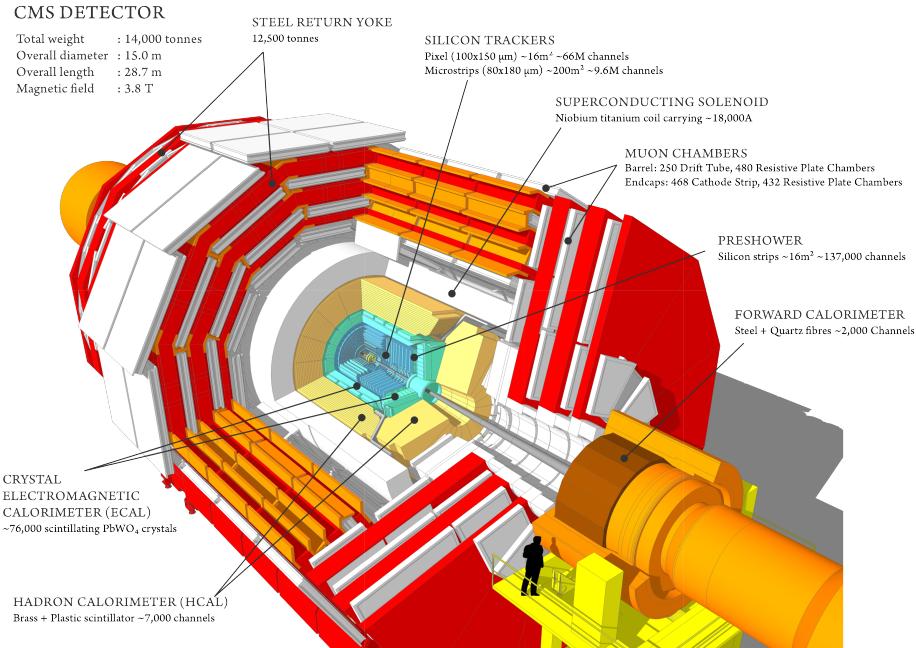


Figure 5.1: A diagram of the CMS detector. Specific detector subsystems are labeled.

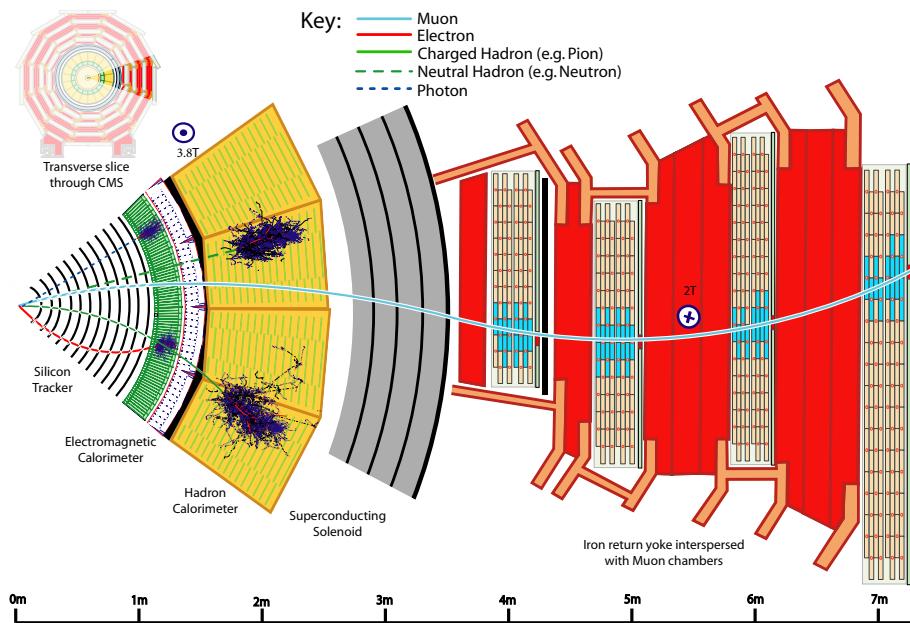


Figure 5.2: A diagram of the CMS detector in the  $r-\phi$  plane; the beam axis is perpendicular to the page; SM particle signatures within the detector are shown.

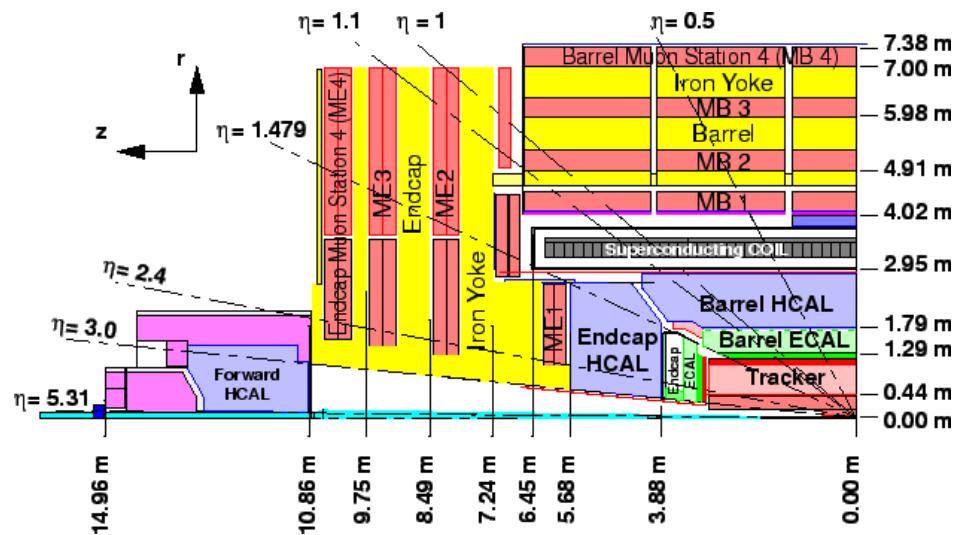


Figure 5.3: A diagram of the CMS detector in the  $r$ - $z$  plane; the beampipe is the thin cyan sliver along the bottom. The detector subtends a large solid angle about the interaction region.

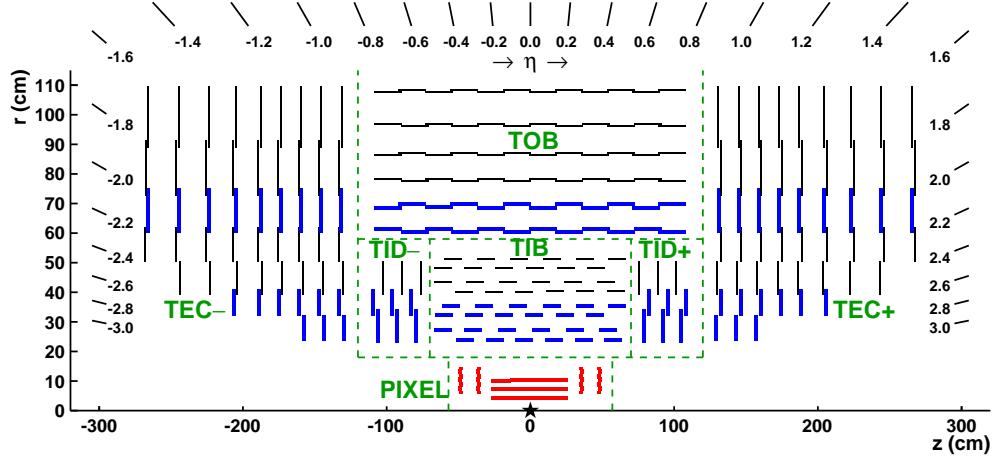


Figure 5.4: The CMS silicon tracker. The vertical and horizontal lines represent layers of silicon modules. Lines in blue represent layers with “stereo” hits, formed from two silicon layers.

electronics. The silicon is arranged as a p-n junction, reversed-biased and fully-depleted. As a charged particle travels through the material, it ionizes the silicon creating electron-hole pairs within the depletion zone. Electric fields accelerate the charge through the silicon to the readout electronics bonded to the back of the sensor. The readout chips amplify, digitize, and store the hit information before being piped outside the detector. The silicon is very thin ( $\sim 300 \mu m$ ), the tracker is constructed of as little material as possible so as not to perturb the trajectory of the particle.

### 5.1.1 Pixel Detector

The task of the pixel detector is to provide the spatial granularity required for precision track vertexing. The barrel region ( $|\eta| < 1.5$ ) of the pixel detector consists of 3 concentric cylinders sitting at radii of 4.4, 7.3, and 10.2 cm from the beamline. The endcaps ( $1.5 < |\eta| < 2.5$ ) consist of two discs on each side ( $\pm z$ ) placed at  $z = \pm 35.5, 48.5$  cm. A pixel module, used to form the detector layers, consists of 16 readout chips glued to a  $2 \times 6$  cm mechanical support structure. Bump-bonded to the readout chips are the  $285 \mu m$  thick silicon sensors. The readout chips and sensor are divided up into  $100 \times 150 \mu m$  pixels which allow the excellent hit resolution (over 65 million channels for the entire detector). The  $100 \mu m$  lengths are oriented to give the most precise measurement of the  $\phi$

coordinate of the track, as that is the direction on bending due to the magnetic field.

A new pixel detector was installed in 2016 to accommodate the ever-increasing instantaneous luminosity provided by the LHC [10]. The increase in the luminosity results in many more additional proton-proton interactions per bunch crossing, called *pileup*. High pileup results in a very dense environment for track reconstruction to operate as all the individual hits in the detector layers must be disentangled correctly. To increase the performance, the new detector therefore included **4** layers in the barrel and **3** endcap discs on either end. The readout chip was upgraded to fully digital readout to give larger hit buffers to accommodate the increased hit rate.

### 5.1.2 Strips Detector

The silicon strips detector sits immediately outside the pixel detector and provides additional hits along a particle’s trajectory. The barrel region ( $|\eta| < 1.5$ ) provides 10 layers of sensor situated between 20 and 116cm from the beamline. The endcap regions ( $1.5 < |\eta| < 2.5$ ) have a total of 12 layers situated between 58 and 282cm from the center of the detector, on each side. The silicon modules are partitioned in roughly 10cm long strips which are oriented parallel to the beamline. The strip pitch varies ranging from 80 – 180  $\mu\text{m}$  wide. The silicon thickness ranges from 320 to 500  $\mu\text{m}$  thick. Some of the layers, indicated in the blue lines in Figure 5.4, have two modules which are slightly rotated relative to each other to give a “stereo-hit”, providing a more precise position measurement.

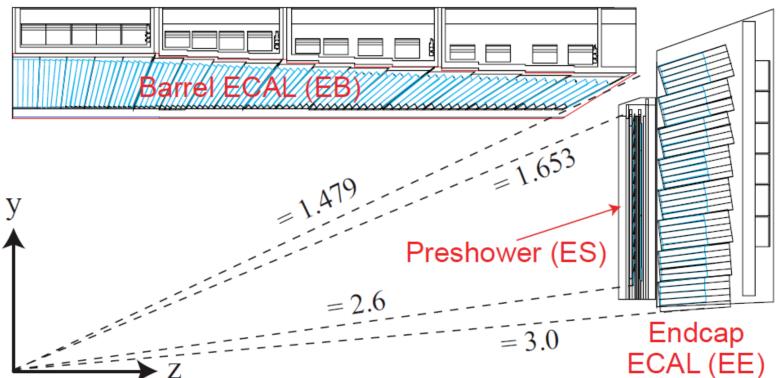
## 5.2 Electromagnetic Calorimeter

The electromagnetic calorimeter is responsible for the reconstruction of electrons  $e^\pm$  and photons  $\gamma$ . The energy is measured by collecting the light generated by an electromagnetic shower as the particle is absorbed in the calorimeter [11, 12].

Electromagnetic showers are created when high energy photons or electrons enter the material comprising the ECAL. The cross section for interactions of these particles with the detector scales with the square of the atomic nucleus, it was for this reason that the material is made of clear



(a) A single  $\text{PbWO}_4$  crystal attached to photomultiplier tube.



(b) Diagram of the ECAL layout, emphasizing the crystal orientation. A small gap in the  $\eta$  coverage is seen.

Figure 5.5: The CMS electromagnetic calorimeter.

$\text{PbWO}_4$  crystals. High energy photons in the detector predominantly lose energy by the creation of  $e^+e^-$  pairs via interactions with detector bulk. The predominant energy loss for high-energy electrons is through the emission of *bremsstrahlung*, or electromagnetic radiation created by the scattering of the electrons. An incident particle will therefore cause a shower in the detector as the interactions proceed until all the particles are sufficiently low energy and absorbed by the photodetectors mounted at the end of each crystal. Avalanche photodiodes with a gain of 50 and vacuum phototriodes with a gain of 10 are used in the barrel and endcap, respectively. The signal is then further amplified and digitized.

The electromagnetic calorimeter is divided into barrel ( $|\eta| < 1.5$ ) and endcap ( $1.5 < |\eta| < 3$ ) regions comprising 75,848 crystals. The crystals measure  $2.2 \times 2.2 \times 23\text{ cm}$  in the barrel and  $3 \times 3 \times 22\text{ cm}$  in the endcaps; they are oriented radially outward from the interaction region. A schematic of the detector is seen in Figure 5.5b. Additionally, the electromagnetic calorimeter serves as an absorber for the hadronic calorimeter, initiating a shower in approximately 1/3 of the hadrons that are headed the HCAL.

### 5.2.1 Preshower

The preshower is an additional detector which allows for greater spatial hit resolution of calorimeter clusters in the  $1.7 < |\eta| < 2.6$  region. Placed in front of the crystal calorimeter, it consists of a lead absorber, followed by a plane of silicon-strip sensors, followed by another lead absorber, followed by an orthogonal plane of silicon strip sensors. The silicon sensors are  $320\mu\text{m}$  thick and measure  $6.1 \times 6.1\text{cm}^2$  per sensor module. Measurements in the two orthogonal directions of each silicon layer are combined to provide more precise shower shape measurements.

The preshower is designed to properly identify photon pairs resulting from high- $p_T$  neutral pion decay (98.8% branching fraction). When a sufficiently high  $p_T$  pion decays, the daughter photons become collimated to the extent they can not be separately resolved within the ECAL crystals. Proper identification of neutral pions are important as they are copiously produced in high-energy particle collisions. There is a physics motivation for instrumenting this particular region of the ECAL: at least one of the photons from  $H \rightarrow \gamma\gamma$  falls within this region of  $\eta$ .

## 5.3 Hadronic Calorimeter

The hadronic calorimeter is responsible for the reconstruction of undecayed hadrons: pions  $\pi^\pm$ , protons  $p$ , neutrons  $n$ , and kaons  $K^\pm, K_L^0$ . It is composed of 4 distinct components with some amount of overlap. It is divided into barrel ( $|\eta| < 1.4$ , HB), endcap ( $1.3 < |\eta| < 3$ , HE), forward ( $3 < |\eta| < 5.2$ , HF), and outer ( $|\eta| < 1.2$ , HO) regions [13]. A diagram of the HCAL geometry is seen in Figure 5.6.

In the HE and HB, a particle will interact with the brass absorber inducing a shower of secondary particles. These secondary particles in turn may interact with the brass, and so on creating a hadronic shower within the detector. Interspaced within the absorber are clear plastic scintillator which create flashes of light after de-excitation of the scintillating molecules embedded in the plastic. Wavelength shifting fibers are routed throughout the plastic to absorb the light, which is then piped to hybrid photodiodes. Light incident on the photodiodes liberates electrons

via the photoelectric effect which are then accelerated onto the surface of a silicon diode which further amplifies and digitizes the signal. The particle energy is therefore measured by collecting light generated by a hadronic shower as the particle is absorbed in the calorimeter.

Each of these detectors consists of 17 alternating layers of 5cm brass and 1cm plastic. The readout of the optical signals are summed into *towers* of size of roughly  $0.09 \times 0.09$  in  $\eta - \phi$ , this results in the  $\eta$  segmentation seen in Figure 5.6, labeled from 1 to 15 for HB and then 18-29 in HE. Layers of constant color represent depth segmentation of the summed light within a single tower. e.g. two energy deposits in the central HE may be formed from a single shower. (Upgrades beginning in 2020 will increase this segmentation, as well as update the signal collection to using silicon photomultipliers [14].)

The HO sits **outside** the magnet in the barrel region  $|\eta| < 1.2$  and collects additional radiation not absorbed in the material in front of it with an additional layer of scintillator planes.

An addition to HB, HO, and HE, there is a steel calorimeter HF which detects radiation in the very forward region  $3 < |\eta| < 5.2$  on each side of the interaction point. This forward region has a very large radiation flux - the majority of the proton-proton interactions are “glancing blows” in which there is very little momentum transfer and the particles are deflected only slightly, directly into this region of the detector - this environment requires a different approach than in the rest of the HCAL. Each detector is comprised of 165cm thick steel interspersed with radiation-hard quartz fibers (parallel to the beamline) to collect light which is read out by photomultipliers. The tower size in this detector is about  $0.175 \times 0.175$  in  $\eta - \phi$ .

## 5.4 Solenoidal Magnet

The solenoidal magnet provides the magnetic field necessary to deflect charged particles within the tracker volume to allow for a measurement of the momentum. The tracker, ECAL, and HCAL all fit inside the magnet bore diameter of 6m and length of 12.5m. It delivers a 3.8T solenoidal field (parallel to the beampipe) within the tracker volume. The field is produced by running current through coils of superconducting NbTi wires cooled to less than 5K. The magnetic

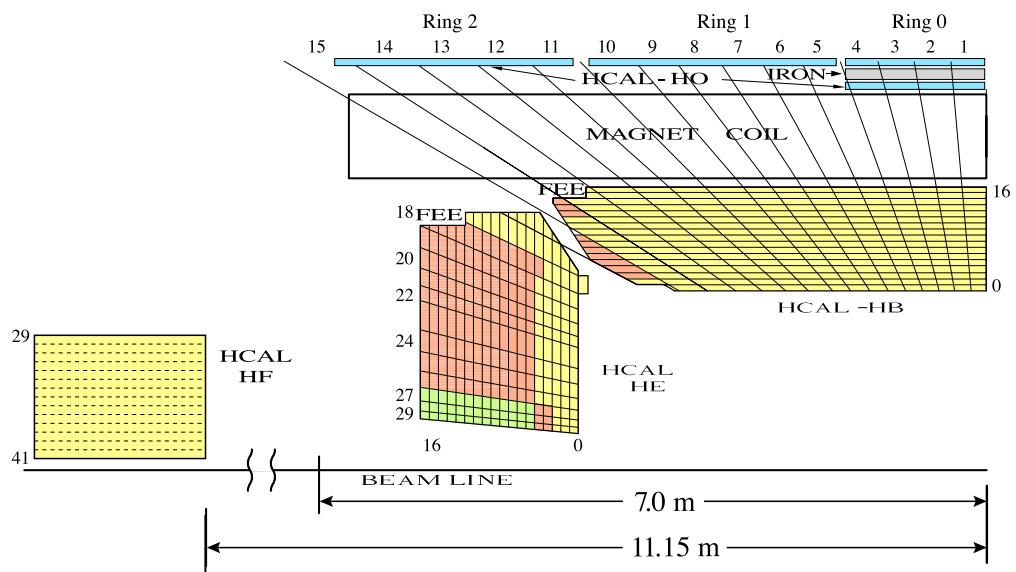


Figure 5.6: The CMS hadron calorimeter [15].

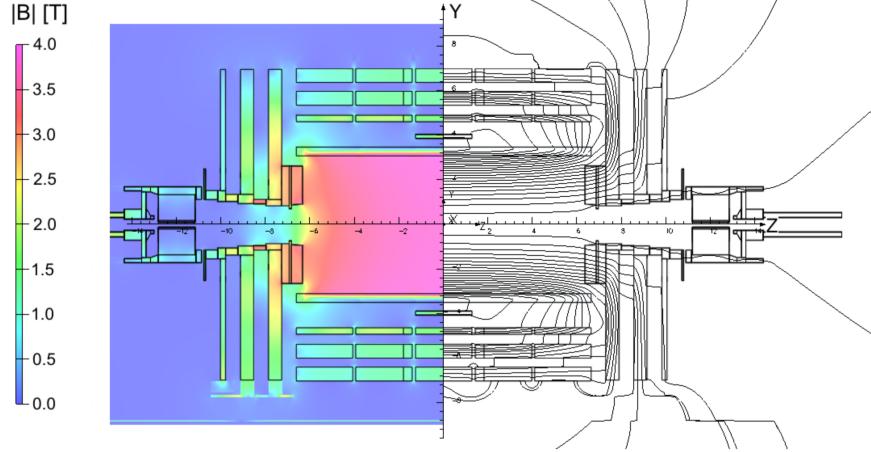


Figure 5.7: A simulation of the 4T CMS magnetic field. Note the uniformity within the tracker volume.

field lines are returned via steel yokes sitting outside the magnet interspaced within the muon tracker volume, the field strength throughout the muon system is approximately 2T [16]. A simulation of the magnetic field within the whole of CMS is seen in Figure 5.7 [17].

## 5.5 Muon System

The muon system is responsible for the reconstruction (and triggering) of muons  $\mu^\pm$ . The muon trajectory is reconstructed using ionization deposits left in layers of gaseous detectors. The muon momentum is measured by the curvature of the trajectory when immersed in the magnetic field [18].

The muon system sits at the furthest distance from the beamline, high performance reconstruction is made possible by the balanced combination of the muon and detector properties. Any particles which have made the journey to the muon system has traveled far and through many layers of detector material (e.g. Si, PbWO<sub>4</sub>, Cu), Fe) before finally being detected. Many particles are unstable and decay before reaching the muon system; other particles are absorbed in either of the calorimeters. But the muon has a relatively large mass (compared to an electron) and is not very likely to initiate electromagnetic showers in the ECAL. Nor does the muon interact strongly,

and so there will be no hadron showers within the HCAL. Muons have a sufficiently long lifetime to make it to the outer detector. This combination yields a very pure sample of reconstructed muons.

There are three components of the muon system: The drift tubes (DT) are in the barrel ( $|\eta| < 1.3$ ), the cathode strip chambers (CSC) are in the endcaps ( $0.9 < |\eta| < 2.4$ ), and resistive plate chambers (RPC) are in both regions ( $|\eta| < 1.6$ ). These detectors rely on different technology and have some amount of overlap with each other. All detectors participate in triggering and track reconstruction, but the DTs and CSCs provide greater position and momentum resolution, while the RPCs have excellent timing resolution allowing for more precise bunch crossing tagging.

### 5.5.1 Drift Tubes

The drift tubes are used for muon tracking in the barrel portion of the detector ( $|\eta| < 1.3$ ). The basic element is a gas tube  $4 \times 1.3$  cm in transverse size and 2-4 m long (depending on its position). High-voltage is applied to a wire strung the length of the cylinder and collects charge released when an incident muon ionizes an 85/15% Ar/CO<sub>2</sub> gas mixture [19].

The drift tubes are divided into four barrel regions (each called a station) at different radii within the magnetic return yoke. Each station contains 3 *superlayers*, where a superlayer is composed of four layers of stacked tubes, each layer staggered by one half width. For each station, two of the superlayers are oriented parallel to the beamline for  $r - \phi$  measurements and one superlayer is perpendicular to the beamline to allow for measurements of the r-z position. An image of a DT station is seen in Figure 5.8.

### 5.5.2 Cathode Strip Chambers

The cathode strip chambers are used for muon tracking in the endcap portion of the detector ( $0.9 < |\eta| < 2.4$ ). The system is divided up into 468 trapezoidal chambers arranged in 2 or 3 concentric rings on a disk. There are 4 discs on either side of the detector ( $\pm z$ ). The geometry of the chambers on a disk are seen in Figure 5.9a (an example image of the hit occupancy of a disc during cosmic ray runs). Each chamber (diagram in Figure 5.9b) consists of 6 layers of electrode

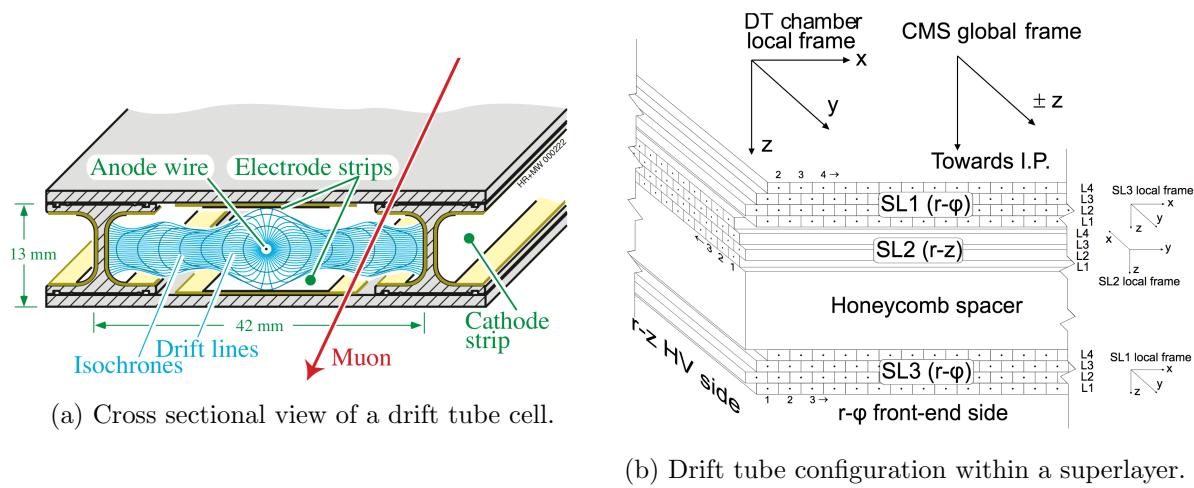


Figure 5.8: The CMS muon drift tube detector [20].

planes separated by a gas layer of  $\text{C}_2\text{H}_2\text{F}_4$  (freon) and  $\text{C}_4\text{H}_{10}$  (isobutane). Wires are strung in the phi direction (concentric circles about the z axis) and therefore make a measurement of the radial coordinate of the hit. The electron shower generates an image charge in cathode planes. For each layer, one of the planes is segmented into strips which are perpendicular to the wires providing a good measure of the  $\phi$  coordinate [21].

### 5.5.3 Resistive Plate Chambers

Resistive plate chambers cover the region  $|\eta| < 1.6$  and are interspersed within CSC and DT and the magnetic return yoke [22]. They have an excellent timing resolution of about 3 ns which allows for fast muon triggering and identification of the different bunch crossings. Pattern matching across the hits in the different layers allows for estimates of the muon  $p_T$  to be used in further trigger processing. Hits created in the resistive plate chambers are additionally used for global fitting of the muon tracks.

The resistive plate chambers consist of an airtight system of two parallel high-resistivity planes separated by a 1 cm gas gap. The outside of each plate is coated to form an electrode for the high-voltage bias. On top of each electrode sits aluminum strips which are insulated from the electrode and serve as the readout. Electron showers created in the gas bulk induce an image charge on the strips which is then recorded. The gas mixture is 95/5%  $\text{C}_2\text{H}_2\text{F}_4$  (freon) and  $\text{C}_4\text{H}_{10}$  (isobutane), with trace amounts of  $\text{SF}_6$ . A diagram of an RPC chamber is seen in Figure 5.10.

## 5.6 Trigger System

While in operation mode, the LHC provides collisions at a rate of 40 MHz (25 ns per bunchcrossing). This is a phenomenal rate which the CMS detector bandwidth is not able to accommodate, nor does the experiment have access to the amount of storage space necessary to store all this information. Therefore, the CMS detector makes use of a trigger system to quickly determine if the event is 'interesting' and will be saved for storage - events which are not triggered are lost forever. Examples of interesting events are those with high- $p_T$  muons, or a large imbalance in the

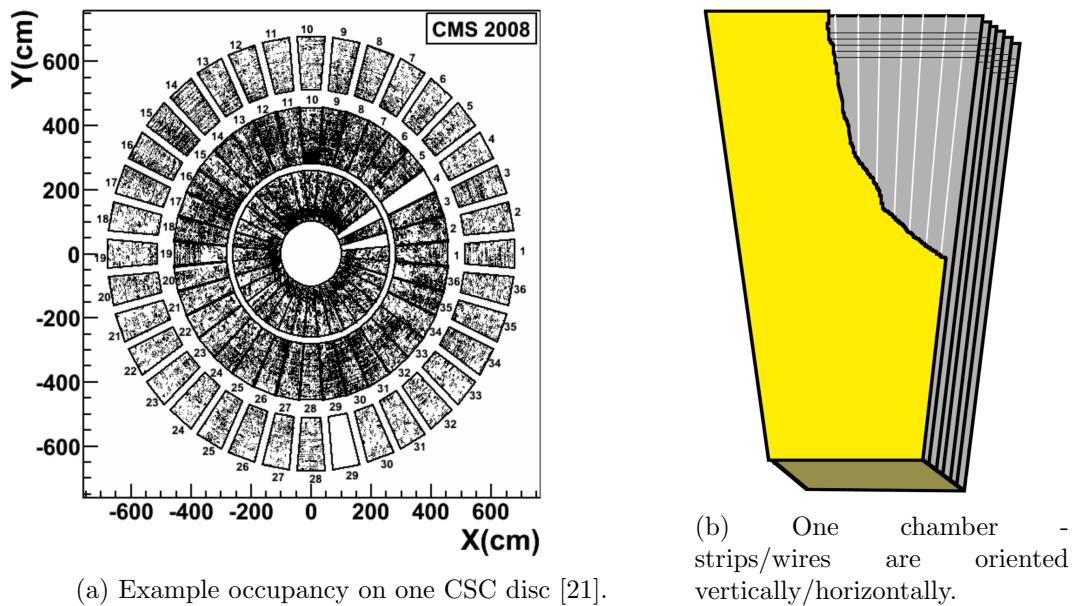


Figure 5.9: The CMS muon cathode strip chambers.

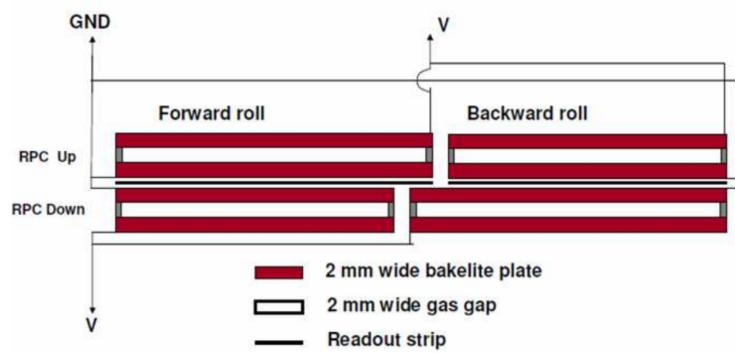


Figure 5.10: The CMS muon resistive plate chamber [22].

total momentum of the event [23].

The trigger consists of two stages known as the Level-1 (L1) and High-Level Trigger (HLT). L1 is a hardware based trigger which combines information from the calorimeters and muon systems to make a decision if the event will be passed to HLT for further processing. L1 is able to reduce the event rate from 40 MHz to 100 kHz and must make the decision within  $4 \mu s$ . Primitive objects such as calorimeter energy deposits or muon track segments are first constructed locally within the detector before being combined to form the global decision at L1. If the decision is made at L1 that the event is of potential interest, it is passed to HLT. HLT is a software based trigger which makes use of more sophisticated reconstruction algorithms which can be tuned to select events of choice.

## Chapter 6

### Event Reconstruction

#### 6.1 Basic Elements From the Detector

Depending on their nature, the particles emanating from the collision leave various forms of energy deposits in the different subdetectors. All these signals need to be aggregated and processed to allow for the reconstruction of what could be considered particle-level information. The first step in this process consists of building *Particle Flow elements* using information only locally within each subdetector. There are four primary elements from the subdetectors: tracks from the tracker, calorimeter deposits in each of the ECAL and HCAL, and tracks from the muon detector. The elements are eventually combined via the *Particle Flow algorithm*, yielding reconstructed particles used for physics analysis.

Note that the definition of any object within the detector makes use of additional selection criteria which are not generally discussed here. For instance, one may require that a track in the tracker have at least 3 hits in the pixel detector, or that a calorimeter hit is above some minimum threshold energy. The effects of these criteria are generally a balance between the reconstruction efficiency of any given particle and the probability to misidentify a particle (purity).

##### 6.1.1 Tracks

Tracker hits are formed in the pixel and strips detectors by clustering any hits in neighboring elements of the detector plane. The cluster position is measured by weighted average of the individual channel positions. Charge sharing among detector channels allow for a finer spatial resolution

in the position measurement. Track reconstruction first begins by forming track seeds consisting of a small number of detector hits. These track seeds are then projected onto successive detector layers looking for nearby additional hits. This follows the Kalman filtering procedure, in which track information is updated after the addition of each hit [24].

The track building procedure follows an iterative procedure, where the requirements on the quality of the track seed decrease as the iterations proceed. The first iterations begin with seeds consisting of 3 pixel hits and lead to high performance reconstruction of high  $p_T$  tracks emanating from the collision region. The iterations proceed until essentially only requiring hits in the outer tracker and reconstructs displaced tracks or those missing hits. The iteration procedure provides a balance between reconstruction efficiency, track purity, and computation economics. Table 6.1 lists these iterations - the requirements on the seed and types of tracks that iteration targets.

Table 6.1: Seeding requirements for each step in the iterative track reconstruction [25].

Iteration	Name	Seeding	Targeted Tracks
1	InitialStep	pixel triplets	prompt, high $p_T$
2	DetachedTriplet	pixel triplets	from b hadron decays, $R \lesssim 5$ cm
3	LowPtTriplet	pixel triplets	prompt, low $p_T$
4	PixelPair	pixel pairs	recover high $p_T$
5	MixedTriplet	pixel+strip triplets	displaced, $R \lesssim 7$ cm
6	PixelLess	strip triplets/pairs	very displaced, $R \lesssim 25$ cm
7	TobTec	strip triplets/pairs	very displaced, $R \lesssim 60$ cm
8	JetCoreRegional	pixel+strip pairs	inside high $p_T$ jets
9	MuonSeededInOut	muon-tagged tracks	muons
10	MuonSeededOutIn	muon detectors	muons

Electron tracking is performed with a slightly modified algorithm to better account for the electron's energy loss in the detector. Electrons are very light particles and more susceptible to emitting *bremsstrahlung* radiation than ionizing the silicon. Bremsstrahlung radiation occurs when the electron scatters from a charged nucleus within the tracker bulk and emits a photon. This results in both energy loss and momentum changes as the electron traverses the detector. This results in very non-Gaussian energy loss mechanisms in which the Kalman filtering is non-optimal.

### 6.1.2 ECAL & HCAL Clusters

*Superclusters* are built by first identifying a crystal with the largest energy deposit, this is called a *seed*. The supercluster is then formed by aggregating any hits among the neighbors (8) of the hits already in the cluster. This process then proceeds building all the superclusters and consuming all the calorimeter hits. Superclusters are built separately in the barrel and endcap. Within a given supercluster, N clusters are identified using an iterative algorithm assuming the observed hits arise from N Gaussian-distributed energy deposits; each of energy E, position in the  $\eta - \phi$  plane  $\vec{\mu}$ , and width  $\sigma$  scale set by the crystal size.

### 6.1.3 Muon Tracks

As there are multiple detector layers within a single muon station, track segments are locally formed within a chamber for both the DTs and CSC. These track segments represent the muon momentum at that station; pattern recognition is able to provide to measurement and its high speed allows the segments to be used trigger primitives for the muon systems. For track reconstruction, the track segments act as seeds for the track finding algorithm. The hits in each the DT, CSC, and RPC subdetectors are used in the final track reconstruction.

## 6.2 Obtaining a Particle-level Description

Once the elements have been built, the Particle Flow algorithm exploits the information from each of the detectors to form the best possible particle candidate [25]. As different varieties of particles have unique signatures in the detector, particle identification is aided by the particular combination of elements *linked* with one another. An illustrative example of these combinations are shown in Figure 5.2. Elements are linked when projections from one element to the other are spatially consistent. There are six primary links:

- Tracks formed in the tracker are linked to an ECAL or HCAL cluster if the projection of the track, at a depth of the expected maximum of a shower in the ECAL or at one interaction

length inside the HCAL, lies within the cluster area.

- ECAL and HCAL clusters are linked if the ECAL cluster falls within the envelope of the HCAL cluster; ECAL provides finer spatial resolution compared to the HCAL.
- If a Preshower cluster is within the envelope of an ECAL cluster the two are linked; Preshower has finer spatial granularity.
- A tracker track and a muon track are linked if their projections onto a common surface are spatially consistent.
- To collect bremsstrahlung radiation (photons) associated to an electron track, an ECAL cluster is linked to tracker tracks if projections tangent to the track at any of the tracker layers lies within the cluster volume. To catch these bremsstrahlung photons which additionally then convert into an  $e^+e^-$  pair in the tracker, track pairs consistent with  $e^+e^-$  conversion are linked if their common momentum points to one of the track tangents.
- Tracks consistent with arising from a secondary vertex are linked to allow for reconstruction of nuclear-interactions.

Particle Flow *blocks* are constructed by aggregating objects directly or indirectly linked with one another. The Particle Flow algorithm then processes each block in turn to create the final reconstructed particles. The algorithm builds the objects in the following order

- (1) **Muons:** There are three types of tracks which can be used for muon reconstruction:

- ***standalone*** muons are built from tracks reconstructed solely in the muon system.
- ***tracker*** muons are built from tracks reconstructed solely in the inner silicon tracker. They are tagged as such if the track projection is consistent with any track segments found in the muon system.

- *global* muons are reconstructed using the hits from both the inner silicon tracker and the muon stations. Global muons are reconstructed when a track in the tracker and a track in the muon system are compatible.

Any ECAL or HCAL clusters associated with the muon track are used as muon selection/definition criteria if those clusters are found to be consistent with the muon hypothesis.

#### (2) Electrons & Photons:

An electron is formed by combining a track in the silicon tracker with a cluster in the ECAL. Its energy assignment uses a combination of both elements. The momentum direction is made using the track in the silicon tracker, as it gives greater spatial resolution. A photon is defined as an ECAL cluster not associated with a track. Surrounding energy in the HCAL must not exceed 10% of the ECAL supercluster energy.

Electrons and isolated photons are reconstructed within the same Particle Flow step to account for similar behavior within the tracker bulk. There is a large probability for both a) an electron to radiate a brehmsstrahlung photon and b) for a photon to convert to an  $e^+e^-$  pair. Therefore in object reconstruction care must be taken to collect the photons radiated from electrons in order to make appropriate measurements of the particles.

#### (3) Hadrons & Photons:

Hadrons & non-isolated photons result from hadronization/fragmentation of jets. ECAL clusters not associated to any tracks are assigned to be photons. Neutral hadrons ( $K_L^0$ , neutrons) are reconstructed from HCAL clusters with no associated track; neutral hadrons leave a very small amount of energy in the ECAL. Charged hadrons ( $\pi^\pm$ ,  $K^\pm$ , protons) are reconstructed using the remaining tracks and HCAL deposits.

#### (4) Nuclear Interactions:

Nuclear interactions can occur when hadrons from the pp collision interact within the detector material causing a shower of secondary particles. If multiple tracks are linked through a common secondary vertex they will be summed to create a single charged hadron particle which replaces its constituents in the particle list of the event.

## 6.3 Additional High-Level Objects

### 6.3.1 Jets

Bare quarks and gluons can never be observed in Nature due to a QCD phenomenon called *color confinement*. Therefore, quark and gluon production manifests as a “jet” of color-neutral particles emanating from the production point. These particles can be clustered together to reconstruct the original parton. The jets used in this analysis are made by clustering particles with the “anti-kt” algorithm with cone sizes of  $\Delta R = 0.4, 0.8$  [26], denoted as AK4 and AK8 jets respectively. This algorithm produces nearly conical jets and is infrared and collinear safe. The 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 that decay to multiple jets (e.g.  $t$ , H, Z, W).

### 6.3.2 b-tagging of Jets

Jets resulting from the production of b quarks (and to some extent c quarks) garner special attention in our experiment. As usual for quarks and gluons, the b-quark will quickly hadronize and form a b hadron. However, the lifetimes of b hadrons are such that it will generally travel hundreds of microns before decaying. Vertexing the tracks resulting from the decay will reveal the presence of a *secondary vertex* which is spatially displaced from the primary vertex from which the other hadrons inside the jet originate. This secondary vertex allows for one handle on being able to identify jets as coming from b quark production. Other handles include the momenta and multiplicity of the other particles clustered into the jet.

In addition to tagging jets as originating from a single b quark, tagging of jets as originating from **two** b quarks is also possible [27].

### 6.3.3 Invisible Particles $\rightarrow p_T^{\text{miss}}$

Neutrinos are so weakly interacting that they leave no energy deposits in CMS and cannot be detected by our experiment. Although direct detection is not possible, we are able to infer their presence. The net momentum of the protons involved in the collisions are zero, but the individual partons (quarks and gluons) within the proton carry unknown fractions of this total (longitudinal) momentum. Only in the transverse direction may we require momentum conservation. We define this imbalance as:

$$p_T^{\text{miss}} \equiv \left| - \sum_i \vec{p}_T^i \right|, \quad \forall \text{ particles i.} \quad (6.1)$$

In the case all the particles in the event are perfectly reconstructed,  $p_T^{\text{miss}}$  would equal zero. Large values indicate the presence of an undetected particle, such as a SM neutrino or something more exotic like the light supersymmetric neutralinos  $\chi_0^0, \chi_1^0$ . This quantity is sometimes labeled as *MET*.

### 6.3.4 $\pi^0$ meson

Charged and neutral  $\pi$  mesons are the lightest of all hadrons with masses of 140 and 135 MeV, respectively. Pions are copiously produced within hadronic interactions and often contain much of the jet energy. As a rule-of-thumb, the charged to neutral energy fraction within a jet is approximately two-thirds to one-third (to match the respective pion multiplicities).

Because the neutral pion branching to photons is over 98% and its lifetime is relatively short (25.5 nm), the detection of neutral pions therefore involves reconstruction of photon pairs with the appropriate invariant mass. This is simple enough in principle, but as the pion  $p_T$  increases the photons from the pion decay become more and more collimated. Eventually, for pions above about 7 GeV, individual photons are not able to be resolved because of the ECAL resolution (crystal

size), the same crystal gets the energy from both photons. This was the primary motivation for the Preshower ECAL detector.

## Chapter 7

### Search for new physics using boosted H bosons and missing energy

#### 7.1 Motivation & Strategy

If a more unifying theory than the SM exists it certainly has not been forthright in its manifestation. One possibility for the lack of discoveries of phenomena not explained within the SM is that there are indeed particles existing in nature which have not been observed, but they have such a large mass that the energy of the proton-proton collisions provided by the LHC is insufficient to directly create them. The outcome of many searches for new particles is thus the setting of limits placed on the cross section of their production mechanism. As these cross sections are dependent on the particle mass, the limits can be interpreted as setting lower bounds on the mass of any new physics - if the particle were any lighter than this limit it would have been produced copiously enough for its unambiguous detection (see for example [5, 6]). As these particles become more massive more momentum is imparted upon the particles involved in the final state; any SM particles resulting from the decay of higher mass states will be produced with large momentum (this is called high boost). As a particle becomes more boosted its decay products are emitted at smaller angles, eventually collimating enough to be reconstructed as a single jet. If new physics exists with masses accessible at 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 finding new particles of large mass.

The Minimal Supersymmetric SM 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 [3]. 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 (similar to a neutrino). Therefore, events with a large momentum imbalance are also interesting as potential regions for SUSY.

With this as motivation, we designed an analysis searching for physics beyond the SM 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  $b\bar{b}$  decay mode, 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 seen in Figure 7.1. These are known as *SMS* models and serve to simplify the SUSY parameter space into particle masses, cross sections, and branching fractions [28]. In general, the diagrams begin with sparticle pair production which subsequently cascade decay to final states consisting of stable SM particles and a pair of stable supersymmetric particles which are the lightest in the theory. The blob in the figure indicates we are not interested in the particulars of the gluino production mechanism. Figures 7.1a and 7.1b are known as the *T5HH* and *T5ZH* models respectively. The proton-proton interaction produces a pair of gluinos  $\tilde{g}$  which decay to a neutralino  $\tilde{\chi}_2^0$  along with the emission of SM quarks. This neutralino  $\tilde{\chi}_2^0$  further decays into neutralino  $\tilde{\chi}_1^0$  with the emission of a SM Higgs or Z boson.

## 7.2 Baseline Selection and Object Definition

The salient feature of our analysis is the presence of two high- $p_T$  AK8 jets and large  $p_T^{\text{miss}}$ . Our most significant backgrounds are those that produce  $p_T^{\text{miss}}$  and can be divided into two categories depending on if it comes from a “true” or “fake” source. True  $p_T^{\text{miss}}$  arises from a neutrino or otherwise unobservable particle escaping our detector - a genuine imbalance in the detectable final-state momentum of the event. Fake  $p_T^{\text{miss}}$  arises from some sort of “imperfection” in our reconstruction process, for example the misidentification of a hadron.

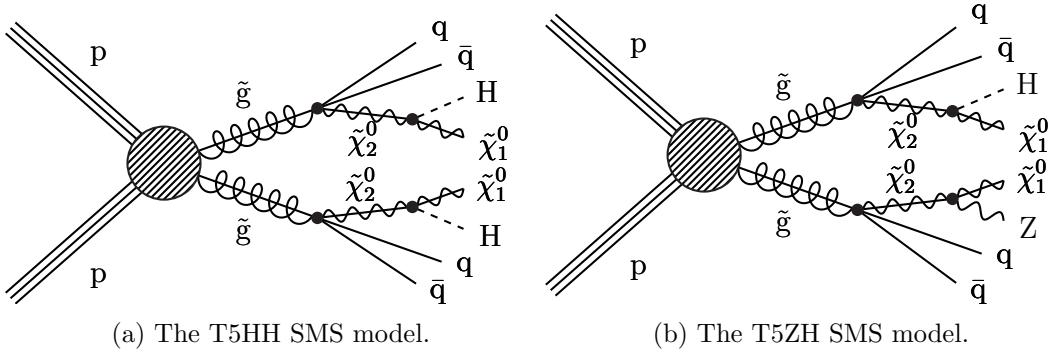


Figure 7.1: Diagrams of the benchmark models used for motivation of the targeted signal.

We approach our background as being constituted of three primary types. Two of them involve the production of neutrinos and the third is a result of jet resolution:

- $Z \rightarrow \nu\bar{\nu}$  in which the neutrinos from the  $Z$  decay results in true  $p_T^{\text{miss}}$  ('Z-invisible').
- $W \rightarrow \ell\nu$  in which the lepton  $\ell$  is not properly identified, resulting in the event not being vetoed, and the associated neutrino from the leptonic decay creating true  $p_T^{\text{miss}}$  ('lost-lepton'). The  $W$  boson can be produced through the decay of a top quark, either in single-top or  $t\bar{t}$  events;  $t\bar{t}$  events generate two  $W$  bosons, in which the other may decay into an AK8 jet. A  $W$  boson may additionally be produced directly in association with other jets.
- Jet production via QCD in which the  $p_T$  of a jet is mis-measured, most commonly under-measured, creating a fake source of  $p_T^{\text{miss}}$ .

To mitigate these backgrounds, we establish a baseline selection choosing events with all-hadronic final states and missing transverse momentum ( $p_T^{\text{miss}}$ ), as motivated by Figure 7.1. The baseline selection is as follows:

- $\geq 2$  AK8 jets, with  $p_T > 300$  GeV and  $50 < \text{mass} < 250$  GeV
- $p_T^{\text{miss}} > 300$  GeV;

- no isolated electrons with  $p_T > 10 \text{ GeV}$ :

Isolation requires the  $p_T$  sum of the particles within a cone of  $\Delta R < 0.2$  to be less than 10% of the electron  $p_T$ . (See [29] for details about “mini-isolation”.)

- no isolated muons with  $p_T > 10 \text{ GeV}$ :

The energy fraction for the isolation requirement is relaxed to 20%.

- no isolated tracks:

To remove events with top or W production in which the W decays to a tau. As an isolated track is defined by looser criteria than that of an electron or muon, this cut also serves to increase the efficiency of the isolated electron and muon vetoes. Leptonic tracks must satisfy  $p_T > 5 \text{ GeV}$  and have 20% isolation. Hadronic tracks must satisfy  $p_T > 10 \text{ GeV}$  and have 10% isolation.

- $\Delta\phi_{1,2,3,4} > 0.5, 0.5, 0.3, 0.3$ ;  $\Delta\phi_i \equiv \Delta\phi(\vec{p_T^{\text{miss}}}, \text{AK4 jet}_i)$

This cut requires that the difference in  $\phi$  between the  $\vec{p_T^{\text{miss}}}$  vector and each of the four highest- $p_T$  jets is sufficiently large to remove events in which a jet has been under-measured giving rise to fake  $\vec{p_T^{\text{miss}}}$ . If less than four AK4 jets are available the additional cuts are removed.

To tag jets from  $H \rightarrow b\bar{b}$  a dedicated MVA algorithm has been developed by the CMS collaboration to discriminate these signal jets from those produced by QCD processes [27]. The algorithm makes use of the kinematics expected by having secondary vertices arising from the b-quark decays. Discriminating variables include the number, mass, energy, and position of the secondary vertices in the jets. Additional variables include the distance of closest approach between the vertices and tracks (when projected backwards); tracks arising from displaced vertices are not expected to project back to the primary interaction point. The distribution of this discriminator for the two highest- $p_T$  AK8 jets are seen in the left and right panels of Figure 7.2; signal-like events peak towards larger values. To  $b\bar{b}$  tag the AK8 jets we choose the loose working-point ( $>0.3$ )

corresponding to an efficiency of approximately  $70 - 80\%$  for  $H \rightarrow b\bar{b}$  (see Figure B.2). The stacked histogram and solid lines shows the distribution after baseline selection for simulation and two representative signal points, respectively. Appendix A details the  $b\bar{b}$ -tagger in more depth.

Additionally, to tag H or Z bosons a requirement is made on the invariant mass of the jet. We use the mass of the so called “pruned” jet - a method involving removal of soft and wide-angle radiation inside the jet [30]. The pruning is very powerful for discriminating QCD jets from those produced by heavy particle decay. Our tagging requires the jet mass to fall within a window [85, 135 GeV] to be consistent with that of the H boson. The distributions of the jet mass are seen in Figure 7.2. The same identification criteria are applied to tag an AK8 jet as either an H or Z boson (there is no distinction made).

### 7.3 Dataset & Trigger

We use a total of  $35.9 \text{ fb}^{-1}$  of data collected by the CMS experiment in 2016. Events are selected with the requirement of at least 100 GeV of  $p_T^{\text{miss}}$  calculated at high-level trigger (`HLT_PFMET100_PFMHT100_IDTight`, HLT). To improve the trigger efficiency, the logical OR of similar triggers of thresholds 110, 120 GeV are included.

The HLT is seeded at level-1 (L1) by the requirement of at least 100 GeV of missing transverse energy (`L1_ETM100`). Missing transverse energy is calculated by the vectorial sums of the regional energy deposits in the calorimeters (and rotation by  $180^\circ$ ). This is additionally ORed with similar seeds of thresholds 90, 80, 70, 60, 50 GeV. Two additional ORed seeds require calorimeter jets with  $p_T$  of at least 60 GeV (`L1_ETM60_Jet60_dPhi_Min0p4`, `L1_DoubleJetC60_ETM60`).

The trigger efficiency is defined as  $\epsilon = N_1/N_2$ , where  $N_2$  is the total number of events passing baseline selection and  $N_1$  is the total number of events selected by the trigger and passing baseline selection. In order to calculate this efficiency, we need to form an additional (hopefully independent) dataset by selecting events via a *reference trigger*. Our reference trigger requires a single electron of  $p_T > 27 \text{ GeV}$ , in addition to the baseline selection we require these events to contain at least three AK4 jets and exactly one reconstructed electron of  $p_T > 25 \text{ GeV}$ . The signal region trigger

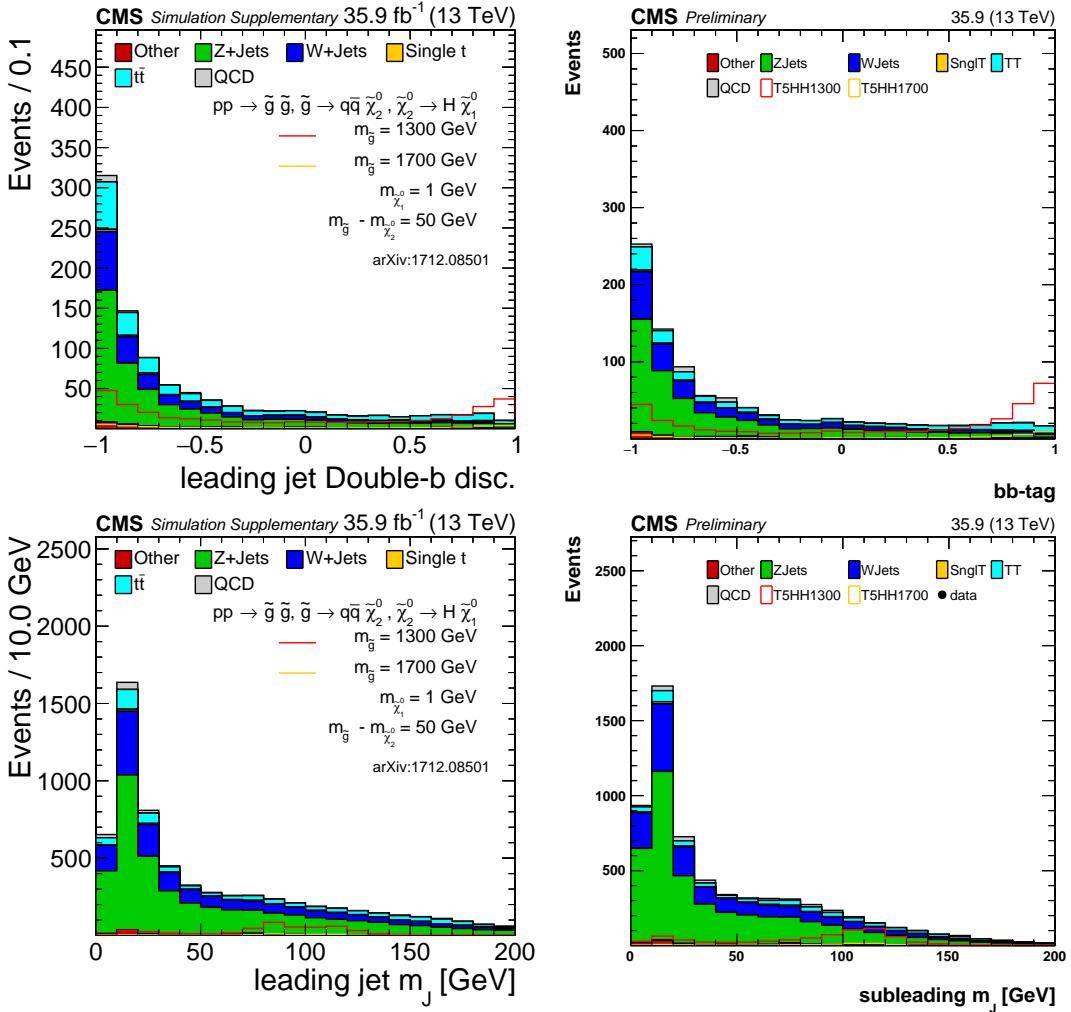


Figure 7.2: Distributions of the bb-tagging discriminator (top row) and the jet mass (bottom row) for the highest (left column) and second highest (right column)  $p_T$  AK8 jets. In the signal models, the H bosons are allowed to decay inclusively.

is found to be greater than 98% for events with  $p_T^{\text{miss}} > 250 \text{ GeV}$  and  $H_T > 300 \text{ GeV}$  [5]. This is applied as a systematic error to the final results.

## 7.4 Event Simulation

Event simulation of the proton-proton collisions proceeds in several steps. MadGraph@NLO2.2.2 [31] is a MC event generator used to simulate proton-proton events. Parton distribution functions, used to model the quark and gluon momenta distribution within the proton, are taken from NNPDF 3.0 [32]. Parton showering, which models quark and gluon evolution into hadrons, or final/initial state radiation, is performed with Pythia [33]. The final-state particles are traced through the detector using GEANT. For example, this includes the effect of the magnetic field, interactions with both the dead and sensitive detector material, and additional particle decay [34]. This simulated data is processed in the same manner as that of the physical experiment.

### 7.4.1 Standard Model Processes

The SM samples which enter as the primary backgrounds are listed in Table 7.1 (see Section 7.5.1 for a discussion of the background). All samples are generated with a pileup distribution with an average of 25 interactions per bunch crossing and a 25ns interval between bunches. For acceptable statistics over a wide range of parameter space, the samples are often binned in  $H_T \equiv \sum_{\text{AK4 jets}} p_T$ . As our event selection requires at least two AK8 jets with  $p_T > 300 \text{ GeV}$  we roughly operate in the regime of  $H_T > 600 \text{ GeV}$ .

### 7.4.2 Signal Models

For commissioning of the analysis technique (as well as the limit-setting procedure, see Section 7.7) Monte Carlo samples with our final-state signal topology were generated, based on the processes in Figure 7.1. The signal sample follows the same processing chain as the SM samples. The mass splitting between the gluino  $\tilde{g}$  and neutralino  $\tilde{\chi}_2^0$  is fixed to 50 GeV, resulting in low  $p_T$  SM quarks produced in the gluino  $\tilde{g}$  decays. The mass of the neutralino  $\tilde{\chi}_1^0$  (LSP) is fixed to 1

Table 7.1: SM samples used in the analysis.  $H_T \equiv \sum_{AK4\text{jets}} p_T$  is the total hadronic energy in the event.  $\sigma$  is the cross section.  $\int \mathcal{L} = N/\sigma$  is an alternative way to express the number of generated events.

process	final state	$H_T$ (GeV)	$\sigma$ (pb)	$\int \mathcal{L}$ ( $\text{fb}^{-1}$ )
$t\bar{t}$	$t \rightarrow \ell\nu, \bar{t} \rightarrow 2q$	inclusive	182.72	283.90
$t\bar{t}$	$\bar{t} \rightarrow \ell\nu, t \rightarrow 2q$	inclusive	182.72	326.48
$t\bar{t}$	$2\ell$	inclusive	88.34	346.25
$t\bar{t}$	inclusive	[600, 800]	2.734	5231.81
$t\bar{t}$	inclusive	[800, 1200]	1.121	9416.61
$t\bar{t}$	inclusive	[1200, 2500]	0.198	14819.34
$t\bar{t}$	inclusive	[2500, $\infty$ ]	0.002	221088.29
QCD	inclusive	[200, 300]	1735000	0.03
QCD	inclusive	[300, 500]	366800	0.16
QCD	inclusive	[500, 700]	29370	1.95
QCD	inclusive	[700, 1000]	6524	6.68
QCD	inclusive	[1000, 1500]	1064	12.62
QCD	inclusive	[1500, 2000]	121.5	32.63
QCD	inclusive	[2000, $\infty$ ]	25.42	239.30
Z+jets	$\nu\bar{\nu}$	[100, 200]	344.8	54.13
Z+jets	$\nu\bar{\nu}$	[200, 400]	95.53	208.46
Z+jets	$\nu\bar{\nu}$	[400, 600]	13.20	77.30
Z+jets	$\nu\bar{\nu}$	[600, 800]	3.148	1795.26
Z+jets	$\nu\bar{\nu}$	[800, 1200]	1.451	1486.09
Z+jets	$\nu\bar{\nu}$	[1200, 2500]	0.355	1029.81
Z+jets	$\nu\bar{\nu}$	[2500, $\infty$ ]	0.0085	47498.87
W+jets	$\ell\nu$	[100, 200]	1627.45	18.16
W+jets	$\ell\nu$	[200, 400]	435.24	45.88
W+jets	$\ell\nu$	[400, 600]	59.18	123.64
W+jets	$\ell\nu$	[600, 800]	14.58	221.32
W+jets	$\ell\nu$	[800, 1200]	6.66	1123.13
W+jets	$\ell\nu$	[1200, 2500]	1.608	153.44
W+jets	$\ell\nu$	[2500, $\infty$ ]	0.039	6497.28

GeV. We have samples with a range of gluino  $\tilde{g}$  masses from 750 to 2200 GeV. The  $p_T$  distribution for the generated H bosons in these samples is seen in Figure 7.3 for a number of gluino  $\tilde{g}$  masses. Additionally the angular separation  $\Delta R \equiv \sqrt{\Delta\phi^2 + \Delta\eta^2}$  between the  $b\bar{b}$  pair is shown. As the  $p_T$  of a parent boson increases the  $b\bar{b}$  pair from its decay tend to align, allowing complete reconstruction with a single AK8 jet. The distributions in Figure 7.3 motivate our use of two such high- $p_T$  AK8 jets.

## 7.5 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 background in the signal regions. The events are categorized according to whether the two highest  $p_T$  AK8 jets are a) in the signal or sideband mass region and b) have or have not been  $b\bar{b}$  tagged. A diagram of this partitioning is seen in Figure 7.4. An additional dimension is added by binning in  $p_T^{\text{miss}}$ : [300, 500 GeV], [500, 700 GeV], [700,  $\infty$  GeV]. This additional binning, motivated by the  $p_T^{\text{miss}}$  distributions of Figure 7.5, allows us to maximize sensitivity for both low and high  $p_T^{\text{miss}}$  signal. This gives a total of  $2 \times 3 = 6$  signal and  $4 \times 3 = 12$  sideband bins. The two signal regions  $A_1$  and  $A_2$  contain events with exactly one and exactly two jets being consistent with H/Z boson decay, respectively.

Assuming that there is no correlation between the jet mass and the  $b\bar{b}$ -tagging one would expect that:

$$\frac{A_1}{B_1} = \frac{A_2}{B_2} = \frac{C}{D} \quad (7.1)$$

Rearranging this gives a prediction for the events in the signal regions

$$A_{1,2}^{\text{predicted}} = \left( B_{1,2} \cdot \frac{C}{D} \right)^{\text{observed}} \quad (7.2)$$

The expected  $p_T^{\text{miss}}$  distribution from simulation is seen in the stacked histograms of Figure 7.6.

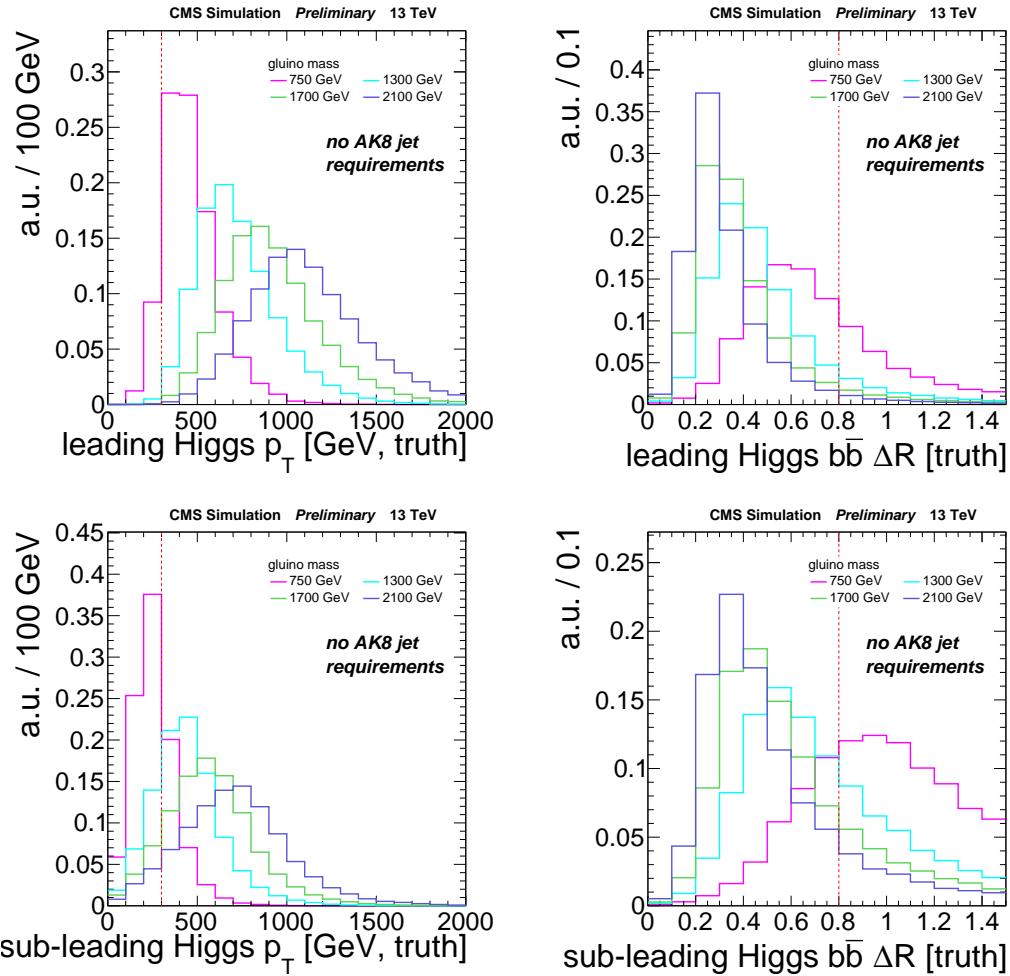


Figure 7.3: Generator level distributions for the highest (top row) and second highest (bottom row)  $p_T$  H boson in the T5HH model. The plot on the left shows the  $p_T$  of the H boson. The plot on right shows  $\Delta R$  between the b-quark daughters - for large H  $p_T$  the daughters become collimated.

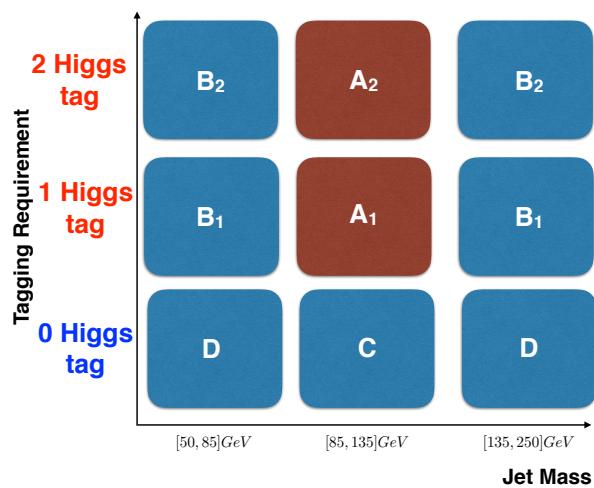


Figure 7.4: Partitioning of the signal and sideband regions for event binning. Additional binning in  $p_T^{\text{miss}}$  brings the total to 6x3=18.

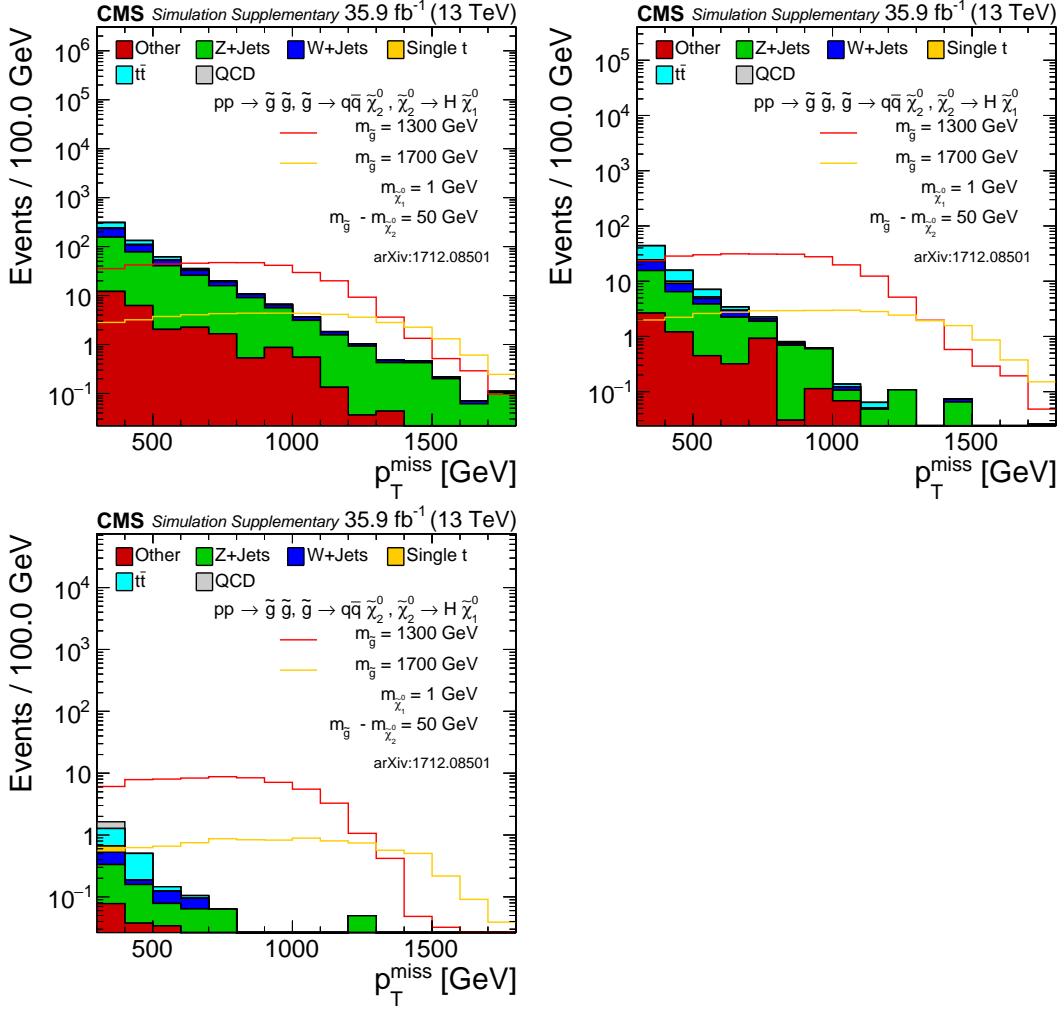


Figure 7.5: MC distributions of  $p_T^{\text{miss}}$  after baseline selection. The top right plot shows the single H-tagged region  $A_1$ , the bottom left plot shows the double H-tagged region  $A_2$ .

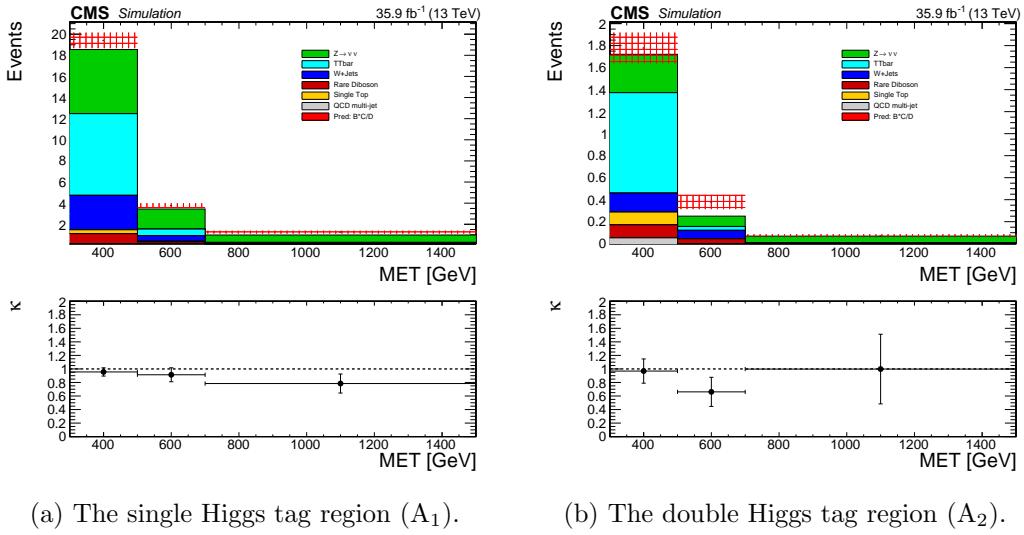


Figure 7.6:  $p_T^{\text{miss}}$  distributions and predictions in the signal regions using simulation only.

The prediction using the ABCD method on the same simulated samples is seen in the red hash. The performance of the method within simulation can be determined by dividing the true content in the signal region with the prediction. This ratio, denoted  $\kappa$ , is seen in the bottom panel of Figure 7.6.  $\kappa = 1$  represents perfect modeling. As will be discussed in Section 7.5.2,  $\kappa$  is used as a correction in the background estimation procedure.

### 7.5.1 Control Regions within Data

In order to study the expected backgrounds in data ( $Z \rightarrow \nu\bar{\nu}$ ,  $t\bar{t} \rightarrow \ell\nu qq' W \rightarrow \ell\nu$ , QCD), we define three control regions which are enhanced in processes representative to the background. The event selection is the same as applied to the nominal signal and sideband regions, but with some defining orthogonal condition:

- A control region with a single photon. For high- $p_T$ , photons and Z bosons are both massless, neutral, gauge bosons and therefore their kinematics are expected to be similar. “Artificial” removal of the photon from the reconstruction (i.e. ignore its calorimeter deposit) results in event topologies phenomenologically similar to  $Z \rightarrow \nu\bar{\nu}$ .
- A control region with a single lepton. The topology of events from direct W or top quark

production is the same regardless if the electron or muon is identified as such.

- A control region defined by the logical inversion of the low- $\Delta\phi$  cut. This explicitly selects events in which the AK4 jet momentum was likely under-measured, resulting in close alignment with  $p_T^{\text{miss}}$ .

As they are orthogonal to the analysis region, we are able to test the validity of the background estimation technique independently **within** each of the three control regions. By comparing the prediction of the yields using the ABCD method with those observed, 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 Figure 7.7.  $\kappa$  in the bottom panel is defined as the ratio of the true event yield to the prediction.  $\kappa = 1$  represents the case in which the prediction perfectly matches the observation. These comparisons are used for commissioning of the background estimation technique only.

### 7.5.2 $\kappa$ as a Correction to the Estimation

A correction factor  $\kappa$  is applied to the prediction to account for the under-prediction of the background estimation procedure as observed in Figure 7.6.  $\kappa$  is obtained by dividing the yields for the signal region by that predicted:

$$\kappa \equiv A^{MC} / \left( B \cdot \frac{C}{D} \right)^{MC} \quad (7.3)$$

There are  $2 \times 3 = 6$  values of  $\kappa$ , one for each signal bin.  $\kappa = 1$  represents the case of a perfect prediction. The corrections are then applied as follows:

$$A_{1,2}^{\text{predicted}} = \kappa \cdot \left( B_{1,2} \cdot \frac{C}{D} \right)^{\text{observed}} \quad (7.4)$$

These values of  $\kappa$  are those which we have already seen in Figure 7.6.

The value of  $\kappa$  is dependent on the yields of each analysis bin and is therefore sensitive to the accuracy of the modeling of in each of the 18 analysis bins. To improve the determination of

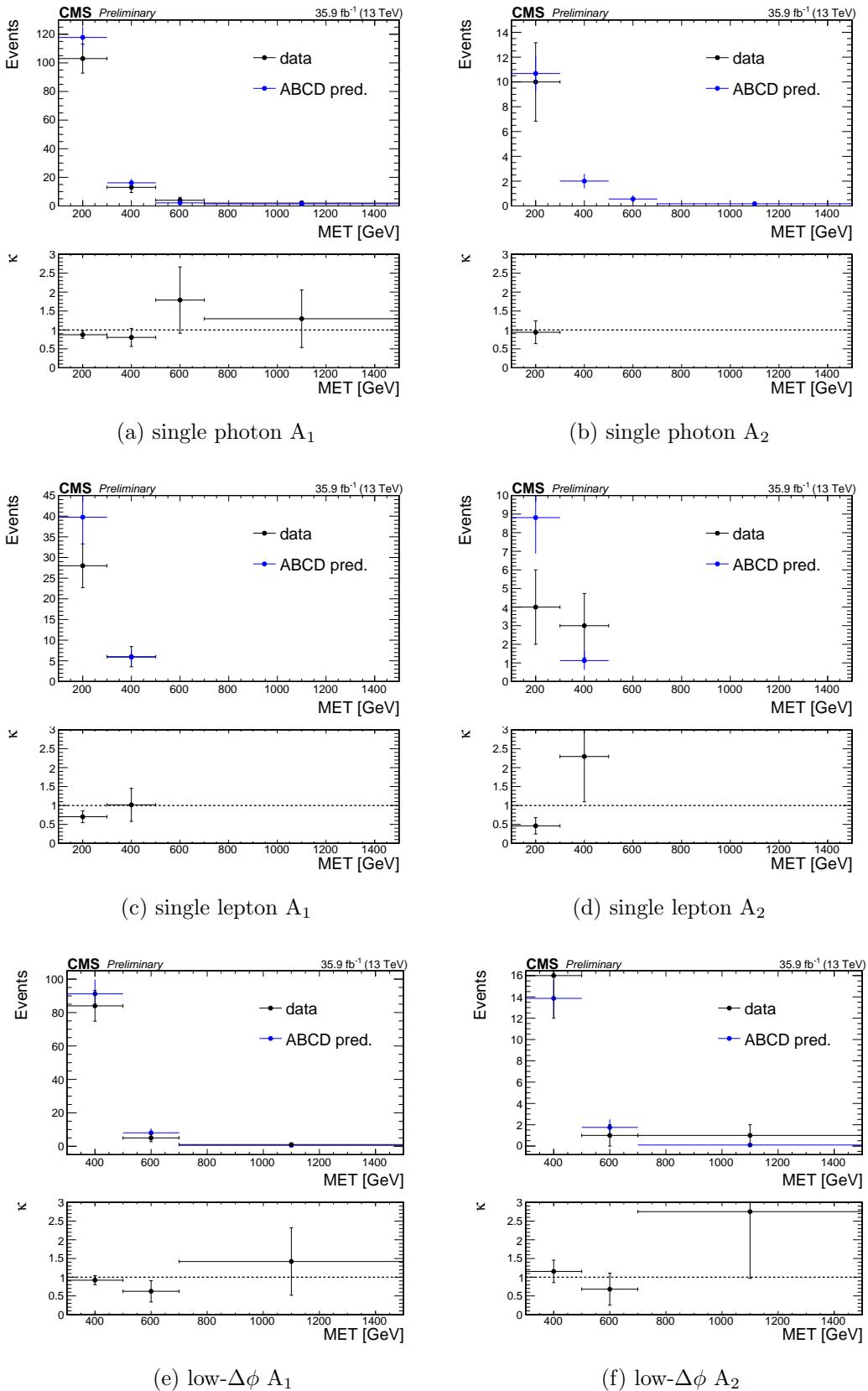


Figure 7.7: Comparisons of the predicted and observed yields within the data control regions.

$\kappa$ , scale factors are derived using the data control regions to correct the normalization of in each of these bins. Different scale factors are assigned separately to the Z-invisible, lost-lepton, and QCD background samples. Rare processes (e.g. diboson) are taken directly from MC

First consider how the yield  $N$  predicted by in an arbitrary bin (of 18) is the sum of the yields in the different datasets ( $t\bar{t}$  and  $W \rightarrow \ell\nu$  are grouped as they together represent the lost-lepton background):

$$N^{MC} = N_{Z \rightarrow \nu\bar{\nu}}^{MC} + N_{t\bar{t}, W \rightarrow \ell\nu}^{MC} + N_{QCD}^{MC} + N_{rare} \quad (7.5)$$

Scale factors are defined for this bin using the corresponding control regions in data and forming the ratio of events in simulation to that observed. They are then applied as follows:

$$N_{corrected}^{MC} = \left( \frac{N_{single-\gamma}^{data}}{N_{single-\gamma}^{MC}} \right) \cdot N_{Z \rightarrow \nu\bar{\nu}}^{MC} + \left( \frac{N_{single-\ell}^{data}}{N_{single-\ell}^{MC}} \right) \cdot N_{t\bar{t}, W \rightarrow \ell\nu}^{MC} + \left( \frac{N_{low-\Delta\phi}^{data}}{N_{low-\Delta\phi}^{MC}} \right) \cdot N_{QCD}^{MC} + N_{rare} \quad (7.6)$$

The  $p_T^{\text{miss}}$  distribution within the control regions is shown for both data and in Figures 7.8, 7.9, 7.10 for the single photon, single lepton, and low- $\Delta\phi$  control regions, respectively. The ratio in the bottom panel of each plot represents the scale factor for that  $p_T^{\text{miss}}$  bin. The dotted horizontal line shows the average scale factor inclusive in  $p_T^{\text{miss}}$ . The scale factors for the single-photon and low- $\Delta\phi$  control regions show no  $p_T^{\text{miss}}$  dependence and are determined integrated over  $p_T^{\text{miss}} > 300 \text{ GeV}$ . The values of the scale factors are summarized in Table 7.2. The Single-lepton region, shown in Figure 7.9, does show  $p_T^{\text{miss}}$  dependence and are summarized in Tables 7.2 and 7.3. In order to improve statistics for the single-lepton region, the low- $\Delta\phi$  requirement has been removed.

The scale factors are then applied to the samples to give yields which better reflect data. The  $p_T^{\text{miss}}$  distributions for the signal regions and expectations from the ABCD background prediction are seen in Figure 7.11 (the data-corrected version of Figure 7.6). The improved value of  $\kappa$  is seen in the lower panel of each plot. The modified values of the yields in the signal region (seen in the calculation of  $\kappa$ ) are seen in Tables 7.6 and 7.6. Since most of the scale factors are less than one the background decreases in Figure 7.6 relative to Figure 7.11 but still preserves the normalization

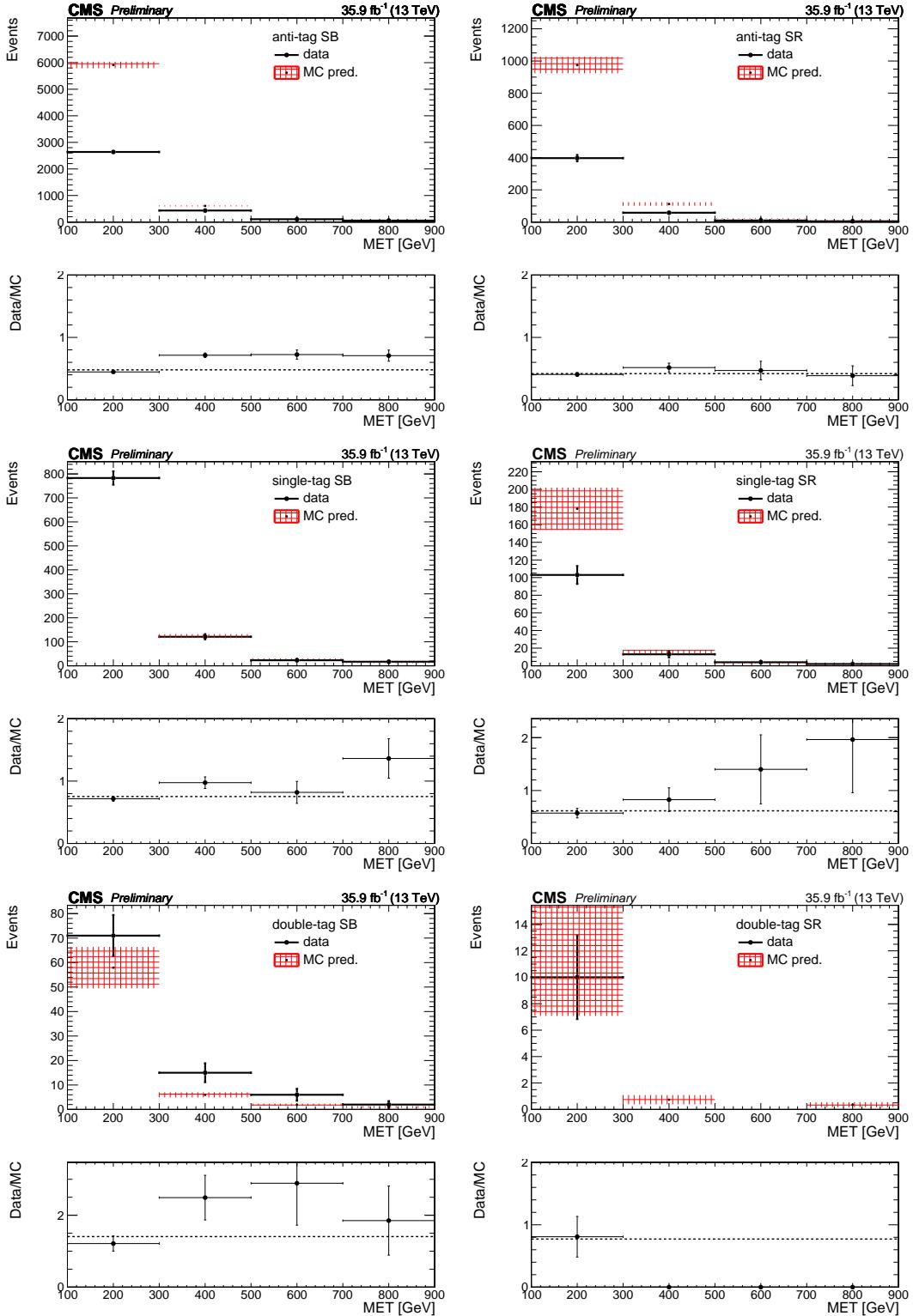


Figure 7.8: Signal and sideband yields in the single photon control region. The hashed red band denotes the prediction from simulation; the solid black points denote the observed yields in data. The Data/ ratio in the lower panel of each plot represents the scale factor for that bin.

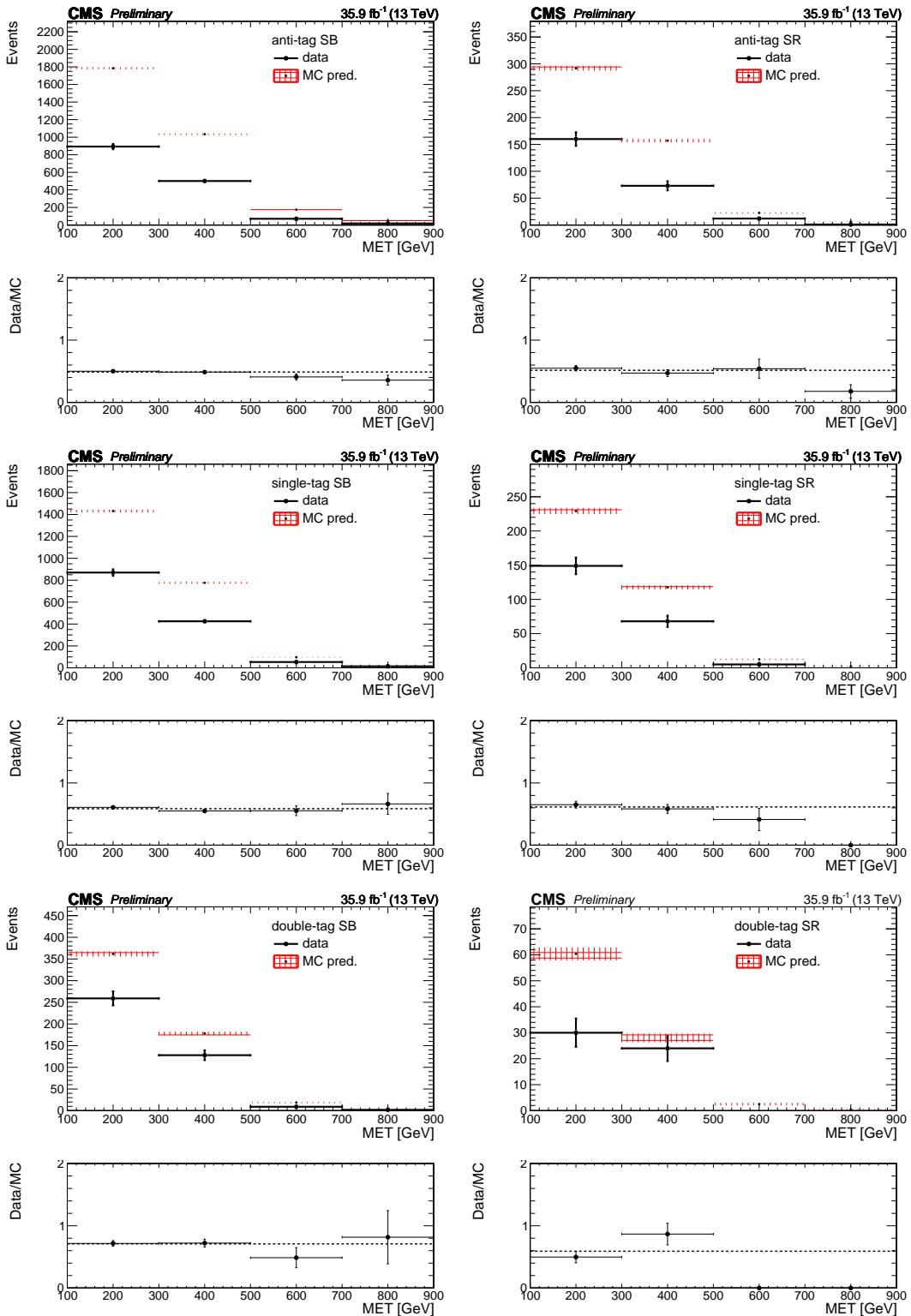


Figure 7.9: Signal and sideband yields in the single lepton control region. The hashed red band denotes the prediction from simulation; the solid black points denote the observed yields in data. The Data/ ratio in the lower panel of each plot represents the scale factor for that bin. The low- $\Delta\phi$  requirement has been removed to improve statistics.

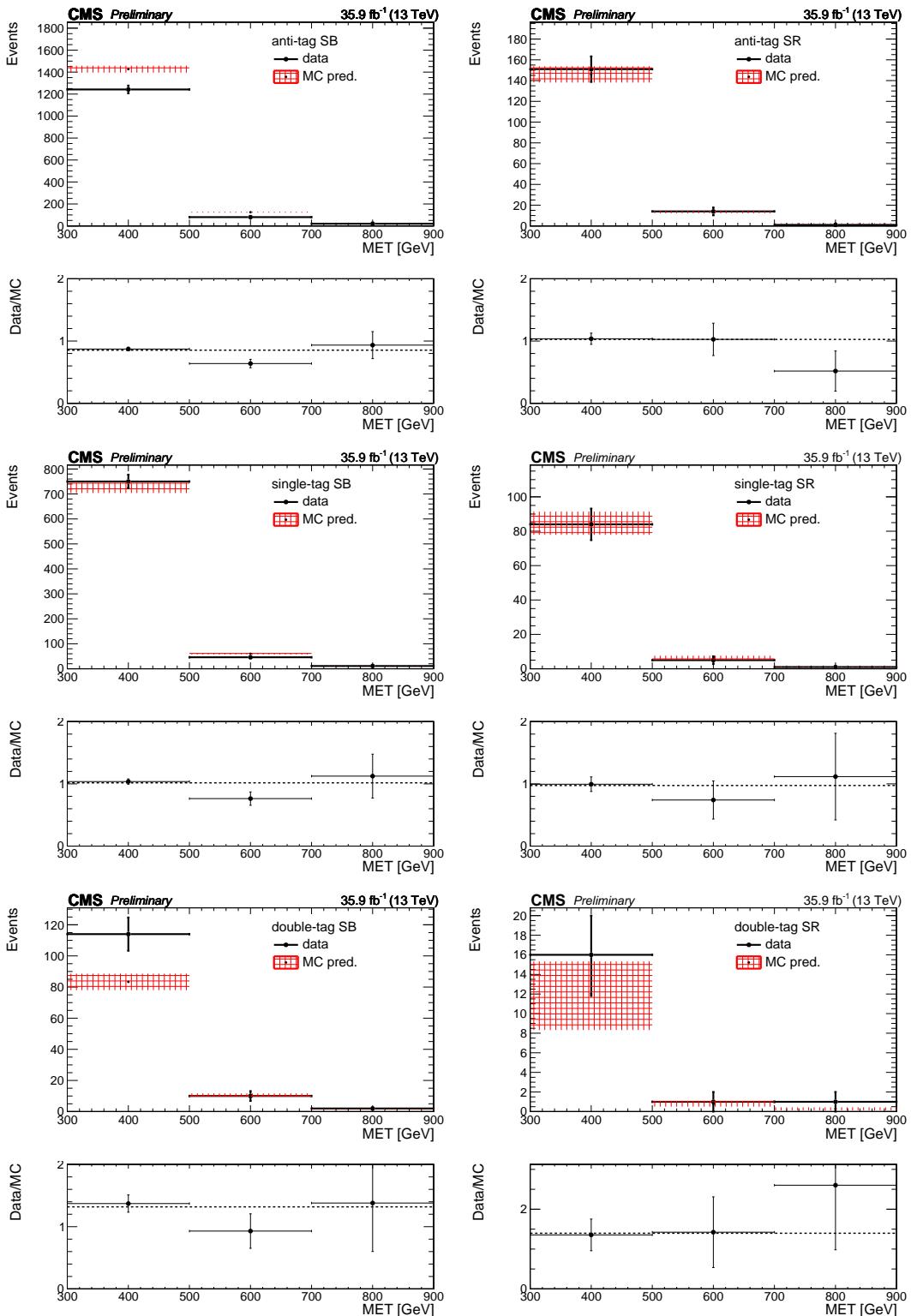


Figure 7.10: Signal and sideband yields in the the low- $\Delta\phi$  control region. The hashed red band denotes the prediction from simulation; the solid black points denote the observed yields in data. The Data/ ratio in the lower panel of each plot represents the scale factor for that bin.

Table 7.2: Summary of the control region scale-factors integrated over  $p_T^{\text{miss}}$ .

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					
$A_{SF}^{1H}$	$A_{SF}^{2H}$	$C_{SF}$	$B_{SF}^1$	$B_{SF}^2$	$D_{SF}$
$0.61 \pm 0.04$	$0.59 \pm 0.08$	$p_T^{\text{miss}}$ dependent	$0.59 \pm 0.016$	$0.71 \pm 0.04$	$p_T^{\text{miss}}$ dependent
Photon					
$A_{SF}^{1H}$	$A_{SF}^{2H}$	$C_{SF}$	$B_{SF}^1$	$B_{SF}^2$	$D_{SF}$
$0.61 \pm 0.088$	$0.75 \pm 0.29$	$0.5 \pm 0.07$	$0.98 \pm 0.094$	$2.58 \pm 0.63$	$0.71 \pm 0.035$

Table 7.3: Single lepton control region scale-factors in the anti-tag sideband region.

Single Lepton $C_{SF}$		
$p_T^{\text{miss}}$ [300, 500]	[500, 700]	[700, $\infty$ ]
$0.47 \pm 0.05$	$0.54 \pm 0.15$	$0.18 \pm 0.1$
Single Lepton $D_{SF}$		
$0.49 \pm 0.02$	$0.40 \pm 0.05$	$0.35 \pm 0.08$

Table 7.4: The  $\kappa$  factor computed by throwing Gaussian toys for the scale factors.

	1-Higgs Tag	2-Higgs Tag
$p_T^{\text{miss}}$	$\kappa$	
[300, 500 GeV]	$0.98 \pm 0.11$	$0.73 \pm 0.14$
[500, 700 GeV]	$0.86 \pm 0.16$	$0.43 \pm 0.12$
[700, $\infty$ GeV]	$0.86 \pm 0.17$	$0.62 \pm 0.30$

so that  $\kappa$  is statistically compatible with unity. A distribution of  $\kappa$  is derived by throwing gaussian toys for each of the scale factors, the final results being summarized in Table 7.4.

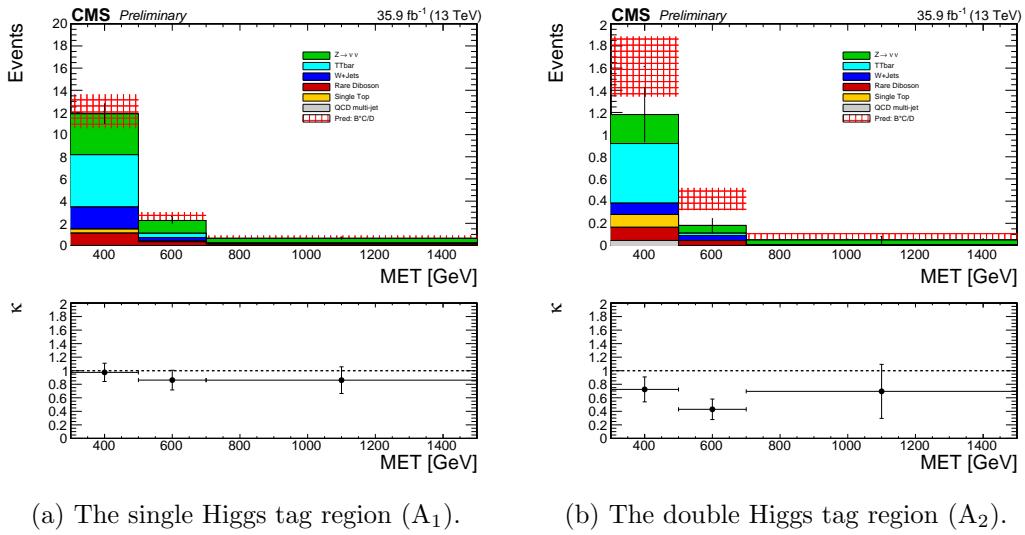


Figure 7.11: Signal region  $p_T^{\text{miss}}$  distributions using scale-factor corrected simulation.

Table 7.5: Corrected yields in the signal regions.  $\kappa = \text{AD} / \text{BC}$ 

$p_T^{\text{miss}}$	Z+Jets	W+Jets	TTbar	QCD	Rare	Total	Data
Region: $A^{1H}$							
[300, 500] GeV	3.76 ± 0.54	2.05 ± 0.30	4.87 ± 0.72	0 ± 0	1.48 ± 0.40	12.17 ± 1.03	15
[500, 700] GeV	1.18 ± 0.21	0.33 ± 0.08	0.40 ± 0.12	0 ± 0	0.39 ± 0.16	2.30 ± 0.29	2
> 700 GeV	0.43 ± 0.10	0.053 ± 0.025	0.046 ± 0.01	0 ± 0	0.13 ± 0.053	0.66 ± 0.12	1
Region: $A^{2H}$							
[300, 500] GeV	0.26 ± 0.12	0.11 ± 0.05	0.58 ± 0.21	0.045 ± 0.05	0.23 ± 0.12	1.24 ± 0.28	1
[500, 700] GeV	0.07 ± 0.045	0.049 ± 0.031	0.022 ± 0.0098	0 ± 0	0.045 ± 0.039	0.19 ± 0.068	0
> 700 GeV	0.044 ± 0.032	0.005 ± 0.005	0 ± 0	0 ± 0	0.002 ± 0.016	0.051 ± 0.036	0

Table 7.6: Corrected yields in the sideband regions.  $\kappa = \text{AD} / \text{BC}$ 

$p_T^{\text{miss}}$	Z+Jets	W+Jets	TTbar	QCD	Rare	Total	Data
Region: $C$							
[300, 500] GeV	17.66 ± 2.52	8.23 ± 1.44	3.87 ± 0.73	0.81 ± 0.49	2.78 ± 1.142	33.36 ± 3.24	44
[500, 700] GeV	5.20 ± 0.77	0.57 ± 0.27	0.22 ± 0.11	0 ± 0	0.63 ± 0.16	6.63 ± 0.84	12
> 700 GeV	2.48 ± 0.39	0.12 ± 0.12	0.028 ± 0.031	0 ± 0	0.14 ± 0.06	2.76 ± 0.41	4
Region: $B^{1H}$							
[300, 500] GeV	42.13 ± 4.17	15.61 ± 10.15	30.99 ± 20.15	1.57 ± 0.54	12.16 ± 1.37	102.47 ± 23.00	112
[500, 700] GeV	12.05 ± 1.28	2.74 ± 1.79	3.04 ± 2.00	0 ± 0	2.55 ± 0.43	20.37 ± 3.00	20
> 700 GeV	5.92 ± 0.69	0.67 ± 0.61	0.49 ± 0.46	0 ± 0	1.93 ± 0.72	9.01 ± 1.25	5
Region: $B^{2H}$							
[300, 500] GeV	5.51 ± 1.47	0.73 ± 0.44	4.46 ± 2.65	0.33 ± 0.23	2.06 ± 0.32	13.09 ± 3.09	13
[500, 700] GeV	1.80 ± 0.56	0.17 ± 0.11	0.59 ± 0.39	0 ± 0	0.62 ± 0.23	3.17 ± 0.73	1
> 700 GeV	0.67 ± 0.27	0.0084 ± 0.009	0.035 ± 0.031	0 ± 0	0.23 ± 0.073	0.94 ± 0.28	1
Region: $D$							
[300, 500] GeV	164.82 ± 8.31	61.24 ± 3.70	33.20 ± 2.16	8.50 ± 2.73	20.64 ± 1.78	288.41 ± 9.90	273
[500, 700] GeV	47.37 ± 2.52	6.36 ± 1.39	2.37 ± 0.55	0 ± 0	4.42 ± 1.46	60.51 ± 3.27	60
> 700 GeV	26.79 ± 1.50	0.99 ± 0.53	0.16 ± 0.086	0 ± 0	3.48 ± 1.01	31.42 ± 1.88	28

Table 7.7: Sideband region yields,  $\kappa$ , and background predictions for the 6 signal bins.

$N_H$	$p_T^{\text{miss}}$ (GeV)	B	C	D	$\kappa$	$\kappa \cdot B \cdot C/D$
$A_1$	[300, 500 GeV]	112	44	273	$0.98 \pm 0.11$	$17.7 \pm 3.8$
$A_1$	[500, 700 GeV]	20	12	60	$0.86 \pm 0.16$	$3.4 \pm 1.5$
$A_1$	$[700, \infty \text{ GeV}]$	5	4	28	$0.86 \pm 0.17$	$0.61 \pm 0.45$
$A_2$	[300, 500 GeV]	13	44	273	$0.73 \pm 0.14$	$1.52 \pm 0.57$
$A_2$	[500, 700 GeV]	1	12	60	$0.43 \pm 0.12$	$0.09 \pm 0.08$
$A_2$	$[700, \infty \text{ GeV}]$	1	4	28	$0.62 \pm 0.30$	$0.09^{+0.11}_{-0.09}$

### 7.5.3 Sideband Yields & Predictions

We have studied the background estimation technique in data and found the results to be acceptable. We then calculated  $\kappa$ , the correction to the ABCD prediction, using scale factors from data to best correct the normalization of the MC components. We can now use the observed yields in the sideband regions to form the background prediction via the ABCD method. Figure 7.12 shows the sideband yields in both data and the scale factor corrected MC - we see that the two agree within statistical errors. Table 7.7 tabulates the data yields alongside the full background prediction including  $\kappa$ .

## 7.6 Signal Systematics

We consider a variety of systematic uncertainties on the signal efficiency and distribution. Some are common to more inclusive SUSY analyses [35] and there are additional systematics related to  $b\bar{b}$  tagging efficiency and the effect of the pruned mass scale and resolution on the signal efficiency.

- **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/ differences based on the study from the 2015 analysis [35].
- **statistics:** The sign fal sample statistical uncertainty is generally 2-4% .
- **Trigger efficiency:** The effect of the uncertainty on the signal yield is about 2%.

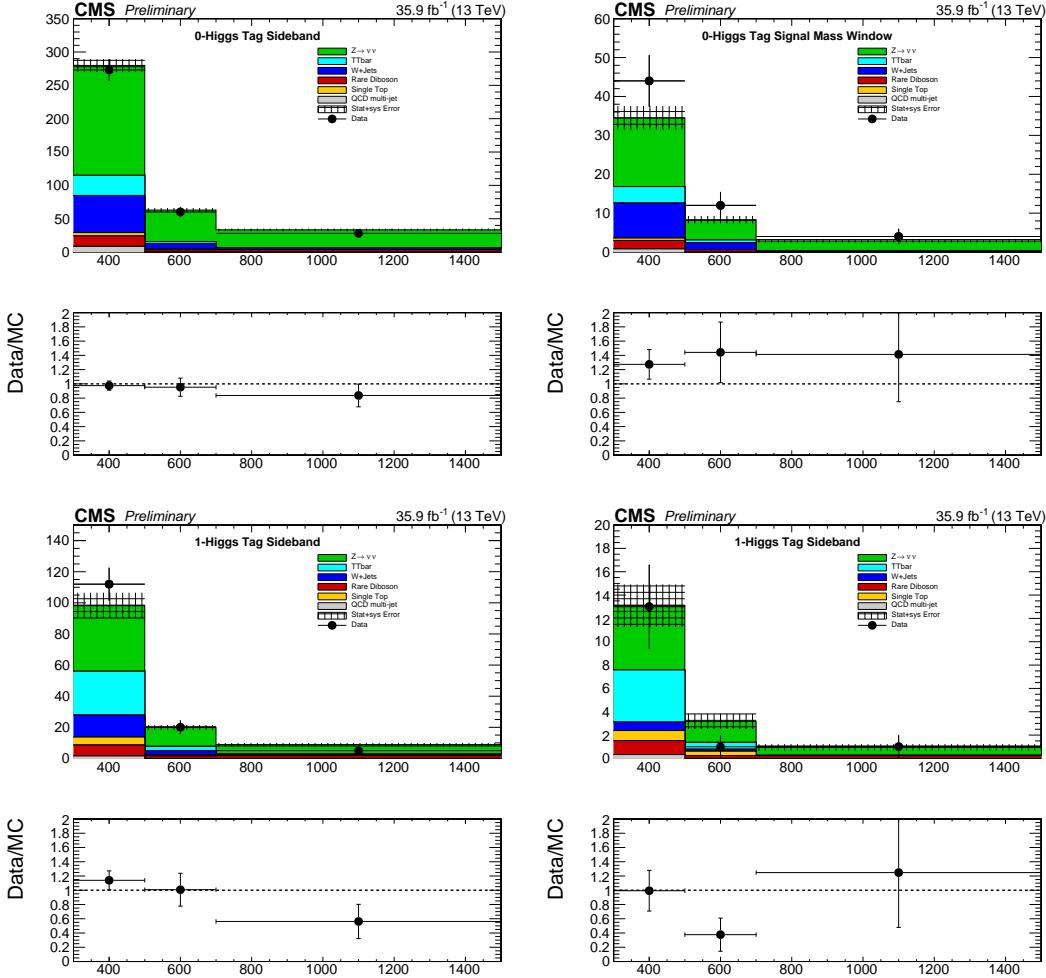


Figure 7.12: Sideband region  $p_T^{\text{miss}}$  distributions comparing data and scale-factor corrected simulation. The hashed red distribution denote the prediction from simulation; the solid points denote the observed yields in data.

- **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 samples and no associated uncertainty is assessed.
- **ISR:** An ISR correction is derived from  $t\bar{t}$  events, with a selection requiring two lepton-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  $\text{NISR} = 0, 1, 2, 3, 4, 5, 6+$ . 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  $p_T^{\text{miss}}$ .
- **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 [36, 37]. The effect on the yield 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:  $H_T$ , MET,  $\Delta\phi(\text{MET}, j_i)$ . The overall effect is less than 1%.
- **Jet Energy Resolution:** The jet momenta in the 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:  $H_T$ , MET,  $\Delta\phi(\text{MET}, j_i)$ . The overall effect ranges from 0.01%.

- **PDFs:** The LHC4PDF prescription for the uncertainty on the total cross section is included as  $\pm 1$  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/ 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 propagated to the final signal efficiency using POG recommendations. The pruned mass scale factor is derived using W-jets in semi-leptonic  $t\bar{t}$  and extrapolating to the H mass. This uncertainty is assigned a shape uncertainty on the signal mass window and the sideband.

- A data/ 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 within a mass window of [65, 85] GeV. The fit for the gaussian resolution of the W-mass peak is shown in Figure 7.13 and the fit results are shown in Table 7.8. The mass scale between and data is consistent though 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 [38].

Table 7.8: Fit results for W-mass resolution in data and simulation.

	Data	$t\bar{t}$
Mean	$78.2 \pm 0.46$	$78.4 \pm 0.35$
Sigma	$10.10 \pm 0.67$	$7.23 \pm 0.48$

The summary of the signal systematics and their effect on the signal yields is shown in Table 7.9. The dominant effect is from the mass resolution uncertainty.

## 7.7 Observed Yields in the Signal Regions

The observed yields, along with the background predictions, are seen in Table 7.10. Our signal region yields are consistent with the background expectation. Additionally, Table 7.10 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.

A visual representation of the one event in the double-H tagged signal bin is seen in Figure 7.15. The purple line represents  $p_T^{\text{miss}} = 426$  GeV. The three yellow cones represent the AK8 jets labeled with  $p_T$ . Note the two additional objects not satisfying our object definition but still plotted in the representation a) the additional low- $p_T$  and low mass AK8 jet b) the  $p_T=18$  GeV muon (red line) suffers from poor reconstruction properties.

## 7.8 Exclusion Curves & Mass Limits

Interpreting our results in the context of the T5HH or T5ZH models, the absence of signal allows us to place lower limits on the cross section of gluino  $\tilde{g}$  pair production. For the statistical treatment, we use a profile-likelihood to combine the observed yields, expected background, and expected signal in all analysis bins to calculate a 95% upper limit on the signal cross section. The ABCD background estimation is explicitly coded in the likelihood, allowing the correlations in the

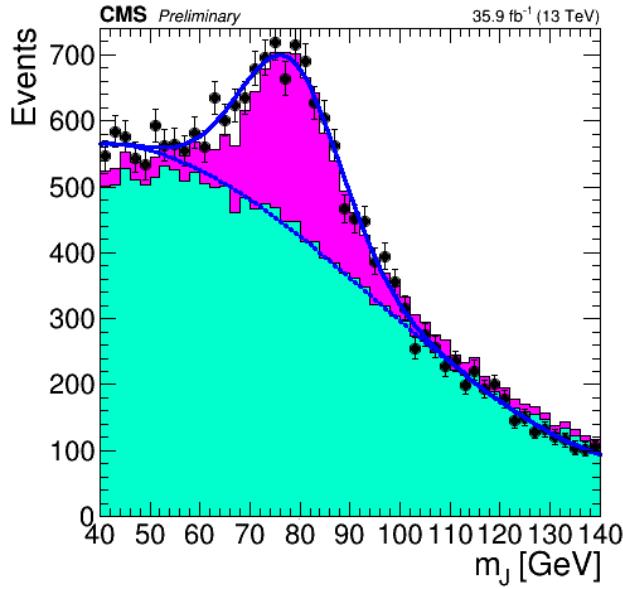


Figure 7.13: Pruned jet mass in semi-leptonic  $t\bar{t}$  events. The mass peak for the W-jets is used to derive the mass resolution uncertainty.

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%
Stat	1-4%
Shape Unc.	
Double-b SF	6%
Mass Resolution	1-15%

Table 7.9: Summary of signal shape and normalization uncertainties.

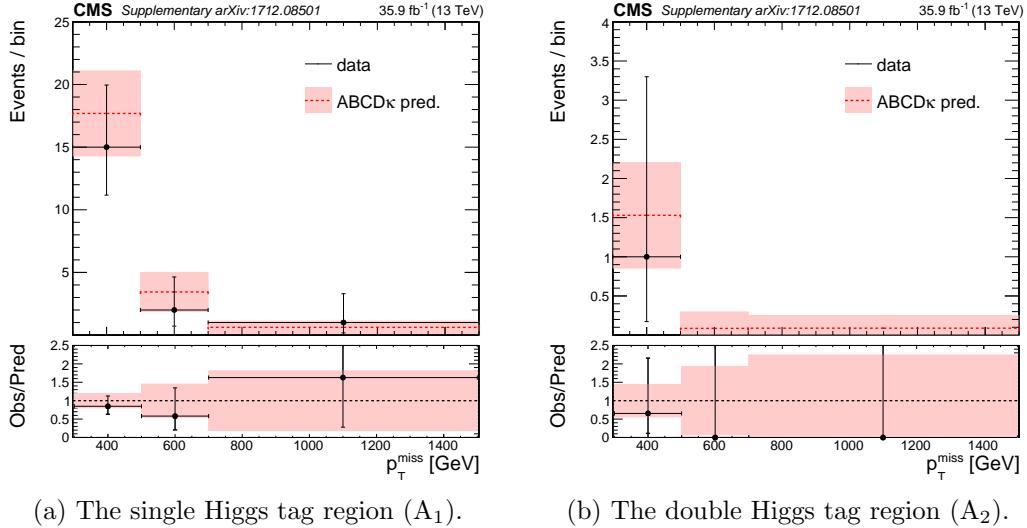


Figure 7.14: Observed yields in the signal regions.

Table 7.10: Yields and predicted background in the signal regions. Columns 2, 3, and 4 form the background prediction. Obs. represents the observed yields in data. The last two columns represent the expected yields from two model points (the gluino mass is in parenthesis).

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

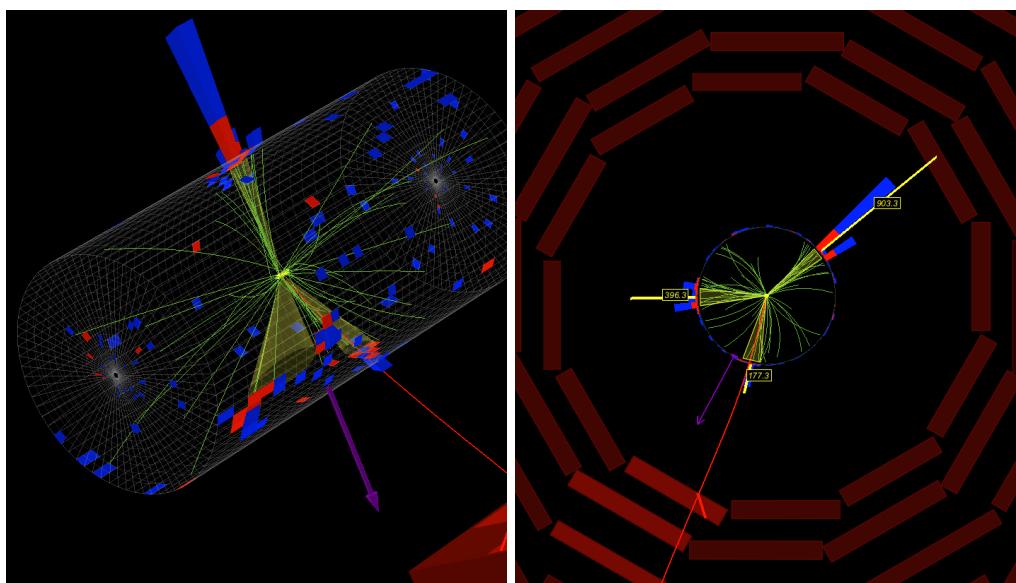


Figure 7.15: The single event in the  $A_2$  region.

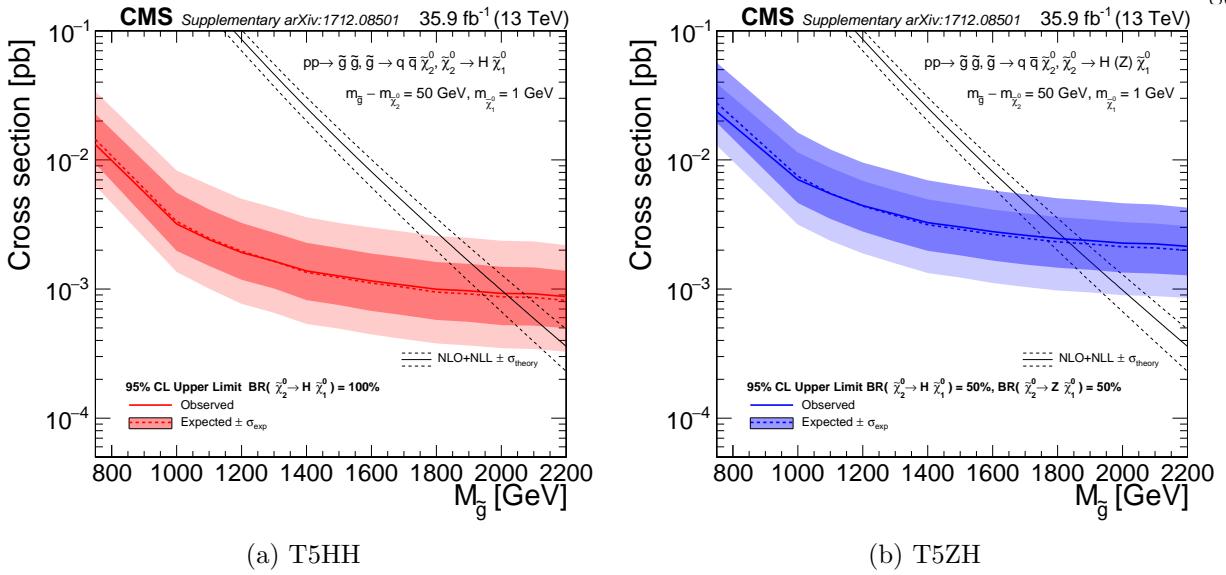


Figure 7.16: Observed and expected limits on the gluino cross section.

estimations to be taken into account. The likelihood function can be expressed as:

$$\mathcal{L} = \prod_{i \in A, B, C, D} \text{Poisson}(n_i | bkg_i + r \cdot sig_i) \times \prod_j^{nuisances} \text{Constraints}(\theta_j, \hat{\theta}_j) \quad (7.7)$$

where the regions are modeled by Poisson distributions,  $n_i$ ,  $bkg_i$ , and  $sig_i$  represent the yield, expected background, and expected signal in bin i. The constraints term encode the nuisance parameters  $\theta_j$ :  $\kappa$  is modeled by a normal distribution, and the signal systematics are modeled as log-normal. The term  $\hat{\theta}_j$  represents the profiled minima of  $\theta_j$ .

The expected and observed limits are then calculated based on approximations of the profile likelihood ratio using the CLs criterion to place upper limits at the 95% confidence level on the production cross section, as seen in Figure 7.16. The allowed regions lie below the red and blue lines. Calculations of the gluino pair production cross section [1], seen in the black line, allow us to rule out gluinos with mass lower than 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.

## Chapter 8

### Conclusions

This thesis presents a search for evidence of physics beyond the SM. The search targets events with two or more wide-angle jets (AK8) being consistent with boosted H or Z bosons decaying to  $b\bar{b}$ . The analysis makes use a dedicated heavy-flavor tagging algorithm used to identify large angle jets composed of two b quarks. We additionally require a large amount of  $p_T^{\text{miss}}$  potentially arising from a supersymmetric particle escaping detection. A data-driven method uses yields in 12 sideband bins to predict the background in 6 signal bins. Events are categorized according to the  $b\bar{b}$ -tagging and masses of the highest two  $p_T$  AK8 jets, and the amount of  $p_T^{\text{miss}}$ . Backgrounds arise from QCD jet production, jets+ $Z \rightarrow \nu\bar{\nu}$ , jets+ $W \rightarrow \ell\nu$ , and  $t\bar{t} \rightarrow qq'\ell\nu$ , in which the leptons are not identified. These backgrounds are studied by defining control regions expected to be enhanced in each of these processes. In combination with simulation, these control regions are additionally used to derive corrections to the background prediction.

The observed yields in the signal bins are statistically compatible with the background expectation and no excess of events is observed. A maximum-likelihood fit allows us to set limits on the cross section of gluino-pair production under the assumption of the T5ZH or T5HH decay models. Comparisons of our limits with the theoretical cross sections allow us to exclude gluinos with masses below 2010/1825 GeV at 95% confidence level for the T5HH/T5ZH models. These limits are made with the assumption that the NLSP mass is 50 GeV less than the gluino mass, and that the LSP has a mass of 1 GeV.

The work presented here has been published in Physical Review Letters in June 2018 [39].

## Appendix A

### b $\bar{b}$ -tagging of AK8 Jets

A novel approach has been studied to identify large-radius jets composed of two b-quarks [40]. A dedicated multivariate (MVA) tagging algorithm is implemented to combine the information from the secondary vertices, the tracks, and the  $\tau$ -axes of the jet and optimize the discrimination between jets from two b-hadron decays and single-parton jets. To generalize the algorithm so that it does not depend on the mass of the b-parents or the boost, the input variables are chosen so that they do not depend strongly on jet mass or jet  $p_T$ . All tracks with  $p_T > 1$  GeV are associated with jets in a cone  $\Delta R < 0.8$  around the jet axis, which is defined by the primary vertex and the direction of the jet momentum. The tracks are then associated with the closest  $\tau$ -axes. For this we take the distance between the sub-structure axis and the track to be the distance of closest approach, which is required to be within  $700\mu m$  and the point of closest approach on the track is required to be 5 cm from the primary vertex. The impact parameter significance (with respect to the primary vertex) is used to discriminate tracks from the b decay from prompt tracks. Several input variables make use of the secondary vertices that are reconstructed using the Inclusive Vertex Finder [41], which identifies secondary vertices independently of the jet clustering. The secondary vertices are then associated with the closest  $\tau$ -axis. For each  $\tau$ -axis, track momenta of the constituent tracks from all the secondary vertices associated with a given  $\tau$ -axis are added to compute the SV mass and transverse momentum. The input variables to the double-b tagger MVA discriminant are detailed in [40]. Figure A.1 shows the discrimination between signal  $H \rightarrow b\bar{b}$  and background jets based on the true number of b-hadrons.

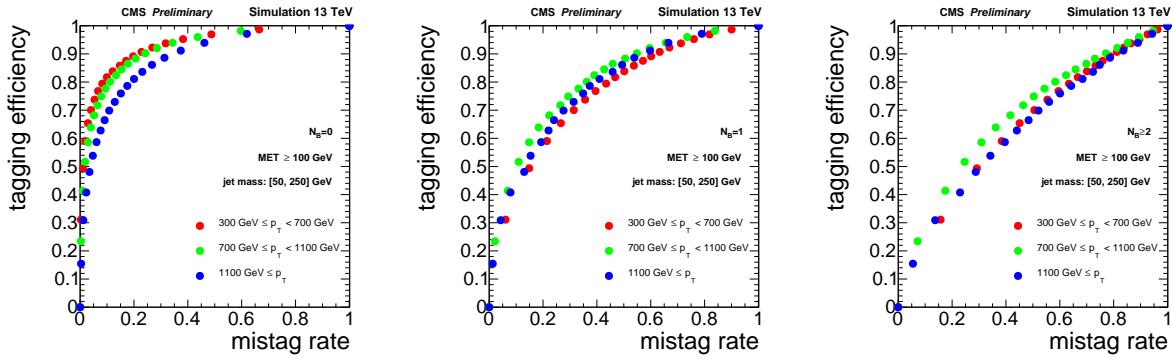


Figure A.1: ROC curve of the signal efficiency and mistag rate for the  $b\bar{b}$  tagger. The mistag rate is calculated using all expected SM backgrounds.

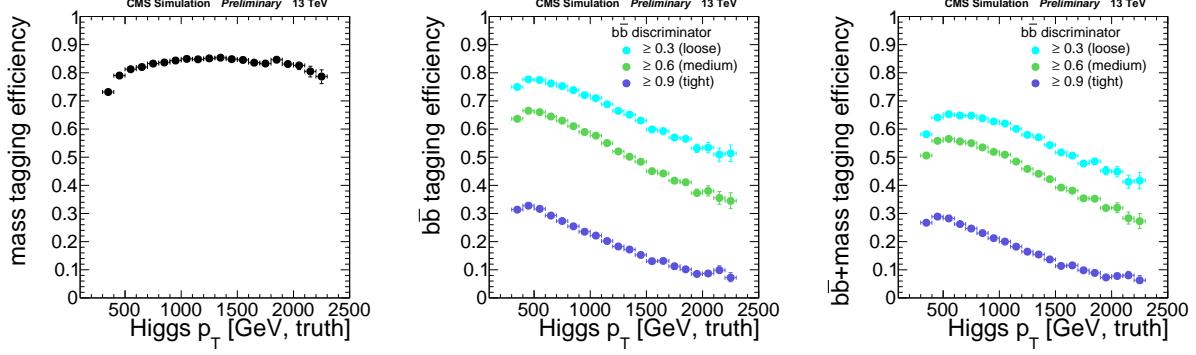


Figure A.2: The efficiency of a  $H \rightarrow b\bar{b}$  jet to have mass [85, 135 GeV], as a function of the generator-level  $p_T$  (left). The efficiency of a  $H \rightarrow b\bar{b}$  jet to pass the double-b tagging requirement, for three different working points (center). The efficiency of a  $H \rightarrow b\bar{b}$  jet to pass both the tight mass and double-b tagging requirements (right).

Figure A.2 shows the signal efficiency for an  $H \rightarrow b\bar{b}$  jet to pass the Double-b tag cut. The efficiencies are computed by matching reconstructed jets to a generator-level Higgs or Z boson based on an angular requirement,  $\Delta R < 0.8$ , between the jet and the appropriate gen-particle and has one or more B-hadrons associated with it via: `Jet→jetFlavourInfo().getbHadrons().size()`. The efficiency is above 80% for the Double-b discriminator alone, and drops to 65% when the pruned mass cut is applied. At high jet  $p_T$  there is an inefficiency in the Double-b tag as tracks from the b quarks become more collimated. The QCD background consists of some mistagged light flavor jets and jets with true b-hadrons which can come from gluon splitting or flavor excitation. The  $t\bar{t}$  background has both true heavy flavor jets and true  $p_T^{\text{miss}}$ , but a good number of the single b-hadron jets can be rejected with the Double-b tag requirement.

Figure A.3 shows the signal efficiency for  $H(b\bar{b})H(b\bar{b})$  and  $H(b\bar{b})Z(q\bar{q})$  signal models as a function of the double-b discriminator value. For the model with  $H(b\bar{b})H(b\bar{b})$ , the efficiency is plotted against the smaller of the double-b discriminators of the leading and sub-leading jet, since both are expected to have two b-quarks. For the model with  $H(b\bar{b})Z(q\bar{q})$ , only one jet is expected to have two b-quarks, so the efficiency is plotted against the larger double-b discriminator value between the leading and sub-leading jet. We use the loose working point of the double-b cut at 0.3

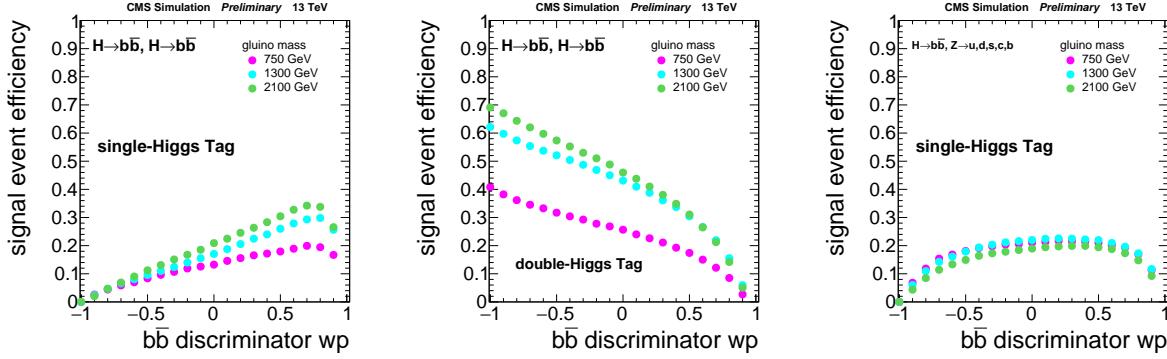


Figure A.3: Signal efficiency to be in the single or double Higgs tag event category, as a function of the double-b discriminator working-point. Efficiencies are relative to baseline selection. The left and center plots show the single and double Higgs tag efficiencies for the Higgs( $b\bar{b}$ )-Higgs( $b\bar{b}$ ) MC. The right plot shows the single Higgs tag efficiency for the Z( $q\bar{q}$ )-Higgs( $b\bar{b}$ ) MC. The curves for three representative gluino masses are shown here.

to consider a jet to be double tagged.

The efficiency of the double-b tagger is measured in a data sample consisting of high  $p_T$  jets enriched in  $b\bar{b}$  from gluon-splitting. To select a boosted topology similar to the signal, the AK8 jet  $p_T$  is required to be above 300 GeV and the AK8 pruned mass is required to be above 50 GeV. The jet must also be matched to at least two muons with  $p_T > 7\text{GeV}$  and within  $\Delta R < 0.4$  of the sub-jet axis so that the jet can be “double-muon tagged”. The difference between data and MC is compared to give a data/MC scale factor which is fairly close to unity. The mis-tag rate is evaluated by comparing data and MC for top-quark jets faking H jets in  $t\bar{t}$  production. The studies are based on single lepton  $t\bar{t}$  events and the event selection requires one isolated muon with  $p_t > 50\text{ GeV}$  and an AK4 jet in the same hemisphere. More details can be found in [40]. Table ?? lists the data/MC scale-factors that correct for the efficiency difference between data and simulation based on the jet  $p_T$ . The final signal efficiency are scaled based on these scale-factors and the double-b efficiency systematic of the signal is based on the uncertainty of these scale-factors. For  $p_T$  values above the limit of 840 GeV, the scale factor at this limit is used with twice the uncertainty for the signal systematic.

Table A.1: Data/MC scale factors for AK8 jet  $b\bar{b}$ -tagging with the loose working point  $\mathcal{D}_{bb} > 0.3$ .

$p_T$ [GeV]	Signal SF
250-350	$0.94 \pm 0.03/0.02$
350-430	$1.00 \pm 0.04/0.03$
430-840	$1.01 \pm 0.02/0.04$

## Appendix B

### Reinterpretation

In Section 7.7, we presented our observed yields in the signal region and found they were consistent with the background expectations. Armed with this information, we were able to set upper limits on the cross sections of the T5HH and T5ZH SMS models (as a function of the gluino mass). As these two models are certainly not the only possibility for new physics to arise, it is important that we provide additional information necessary to cast other models within the context of our analysis. This information allows a model builder to make predictions of the yields for their models of similar construction. They can compare these yields to the ones we observed in our analysis and see if their model has potential for further investigation or would already be ruled out.

We present the efficiencies for mass-tagging and  $b\bar{b}$ -tagging of Higgs and Z bosons reconstructed as AK8 jets. Tagging efficiencies for the five largest decay modes relevant to the analysis for the H boson are seen in B.2. Tagging efficiencies for the hadronic decay modes of the Z boson are seen in Figure B.3, the much lower mass tagging efficiency for the Z boson is due to our choice of signal mass window [85, 135 GeV] not being optimal for Z reconstruction.

In the analysis, each event is categorized depending on the mass and  $b\bar{b}$ -tagging of the two AK8 jets. Among these combinations, three are related to  $b\bar{b}$ -tagging: they can both be tagged, one-and-only-one can be tagged, or neither of them can be tagged. There are two relevant combinations for the mass-tagging among the jets: they are both within the signal mass window [85, 135 GeV], or not. On an event-by-event basis, given the  $p_T$  of the Higgs or Z bosons from the model, we can use the provided efficiency maps to form event weights for each of the 6 analysis bins for the event.

Table B.1: Effective event weights for  $b\bar{b}$  and mass tagging of AK8 jets.

both jets are bb-tagged	$W_{2b} = j_{0,bb} \cdot j_{1,bb}$
only one jet is bb-tagged	$W_{1b} = j_{0,bb} \cdot (1 - j_{1,bb}) + (1 - j_{0,bb}) \cdot j_{1,bb}$
neither jet is bb-tagged	$W_{0b} = (1 - j_{0,bb}) \cdot (1 - j_{1,bb})$
both jets are mass-tagged	$W_m = j_{0,m-sig} \cdot j_{1,m-sig}$
! both jets are mass-tagged	$W_{!m} = (j_{0,m-side} \cdot j_{1,m-sig}) + (j_{1,m-sig} \cdot j_{1,m-side}) + (j_{0,m-side} \cdot j_{1,m-side})$

As these weights are dependent on the decay mode of the boson, we additionally need to make sure to choose the correct map. The total expected yields in each bin are then obtained by summing these weights over the events in the sample.

**For each event**, we form “primitive weights” which combine the status of the mass and  $b\bar{b}$ -tagging, these are seen in Table B.1.  $j_{0,bb}$  represents the probability the leading AK8 jet is  $b\bar{b}$ -tagged,  $j_{0,m-sig}$  represents the probability the leading AK8 jet falls within the signal mass window [85, 135 GeV], and  $j_{0,m-side}$  represents the probability the leading AK8 jet falls within the sideband mass window [50, 85]+[135, 250 GeV].

Now that these weights are formed we can combine them appropriately to make event weights for the individual analysis bins. The weights for a given bin are then summed over the total number of events in the sample. The weights for each analysis bin are shown in Figure B.1.

As a test of closure, the prescription is tested using the T5HH model with a gluino mass of 2200 GeV. The yields in each of the 6 analysis bins for this model point (inclusive in  $p_T^{\text{miss}}$ ) are listed in Table B.2 under the heading “RECO truth”. As the efficiency maps are  $p_T$  dependent, we have the choice of using the momentum of the reconstructed or the generated bosons. The “RECO prediction” column in the table uses the reconstructed momentum as input, “GEN prediction” uses the generated momentum. The largest deficit is in the D region, with a difference of -36% difference from nominal. 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.

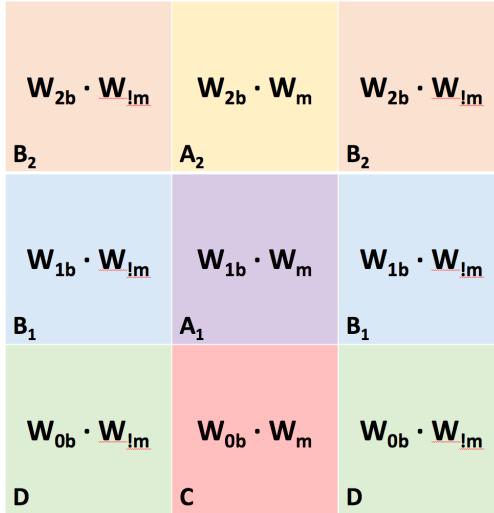


Figure B.1: Diagram for compiling the weights into each of the 6 analysis bins. (See Figure 7.4)

Table B.2: Comparison of the event yields in the 6 analysis regions with those obtained via the prediction. The prediction was made using the T5HH signal MC with a gluino mass of 2200 GeV. Event yields are scaled to a luminosity of  $35.9 \text{ fb}^{-1}$ .

	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%)

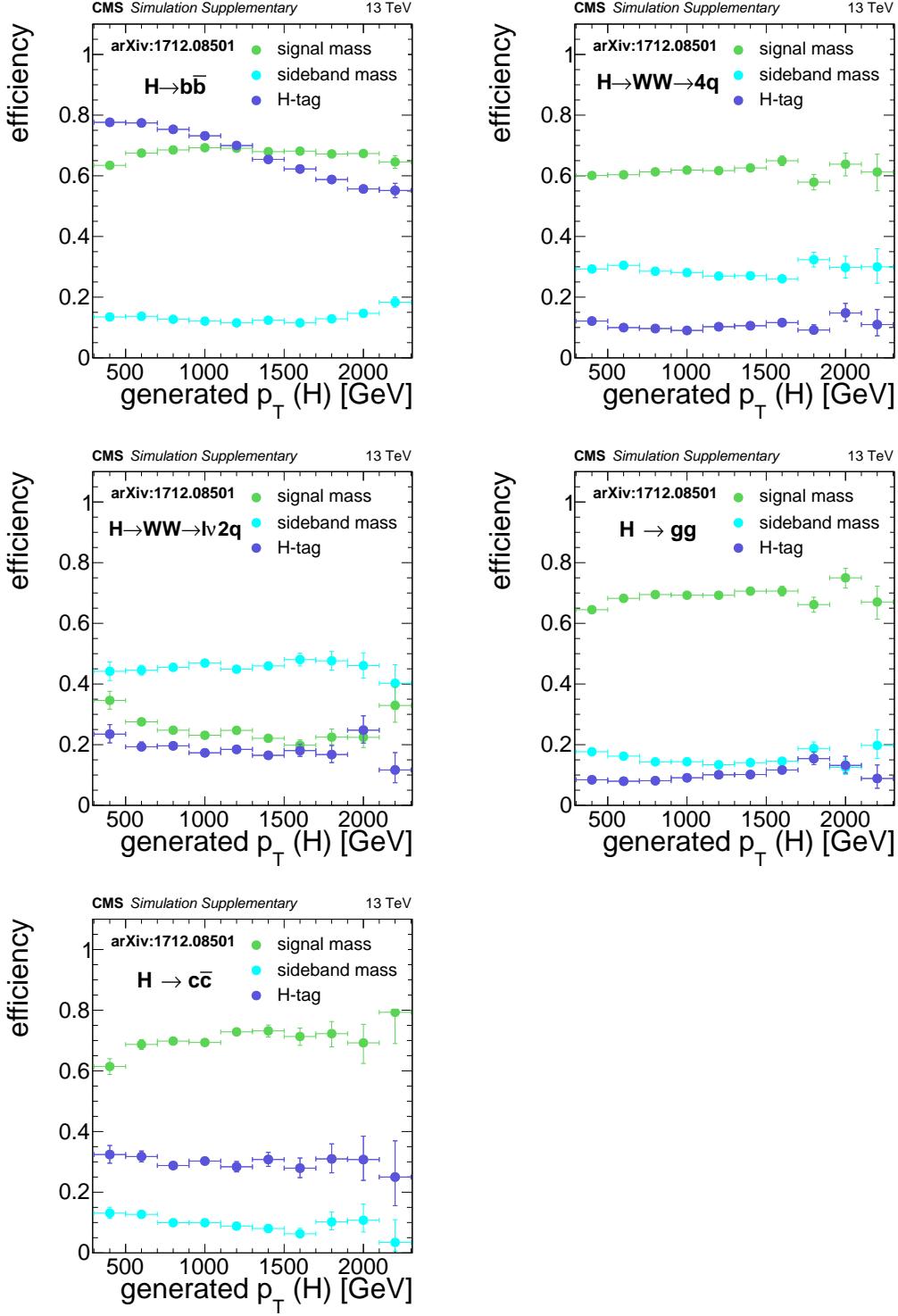


Figure B.2: Efficiencies for an AK8 jet originating from H boson decay, relative to baseline selection. "signal mass" represents the probability the jet will have mass [85, 135 GeV]. "sideband mass" represents the probability the jet will have mass [50, 85 GeV] or [135, 250 GeV]. "H-tag" represents the probability the jet have a double-b discriminator value greater than 0.3, for jets with mass [50, 250 GeV]. Efficiencies were derived using the T5ZH with a gluino mass of 2200 GeV.

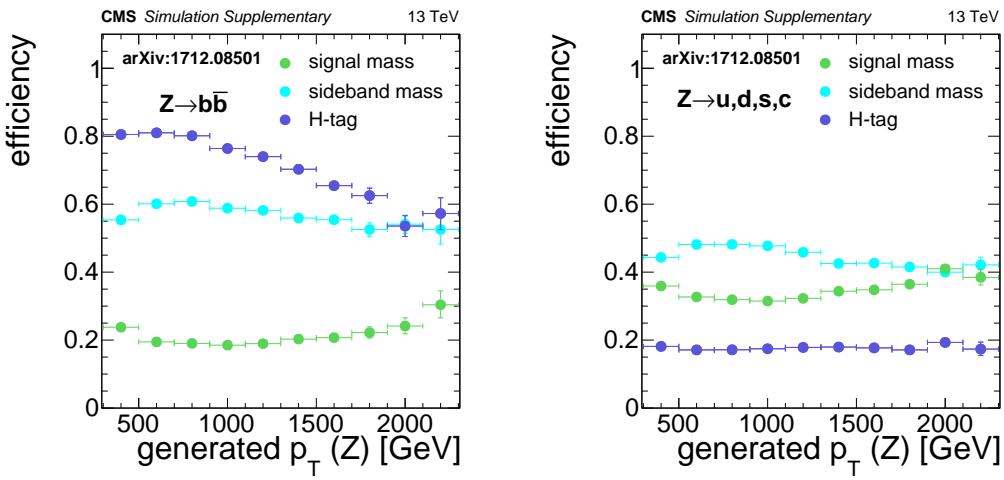


Figure B.3: Efficiencies for an AK8 jet originating from Z boson decay, relative to baseline selection. "signal mass" represents the probability the jet will have mass [85, 135 GeV]. "sideband mass" represents the probability the jet will have mass [50, 85 GeV] or [135, 250 GeV]. "H-tag" represents the probability the jet have a double-b discriminator value greater than 0.3, for jets with mass [50, 250 GeV]. Efficiencies were derived using the T5ZH with a gluino mass of 2200 GeV.

## Appendix C

### Determining $b\bar{b}$ -tagging Scale Factors for W jets in $t\bar{t}$ Events

At LHC energies,  $t\bar{t}$  production has a large production cross section and often makes up a substantial component of the background for physics analyses. These events have a rich final state with two b jets from each of the top decays, and hadrons or leptons from the decay of each of the W bosons. Semileptonic  $t\bar{t}$  events, illustrated in Figure C.1, are events in which one W decays to a lepton and neutrino, and the other decays hadronically. These events have a true source of missing energy ( $p_T^{\text{miss}}$ ), and additionally can result in boosted W bosons.

Although W bosons do not decay to b-quarks (the top quark is too heavy), the  $b\bar{b}$ -tagger, described in Section 7.2 has a non-zero probability to tag one of these jets as having decayed to a  $b\bar{b}$  pair. This *mistag* rate can, in principle, be different in simulation compared to in data. Depending on the use, it may be necessary to apply *scale factors* to the simulation to correct for this difference. This appendix provides a description of how these scale factors and their uncertainties are determined.

To obtain a clean source of hadronically decaying W bosons, events are selected that are consistent with semileptonic  $t\bar{t}$  events. The events are obtained from a data sample with a trigger requirement of at least one muon with  $p_T > 50 \text{ GeV}$  and are reconstructed in the same manner as described in Chapter 6. In particular, jets are reconstructed with the anti- $\text{kt}$  algorithm [26] operated with distance parameters of 0.4 (AK4 jets) and 0.8 (AK8 jets). The events are required to have a single muon with  $p_T > 50 \text{ GeV}$  and  $|\eta| < 2.1$  and be close ( $\Delta\phi < 2\pi/3$ ) to a b-tagged AK4 jet, where the loose b-tagging working point of the CSV algorithm is used [43]. Away from

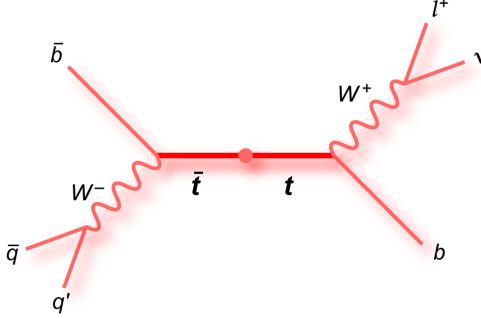


Figure C.1: Diagram of a semileptonic  $t\bar{t}$  event; one  $W$  boson decays leptonically and one  $W$  boson decays hadronically [42].

this muon ( $\Delta\phi > 2\pi/3$ ), we require at least one AK8 jet with  $p_T > 250$  GeV and  $|\eta| < 2.4$ , intended to be the  $W$  jet. To increase the purity of the sample, three additional criteria are applied:

- The AK8 jet must be consistent with having a two-jet substructure. This requirement is made by defining a variable “ $n$ -subjettiness” which parametrizes the degree to which an AK8 jet is consistent with having  $n$  subjets [44, 45]. The variable is calculated by forming a  $p_T$  weighted sum of the  $\Delta R$  distance of every particle in the jet with its nearest subjet, for some fixed number of subjets:  $\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}$ , where  $d_0 = 0.8 \sum_k p_{T,k}$ , is a normalization parameter. The ratio  $\tau_2/\tau_1$  is formed to discriminate between jets with two subjets from those with one.
- The AK8 jet mass must be between 50 and 200 GeV. This is consistent with the  $W$  boson mass of 80.379 GeV, and tends to reject jets arising from QCD-only interactions, which tend to have lower mass.
- To minimize contamination from nearby objects, no AK4 jets are allowed within  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 0.8$  of the selected AK8 jet.

Figure C.2 (left) shows the AK8 jet mass distribution from data and simulation after all selection criteria have been applied. The distribution peaks near the  $W$  boson mass (80.379 GeV) and is dominated by the desired signal. There is also a shoulder near the top quark mass (173.0 GeV), which arises from the merging of a  $W$  boson and a  $b$  quark into a single AK8 jet.

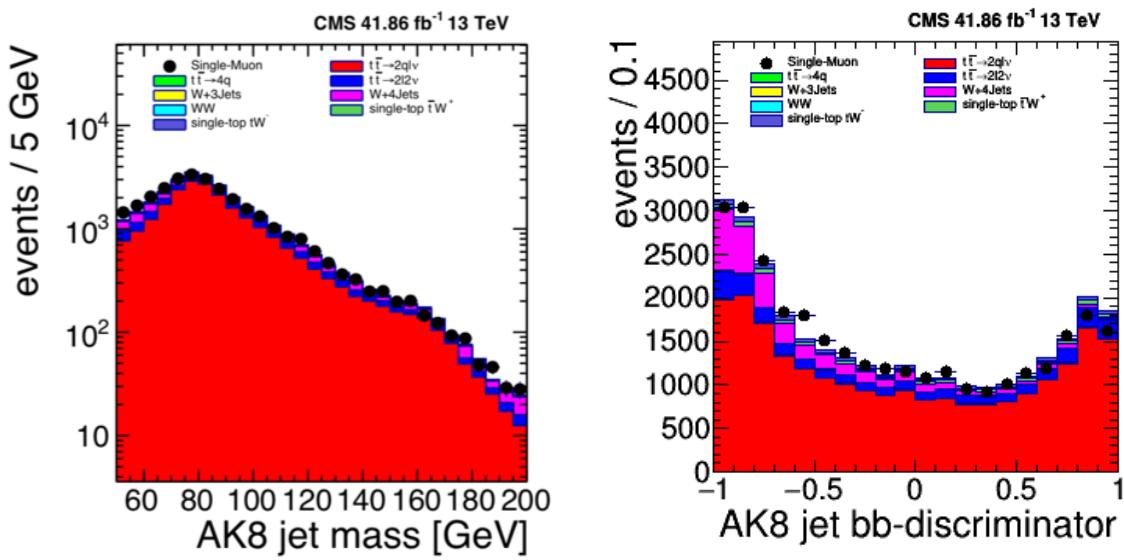


Figure C.2: AK8 jet mass and  $b\bar{b}$ -tagging discriminator distributions from data and simulation. The simulation distributions are normalized by area to the data.

Table C.1: W jet  $b\bar{b}$  mistag rates in data and simulation, inclusive in  $p_T$ . Statistical uncertainties only.

working point	data	MC
disc > 0.3	$0.326 \pm 0.003$	$0.340 \pm 0.003$
disc > 0.6	$0.219 \pm 0.003$	$0.236 \pm 0.002$
disc > 0.8	$0.122 \pm 0.002$	$0.138 \pm 0.002$
disc > 0.9	$0.058 \pm 0.001$	$0.067 \pm 0.001$

Figure C.2 (right) shows the distribution of the  $b\bar{b}$ -tagging discriminator for data and simulation. The simulation contributions are divided into signal and background components, where the signal contributions are inclusive  $t\bar{t}$  events and the background contributions are from events with direct W boson plus jets, WW, or single-top production. For a given working point (i.e. a fixed point along the x-axis), the mistag rate for the MC simulation is the ratio of the number of signal MC events above that point to the total number of signal MC events. The mistag rate for the data is calculated similarly except that the observed yields are corrected by subtracting the expected number of background events, estimated from the MC simulation:

$$\epsilon_{mc} = \frac{N_{b\bar{b}-tagged}^{sig,mc}}{N^{sig,mc}}, \quad \epsilon_{data} = \frac{N_{b\bar{b}-tagged}^{data} - N_{b\bar{b}-tagged}^{bkg,mc}}{N^{data} - N^{bkg,mc}}, \quad (C.1)$$

Following this prescription, we obtain the mistag rates in data and simulation seen in Table C.1. Four working points are defined ranging from disc>0.3 to disc>0.9. It is seen that the mistag rate in MC is higher for all working points than in data. This trend can be seen in the right-hand plot of Figure C.2 - the MC under-predicts the yields with low discriminator value, and over-predicts those with a large discriminator value.

The scale factors are then formed by taking the ratio of the mistag rate in data to that of simulation:

$$\text{scale factor} = \frac{\epsilon_{data}}{\epsilon_{mc}} \quad (C.2)$$

Table C.2 shows the scale factors, binned in jet  $p_T$ , for the four different working points. The uncertainties are the result of summing in quadrature the statistical and two additional systematic

Table C.2: Summary of scale factors for  $b\bar{b}$ -tagging W jets in  $t\bar{t}$  events.

working point	$250 < p_T < 350 \text{ GeV}$	$350 < p_T < 430 \text{ GeV}$	$p_T > 430 \text{ GeV}$
disc > 0.3	$0.939 \pm 0.026$	$1.007 \pm 0.055$	$0.996 \pm 0.079$
disc > 0.6	$0.922 \pm 0.027$	$0.967 \pm 0.057$	$0.902 \pm 0.082$
disc > 0.8	$0.875 \pm 0.030$	$0.939 \pm 0.063$	$0.893 \pm 0.090$
disc > 0.9	$0.855 \pm 0.036$		$0.914 \pm 0.068$

uncertainties: The first systematic uncertainty originates from subtracting the background from the data using the MC. The uncertainty in the cross section of the background contributions is estimated to be 30%, which results in uncertainties of 2, 4, and 6% for the low, medium, and high  $p_T$  bins, respectively. The second systematic uncertainty arises from a known difference in the  $p_T$  spectra of top quarks between data and simulation [42]. We assess the systematic uncertainty by comparing results when the top quark  $p_T$  spectra is and is not reweighted to match the data. This results in 0, 1, and 2% uncertainties for the low, medium, and high  $p_T$  bins, respectively.

## Bibliography

- [1] C. Borschensky, M. Krmer, A. Kulesza, M. Mangano, S. Padhi, T. Plehn, and X. Portell, “Squark and gluino production cross sections in pp collisions at  $\sqrt{s} = 13, 14, 33$  and  $100$  TeV,” *Eur. Phys. J.*, vol. C74, no. 12, p. 3174, 2014. arXiv:1407.5066.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B*, vol. 716, pp. 30–61, September 2012. arXiv:1207.7235.
- [3] S. P. Martin, “A Supersymmetry Primer,” pp. 58–60, January 2016. arXiv:hep-ph/9709356.
- [4] J. Wess and B. Zumino, “Supergauge transformations in four dimensions,” , 1974. arXiv:0910.5530.
- [5] 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. arxiv:1704.07781.
- [6] CMS Collaboration, “Search for supersymmetry in the multijet and missing transverse momentum final state in pp collisions at 13 TeV,” *Phys. Lett. B*, vol. 758, p. 152, 2016. arXiv:1602.06581.
- [7] “Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill.” <https://cds.cern.ch/record/2231915>.
- [8] CMS Collaboration, “The CMS Tracker Project: Technical Design Report,” 1997. CERN-LHCC-98-006 ; CMS-TDR-5.
- [9] CMS Collaboration, “The CMS Tracker Project: Addendum to the Technical Design Report,” 2000. CERN-LHCC-2000-016 ; CMS-TDR-5-add-1.
- [10] D. A. Matzner Dominguez, D. Abbaneo, K. Arndt, N. Bacchetta, A. Ball, E. Bartz, W. Bertl, G. M. Bilei, G. Bolla, H. W. K. Cheung, *et al.*, “CMS Technical Design Report for the Pixel Detector Upgrade,” , September 2012. CERN-LHCC-2012-016 ; CMS-TDR-11.
- [11] CMS Collaboration, “The CMS Electromagnetic Calorimeter Project: Technical Design Report,” 1997. CERN-LHCC-97-033 ; CMS-TDR-4.
- [12] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, “The CMS Electromagnetic Calorimeter Project: Addendum to the Technical Design Report,” 2002. CERN-LHCC-2002-027 ; CMS-TDR-4-add-1.

- [13] CMS Collaboration, “The CMS Hadron Calorimeter Project: Technical Design Report,” 1997. CERN-LHCC-97-031 ; CMS-TDR-2.
- [14] N. Strobbe, “The upgrade of the CMS hadron calorimeter with silicon photomultipliers,” *JINST*, vol. 12, January 2017. FERMILAB-CONF-16-501-CMS-E.
- [15] CMS Collaboration, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data,” *JINST*, vol. 5, March 2010. arXiv:0911.4991.
- [16] CMS Collaboration, “The CMS Magnet Project: Technical Design Report,” 1997. CERN-LHCC-97-010 ; CMS-TDR-1.
- [17] CMS Collaboration, “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays,” . arXiv:0910.5530.
- [18] CMS Collaboration, “The CMS Muon Project: Technical Design Report,” 1997. CERN-LHCC-97-032 ; CMS-TDR-3.
- [19] CMS Collaboration, “Performance of the CMS Drift Tube Chambers with Cosmic Rays,” *JINST*, vol. 5, March 2010. arXiv:0911.4855.
- [20] CMS Collaboration, “The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV at the LHC,” *JINST*, vol. 8, November 2013. arXiv:1306.6905.
- [21] CMS Collaboration, “Performance of the CMS cathode strip chambers with Cosmic Rays,” *JINST*, vol. 5, March 2010. arXiv:0911.4992.
- [22] Costantini et al., “Uniformity and Stability of the CMS RPC Detector at the LHC,” *JINST*, vol. 8, 2013. arXiv:1209.1989.
- [23] CMS Collaboration, “The CMS trigger system,” *JINST*, vol. 12, p. 01020, 2017. arXiv:1609.02366.
- [24] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, October 2014. arXiv:1405.6569.
- [25] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, p. 10003, 2017. arXiv:1706.04965.
- [26] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $\text{kt}$  jet clustering algorithm,” *JHEP*, no. 04, p. 063, 2008. arXiv:0802.1189.
- [27] CMS Collaboration, “Identification of double- $b$  quark jets in boosted event topologies,” 2016. CMS-PAS-BTV-15-002.
- [28] Sabine Kraml, Suchita Kulkarni, Ursula Laa, Andre Lessa, Wolfgang Magerl, Doris Proschofsky-Spindler, Wolfgang Waltenberger, “SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry,” *The European Physical Journal C*, May 2014. arXiv:1312.4175.
- [29] CMS Collaboration, “Search for supersymmetry in  $\text{pp}$  collisions at  $\sqrt{s} = 13\text{ TeV}$  in the single-lepton final state using the sum of masses of large-radius jets,” *JHEP*, August 2016. arXiv:1605.04608.

- [30] J. R. Stephen D. Ellis, Christopher K. Vermilion, “Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches,” *Phys. Rev. D*, May 2010. arXiv:0912.0033.
- [31] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 07, p. 079, 2014. arXiv:1405.0301.
- [32] R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015. arXiv:1410.8849.
- [33] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An introduction to PYTHIA 8.2,” *Computer Physics Communications*, vol. 191, pp. 159–177, June 2015. arXiv:1410.3012.
- [34] S. Agostinelli *et al.*, “GEANT4—a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. 506, p. 250, July 2003. [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [35] CMS Collaboration, “Search for supersymmetry in multijet events with missing transverse momentum in proton-proton collisions at 13 TeV,” CMS AN-16-350.
- [36] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, “The  $t\bar{t}$  cross-section at 1.8 TeV and 1.96 TeV: A Study of the systematics due to parton densities and scale dependence,” *JHEP*, vol. 4, no. 2004, p. 068, 2004. arXiv:hep-ph/0303085.
- [37] 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. arXiv:hep-ph/0306211.
- [38] 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 AN-16-215.
- [39] 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. arXiv:1712.08501.
- [40] “Identification of double-b quark jets in boosted event topologies,” Tech. Rep. CMS-PAS-BTV-15-002, CERN, Geneva, 2016.
- [41] “Identification of b quark jets at the CMS Experiment in the LHC Run 2,” Tech. Rep. CMS-PAS-BTV-15-001, CERN, Geneva, 2016.
- [42] [https://www-d0.fnal.gov/Run2Physics/top/top\\_public\\_web\\_pages/top\\_feynman\\_diagrams.html](https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html).
- [43] A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST*, vol. 13, p. P05011, 2018.
- [44] Jesse Thaler, Ken Van Tilburg, “Identifying Boosted Objects with N-subjettiness,” *JHEP*, March 2011. arXiv:1011.2268.
- [45] Jesse Thaler, Ken Van Tilburg, “Maximizing Boosted Top Identification by Minimizing N-subjettiness,” *JHEP*, February 2012. arXiv:1108.2701.