

This Dissertation  
entitled  
SEARCH FOR GAUGE-MEDIATED SUPERSYMMETRY BREAKING  
typeset with  $\text{NDdiss2}_{\mathcal{E}}$  v3.2013 (2013/04/16) on June 12, 2018 for  
Allison Reinsvold Hall

This L<sup>A</sup>T<sub>E</sub>X 2 <sub>$\mathcal{E}$</sub>  classfile conforms to the University of Notre Dame style guidelines as of Fall 2012. However it is still possible to generate a non-conformant document if the instructions in the class file documentation are not followed!

Be sure to refer to the published Graduate School guidelines at <http://graduateschool.nd.edu> as well. Those guidelines override everything mentioned about formatting in the documentation for this  $\text{NDdiss2}_{\mathcal{E}}$  class file.

It is YOUR responsibility to ensure that the Chapter titles and Table caption titles are put in CAPS LETTERS. This classfile does *NOT* do that!

*This page can be disabled by specifying the “noinfo” option to the class invocation.  
(i.e., \documentclass[... , noinfo]{nddiss2e} )*

**This page is *NOT* part of the dissertation/thesis. It should be disabled before making final, formal submission, but should be included in the version submitted for format check.**

$\text{NDdiss2}_{\mathcal{E}}$  documentation can be found at these locations:

<http://www.gsu.nd.edu>  
<http://graduateschool.nd.edu>

SEARCH FOR GAUGE-MEDIATED SUPERSYMMETRY BREAKING

A Dissertation

Submitted to the Graduate School  
of the University of Notre Dame  
in Partial Fulfillment of the Requirements  
for the Degree of

Doctor of Philosophy

by  
Allison Reinsvold Hall

---

Michael Hildreth, Director

Graduate Program in Physics

Notre Dame, Indiana

June 2018

# SEARCH FOR GAUGE-MEDIATED SUPERSYMMETRY BREAKING

Abstract

by

Allison Reinsvold Hall

This dissertation describes a search for new physics in final states with two photons and missing transverse momentum,  $E_T^{\text{miss}}$ . The analysis was performed using proton-proton collisions at a center-of-mass energy  $\sqrt{s} = 13$  TeV. The data were collected with the Compact Muon Solenoid (CMS) detector at the CERN LHC in 2016. The total integrated luminosity of the data set is  $35.9 \text{ fb}^{-1}$ . The results are interpreted in the context of gauge-mediated supersymmetry breaking (GMSB). Supersymmetry (SUSY) is an attractive extension to the standard model of particle physics. It addresses several known limitations of the standard model, including solving the hierarchy problem and providing potential dark matter candidates. In GMSB models, the lightest supersymmetric particle is the gravitino, and the next-to-lightest supersymmetric particle is often taken to be the neutralino. The neutralino can decay to a photon and a gravitino, leading to final state signatures with one or more photons and significant  $E_T^{\text{miss}}$ . No excess above the expected standard model backgrounds is observed, and limits are placed on the masses of SUSY particles in two simplified GMSB models. Gluino masses below 1.87 TeV and squark masses below 1.60 TeV are excluded at a 95% confidence level.

To my husband, Matt, for his unwavering love.

## CONTENTS

FIGURES . . . . .	vi
TABLES . . . . .	ix
ACKNOWLEDGMENTS . . . . .	x
CHAPTER 1: INTRODUCTION . . . . .	1
CHAPTER 2: THE STANDARD MODEL AND SUPERSYMMETRY . . . . .	3
2.1 The standard model of particle physics . . . . .	3
2.2 Supersymmetry . . . . .	12
2.3 Gauge-mediated supersymmetry breaking . . . . .	16
2.4 Experimental bounds on supersymmetry . . . . .	18
CHAPTER 3: THE LARGE HADRON COLLIDER . . . . .	23
3.1 Introduction . . . . .	23
3.2 Injector chain . . . . .	24
3.3 LHC tunnel and magnets . . . . .	24
3.4 Machine luminosity . . . . .	26
CHAPTER 4: CMS DETECTOR . . . . .	28
4.1 Coordinate system . . . . .	28
4.2 Superconducting solenoid . . . . .	29
4.3 Tracker . . . . .	31
4.4 Electromagnetic calorimeter (ECAL) . . . . .	33
4.5 Hadron calorimeter (HCAL) . . . . .	36
4.6 Calorimeter performance . . . . .	38
4.7 Muon system . . . . .	40
4.8 Instantaneous luminosity measurement . . . . .	41
CHAPTER 5: TRIGGER SYSTEM . . . . .	44
5.1 Overview of the CMS trigger . . . . .	44
5.2 Level 1 trigger . . . . .	45
5.3 High Level Trigger . . . . .	47
5.4 Analysis triggers . . . . .	48

5.5 Trigger efficiency . . . . .	51
<b>CHAPTER 6: CMS RECONSTRUCTION . . . . .</b>	<b>60</b>
6.1 Overview of the CMS reconstruction process . . . . .	60
6.2 Particle flow algorithm . . . . .	61
6.3 Jet reconstruction . . . . .	66
6.4 Missing transverse momentum . . . . .	68
6.5 Pileup subtraction . . . . .	71
6.6 Photon reconstruction . . . . .	73
<b>CHAPTER 7: DATA SETS AND SAMPLE DEFINITIONS . . . . .</b>	<b>76</b>
7.1 Data samples . . . . .	76
7.2 MC samples . . . . .	76
7.3 Object definitions . . . . .	78
7.4 Signal region and control samples . . . . .	81
7.5 Photon ID scale factor . . . . .	82
<b>CHAPTER 8: DATA ANALYSIS . . . . .</b>	<b>89</b>
8.1 Signal region . . . . .	89
8.2 Overview of backgrounds . . . . .	89
8.3 QCD background . . . . .	91
8.4 Electroweak background . . . . .	95
8.5 Irreducible background . . . . .	101
8.6 Systematic uncertainties . . . . .	101
<b>CHAPTER 9: RESULTS AND INTERPRETATIONS . . . . .</b>	<b>109</b>
9.1 Prediction versus observation . . . . .	109
9.2 Simplified models . . . . .	109
9.3 Signal acceptance and efficiency . . . . .	111
9.4 Limits and interpretations . . . . .	112
<b>CHAPTER 10: SUMMARY AND FUTURE PLANS . . . . .</b>	<b>121</b>
10.1 Future improvements . . . . .	121
10.2 GMSB combination paper . . . . .	121
10.3 Conclusions . . . . .	122
<b>APPENDIX A: ALTERNATE CONTROL SAMPLE FOR QCD BACKGROUND ESTIMATION . . . . .</b>	<b>124</b>
A.1 Problems with using $ee$ . . . . .	124
A.2 Results using ee sample . . . . .	127
<b>APPENDIX B: MOVING TOWARD A 2D ANALYSIS . . . . .</b>	<b>130</b>
B.1 Bin in number of jets . . . . .	130
B.2 ABCD method . . . . .	131

BIBLIOGRAPHY . . . . .	135
------------------------	-----

## FIGURES

2.1	Particle content of the standard model. . . . .	6
2.2	Cartoon illustrating the Higgs potential. . . . .	8
2.3	One-loop quantum corrections to the Higgs mass parameter. . . . .	10
2.4	Two-loop quantum corrections to the Higgs mass parameter. . . . .	11
2.5	Feynman diagram illustrating how the messenger particles in SUSY models with general gauge mediation give rise to mass terms for the MSSM gauginos. . . . .	17
2.6	Feynman diagrams illustrating how the messenger particles in SUSY models with general gauge mediation give rise to mass terms for the MSSM scalars. . . . .	18
2.7	Example GMSB processes leading to final states with two photons and $E_T^{\text{miss}}$ in $p$ - $p$ collisions. . . . .	19
2.8	Excluded sparticle mass ranges for a selection of CMS results using the 2016 data set. . . . .	20
2.9	The 95% CL upper limits on the gluino and squark pair production cross sections as a function of neutralino versus gluino or squark mass. . . . .	21
2.10	95% confidence level exclusion contours in the neutralino-gluino mass plane derived using $3.2 \text{ fb}^{-1}$ of data collected with the ATLAS detector in 2015. . . . .	22
3.1	Schematic of the CERN accelerator complex. . . . .	25
3.2	Peak instantaneous luminosity per day achieved by the LHC during the 2016 data-taking period. . . . .	27
4.1	Schematic of the CMS detector. . . . .	29
4.2	Calculated magnetic field $ \vec{B} $ for a longitudinal slice of the CMS detector. . . . .	30
4.3	Schematic of the CMS tracker system. . . . .	32
4.4	Relative response of the ECAL crystals for the 2011 to 2016 data-taking periods. . . . .	35
4.5	Longitudinal slice of the CMS HCAL. . . . .	37
4.6	Relative electron energy resolution in bins of pseudorapidity. . . . .	39
4.7	Integrated luminosity delivered to and recorded by CMS during $p$ - $p$ collisions at $\sqrt{s} = 13 \text{ TeV}$ in 2016. . . . .	43
5.1	Schematic showing the overall structure of the calorimeter half of the L1 trigger. . . . .	54

5.2	Schematic showing the overall structure of the muon half of the L1 trigger. . . . .	55
5.3	Efficiency of the L1 seed and the leading leg of the primary analysis trigger with respect to photon $p_T$ . . . . .	56
5.4	Efficiency of the trailing leg of the primary analysis trigger with respect to photon $p_T$ . . . . .	57
5.5	Efficiency of the primary analysis trigger with respect to the invariant mass of the diphoton system. . . . .	58
5.6	Efficiency of the trailing leg of the control trigger with respect to photon $p_T$ . . . . .	59
6.1	Slice of the transverse plane of the CMS detector showing how different particles interact with the detector subsystems. . . . .	61
6.2	Behavior of several jet-clustering algorithms illustrated with a sample parton-level event. . . . .	67
6.3	Ratio of the average reconstructed jet $p_T$ to the particle-level jet $p_T$ in QCD Monte Carlo simulation. . . . .	69
6.4	Jet energy resolution in 8 TeV simulation as a function of particle-level jet $p_T$ . . . . .	70
6.5	Number of pileup interactions per bunch crossing as observed in $p$ - $p$ collisions with the CMS detector in 2016. . . . .	71
6.6	Energy composition as a function of the number of pileup interactions for jets in data and simulation. . . . .	73
6.7	Average transverse momentum density $\rho$ for the 2016 $p$ - $p$ data set. . . . .	74
7.1	Calculated scale factors and uncertainties in bins of photon $p_T$ and $\eta$ . . . . .	83
7.2	Scale factors and uncertainties averaged over all photons in each bin in the neutralino versus gluino mass plane. . . . .	84
7.3	Scale factor for applying $R_9 > 0.5$ on top of the cut-based medium photon ID. . . . .	85
8.1	Event display of one of the candidate $\gamma\gamma$ events. . . . .	90
8.2	Illustration showing how the di-EM $p_T$ vector is defined. . . . .	92
8.3	Di-EM $p_T$ distributions of the $ff$ control sample and the $\gamma\gamma$ candidate sample. . . . .	93
8.4	$E_T^{\text{miss}}$ distributions of the unweighted $ff$ control sample and the $\gamma\gamma$ candidate sample. . . . .	94
8.5	$E_T^{\text{miss}}$ distributions of the di-EM $p_T$ reweighted $ff$ control sample and the $\gamma\gamma$ candidate sample. . . . .	95
8.6	Ratio of $\gamma\gamma$ to $ff$ events for $E_T^{\text{miss}} < 100$ GeV. . . . .	96
8.7	Data versus MC comparison of the $e\gamma$ control sample $E_T^{\text{miss}}$ distribution. . . . .	98
8.8	Invariant mass distributions for the $e\gamma$ and $ee$ samples. . . . .	106
8.9	Value of $R_{\gamma/e}$ for probes in the barrel as a function of various kinematic variables: probe $p_T$ , $E_T^{\text{miss}}$ , isolation $I_{\pm}$ , and track multiplicity of the primary vertex. . . . .	107

8.10	Value of $R_{\gamma/e}$ for probes in the barrel as a function of various kinematic variables: $\sigma_{i\eta i\eta}$ of the probe, vertex multiplicity, jet multiplicity, and probe $ \eta $ .	108
9.1	$E_T^{\text{miss}}$ distributions for the full background estimation and the observed data.	116
9.2	Diagrams showing the production of signal events in the collision of two protons.	117
9.3	NLO+NLL cross sections for the simplified gluino pair production and squark pair production models used in this analysis.	117
9.4	Acceptance $\times$ efficiency for the T5gg simplified model.	118
9.5	Acceptance $\times$ efficiency for the T6gg simplified model.	118
9.6	The 95% confidence level upper limits on the gluino pair production cross sections as a function of gluino and neutralino masses.	119
9.7	The 95% confidence level upper limits on the squark pair production cross sections as a function of squark and neutralino masses.	120
A.1	Unweighted $E_T^{\text{miss}}$ distributions of the $ee$ control sample and the $\gamma\gamma$ candidate sample.	127
A.2	Comparison of the di-EM $p_T$ distributions of the $\gamma\gamma$ and $ee$ samples.	128
A.3	$E_T^{\text{miss}}$ distributions of the candidate sample and the reweighted $ee$ sample.	129
B.1	The di-EM $p_T$ distributions for $\gamma\gamma$ and $ee$ events with $\geq 2$ jets.	131
B.2	Example of the ABCD analysis method using $ee$ events in data.	132
B.3	Exploring the correlation between $S_T$ and $E_T^{\text{miss}}$ significance using the ratio method.	134

## TABLES

2.1	FERMIONS IN THE STANDARD MODEL . . . . .	5
2.2	GAUGE BOSONS IN THE STANDARD MODEL . . . . .	7
2.3	CHIRAL SUPERMULTIPLETS IN MSSM . . . . .	14
2.4	GAUGE SUPERMULTIPLETS IN THE MSSM . . . . .	15
5.1	HLT TRIGGER PATHS . . . . .	49
5.2	PRIMARY TRIGGER REQUIREMENTS . . . . .	50
7.1	DATA SAMPLES . . . . .	77
7.2	MC SAMPLES FOR SIGNAL AND BACKGROUND . . . . .	86
7.3	EFFECTIVE AREAS FOR ISOLATION CORRECTIONS . . . . .	87
7.4	PHOTON AND ELECTRON ID CRITERIA . . . . .	87
7.5	FAKE ID CRITERIA . . . . .	88
8.1	COMPARISON BETWEEN REWEIGHTING METHOD AND RATIO METHOD FOR THE QCD BACKGROUND ESTIMATE . . .	97
8.2	EWK BACKGROUND ESTIMATE . . . . .	101
8.3	$Z\gamma\gamma$ BACKGROUND ESTIMATE . . . . .	102
8.4	UNCERTAINTIES ON THE QCD BACKGROUND ESTIMATE . .	103
8.5	UNCERTAINTIES ON SIGNAL YIELD . . . . .	105
9.1	FINAL BACKGROUND ESTIMATE . . . . .	110
9.2	EXPECTED AND OBSERVED EVENTS IN THE SIGNAL REGION	111
A.1	BACKGROUND MC SAMPLES FOR EE SAMPLE . . . . .	125
A.2	CONTRIBUTIONS TO EE SAMPLE . . . . .	126

## ACKNOWLEDGMENTS

There are many people whose guidance and support were instrumental in making this a reality. Thank you to my advisor, Mike Hildreth, for your patience and encouragement over the last few years. Thank you for always being willing to sit down with me to talk through whatever issue I was facing in the analysis. Thank you to the rest of my analysis group, both current and previous members, who contributed code, plots, and ideas to pull all of this together: Andrew Askew, Lisa Paspalaki, Tom McCauley, Marc Weinberg, Iasonas Topsis Giotis, and Arka Santra.

I also need acknowledge my committee members—Antonio Delgado, Colin Jessop, and Sylwia Ptasinska—for their guidance and suggestions, and Shari Herman, for all of the work she does to keep the department running and to look out for us graduate students. Thank you to Richard and Peggy Notebaert, whose generous gift to the university to found the Richard and Peggy Notebaert Premier Fellowship gave me the freedom to pursue this research.

Thank you to my classmates—Ek Adhikari, Rodolfo Capdevilla, Laura Collins, Ali Deatsch, Nabarun Dev, James Kelly, Liz Louden, Anthony Ruth and honorary member Patrick Louden—for an amazing five years. We've come a long way since sitting in awkward silence in the first year office working alone on prelim prep problems.

Many thanks need to go to my friends and family who have been supportive in so many ways. Thank you to my dad, for encouraging me to apply to Notre Dame in the first place. I am very glad I listened to you. I always appreciate your genuine interest and curiosity when I tell you about my research. Thank you to my mom, for

your constant love and encouragement. You do so many things for our family that often go unacknowledged. To my siblings, thank you for always keeping me humble.

Finally, thank you to my husband, Matt. Thank you for finding the balance between sympathetically listening to my complaints and ordering me to stop procrastinating and get to work. Thank you for all of your suggested edits and your patience and insights when I constantly asked your opinion on everything from how a plot looked to the basic principles of the analysis. This would not have been possible without you. I am excited for the adventures ahead of us.

## CHAPTER 1

### INTRODUCTION

The standard model of particle physics is one of the biggest success stories of modern physics. It provides a framework for understanding the basic building blocks of matter: the gauge bosons, quarks, leptons, and lastly the Higgs boson, discovered in 2012. The standard model has incredible predictive power to describe the interactions of these elementary particles. Over the last several decades, progressively more complex accelerators and detectors have been used to probe the standard model and to search for evidence of new physics. At the same time, theorists have made progress understanding the fundamental limitations of the theory and have proposed many potential extensions.

The state-of-the-art tool to explore the standard model and test new models is the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN), the largest machine in the world and a remarkable feat of engineering and science. The LHC is a proton-proton ( $p\text{-}p$ ) collider that can reach center-of-mass energies up to 13 TeV. The Compact Muon Solenoid (CMS) detector at the LHC is an all-purpose detector that measures the properties of particles produced in the collisions.

One of the most active research areas of the CMS Collaboration are searches for supersymmetric extensions of the standard model. Supersymmetry (SUSY) addresses several known limitations of the standard model, including solving the hierarchy problem and providing potential dark matter candidates. In this dissertation, a search is presented for new physics in final states with two photons and significant

missing transverse momentum. The results are interpreted as a search for SUSY appearing in the guise of gauge-mediated supersymmetry breaking (GMSB) models.

Chapter 2 will review the standard model and its supersymmetric extensions, focusing primarily on the phenomenology of the standard model and GMSB as it relates to this analysis. The LHC accelerator complex will be described in Chapter 3, and the CMS detector and subsystems will be described in Chapter 4. Chapters 5 and 6 will outline the CMS trigger system and reconstruction algorithms, respectively.

The primary analysis methods will be discussed in Chapters 7 and 8. Finally, Chapter 9 will include the results and interpretations. No excess above the expected standard model backgrounds is observed, and limits are set on two simplified GMSB models.

## CHAPTER 2

### THE STANDARD MODEL AND SUPERSYMMETRY

#### 2.1 The standard model of particle physics

The standard model (SM) of particle physics is a Lorentz-invariant quantum field theory that describes elementary particles and their allowed interactions. The first piece of the standard model was developed in 1961 with the unification of the theoretical descriptions of the electromagnetic and weak interactions [49, 50]. The Higgs mechanism was incorporated a few years later [51, 53, 42, 52], and the standard model took on the form we know today with the inclusion of the strong force and quantum chromodynamics (QCD) in the 1970's [68]. The standard model is incredibly successful, and it has made absurdly precise predictions that have held up to experimental scrutiny.

The standard model is a non-Abelian gauge theory. The symmetry group of the standard model is

$$G_{SM} = SU(3)_C \otimes SU(2)_L \otimes SU(1)_Y \quad (2.1)$$

where  $SU(3)_C$  is responsible for mixing the 3 *colors* of quarks and antiquarks in QCD,  $SU(2)_L$  represents *weak isospin*, and  $U(1)_Y$  couples to the *weak hypercharge*  $Y$ . The combination  $SU(2)_L \otimes SU(1)_Y$  corresponds to the electroweak interactions.

##### 2.1.1 Particle content of the standard model

The fields in the SM are identified by their representation in the symmetry group of Equation 2.1. These are listed for the SM fermions in Table 2.1 and for the gauge

bosons in Table 2.2. There are three generations of quarks and three generations of leptons in the SM. The quarks are in a non-trivial representation of all three SM symmetries and therefore interact via the strong, weak, and electromagnetic interactions. The three “up-type” quarks are the up quark  $u$ , the charm quark  $c$ , and the top quark  $t$ . The three “down-type” quarks are the down quark  $d$ , the strange quark  $s$ , and the bottom quark  $b$ .

The three generations of leptons include the electron  $e$ , muon  $\mu$ , and tau  $\tau$  and the corresponding neutrinos. In the SM, the neutrinos are massless, even though it has been experimentally established that at least two of the neutrinos must have non-zero mass [65, 63, 64]. The electron, muon and tau particles participate in the weak interaction and the electromagnetic interaction, but the neutrinos, being electromagnetically neutral, only participate in the weak interaction.

As can be seen in Table 2.1, the weak interaction does not respect parity. The left-handed quarks and leptons belong to  $SU(2)_L$  doublets, whereas the right-handed quarks and leptons form  $SU(2)_L$  singlets.

The gauge bosons are responsible for mediating the interactions between the fermions. Before electroweak symmetry breaking (EWSB), the three gauge bosons of  $SU(2)_L$  are the  $W^+$ ,  $W^-$ , and  $W^0$ , and the  $U(1)_Y$  gauge boson is the  $B$  boson. After EWSB, the  $W^+$  and  $W^-$  bosons gain a mass, and the  $W_0$  and  $B$  bosons mix to form the massive  $Z$  boson and the massless photon  $\gamma$ . EWSB will be described in more detail in Section 2.1.2.

For the strong force  $SU(3)_C$ , the gauge bosons are the gluons. Processes involving the interactions of quarks and gluons are referred to as QCD processes. The QCD interaction is different from the electroweak interactions in that the strong force grows in strength at larger distances. This leads to the phenomenon of *confinement*: at low energies and large distances (greater than  $1 \text{ GeV}^{-1}$ ), we cannot identify individual

TABLE 2.1  
FERMIIONS IN THE STANDARD MODEL

Name	Symbol	$SU(3)_C, SU(2)_L, U(1)_Y$
quarks (3 families)	$(u_L \ d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
	$u_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	$d_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
leptons (3 families)	$(\nu \ e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
	$e_R^\dagger$	$(\mathbf{1}, \mathbf{1}, 1)$

Fermions in the standard model and their representations under  $G_{SM}$ .

quarks or gluons. Instead, these particles can only be found in  $SU(3)_C$  singlets, typically either mesons ( $q\bar{q}$ ) or baryons ( $qqq$  or  $\bar{q}\bar{q}\bar{q}$ ).

Finally, the only particle not included in Table 2.1 or 2.2 is the Higgs boson. The Higgs boson was discovered by the CMS and ATLAS collaborations in 2012 [7, 16]. It is the only fundamental scalar in the SM and has a mass of approximately 125 GeV. The Higgs boson is in the  $(\mathbf{1}, \mathbf{2}, +\frac{1}{2})$  representation of  $G_{SM}$ . It is uncharged under the strong and electromagnetic interactions.

Figure 2.1 summarizes the SM particles and their allowed interactions.

### 2.1.2 Electroweak symmetry breaking

Electroweak symmetry breaking is a crucial feature of the standard model. The electroweak symmetry is spontaneously broken through the Higgs mechanism. The scalar potential of the complex Higgs field  $H$  is

$$V(H^\dagger H) = \mu^2(H^\dagger H) + \lambda(H^\dagger H)^2 \quad (2.2)$$

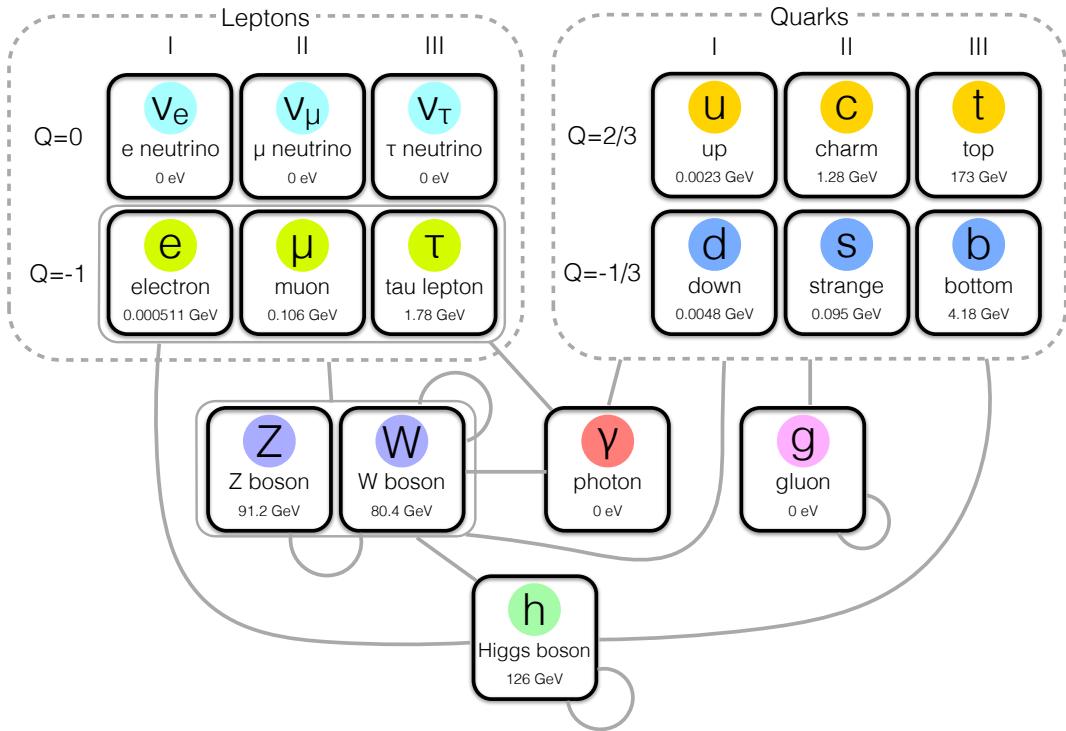


Figure 2.1: Particle content of the standard model. Allowed interactions are indicated with gray lines between particles or particle groups. The mass of each particle is listed underneath its symbol and name. The electric charge  $Q$  of the leptons and quarks is also indicated. Particles connected to themselves can have self-interactions. Note that in the standard model, neutrinos are still treated as being massless, even though it has been established that at least two neutrinos must have non-zero mass. Reprinted from Reference [54].

TABLE 2.2  
GAUGE BOSONS IN THE STANDARD MODEL

Name	Symbol	$SU(3)_C, SU(2)_L, U(1)_Y$
gluon	$g$	(8, 1, 0)
$W$ bosons	$W^\pm, W^0$	(1, 3, 0)
$B$ boson	$B^0$	(1, 1, 0)

Gauge bosons in the standard model before electroweak symmetry breaking and their representations under  $G_{SM}$ .

The mass parameter  $\mu^2$  is negative, which means that the vacuum expectation value (vev) of the Higgs (written as  $\langle H \rangle$  or  $v$ ) is non-zero. The Higgs potential is illustrated in Figure 2.2. Spontaneous symmetry breaking occurs when the value of the Higgs field “rolls” from the unstable position at the origin and comes to rest at the stable minima:

$$\langle H \rangle = v = \sqrt{\frac{-\mu^2}{2\lambda}} \quad (2.3)$$

The value of  $v$  has been measured in electroweak interactions to be about 174 GeV.

During EWSB, three of the four degrees of freedom in the Higgs complex scalar doublet get “eaten” to give mass to the  $W^\pm$  and  $Z$  bosons. The last degree of freedom becomes the 125 GeV scalar Higgs boson. The mass  $m_H$  of the Higgs boson can be written in terms of the parameters  $\mu$  and  $\lambda$ :

$$m_H = \sqrt{-\mu^2} = v\sqrt{2\lambda} \quad (2.4)$$

The remaining symmetry is  $U(1)_{EM}$ , which corresponds to the electromagnetic interaction. The massless photon is the gauge boson of  $U(1)_{EM}$  and couples to the

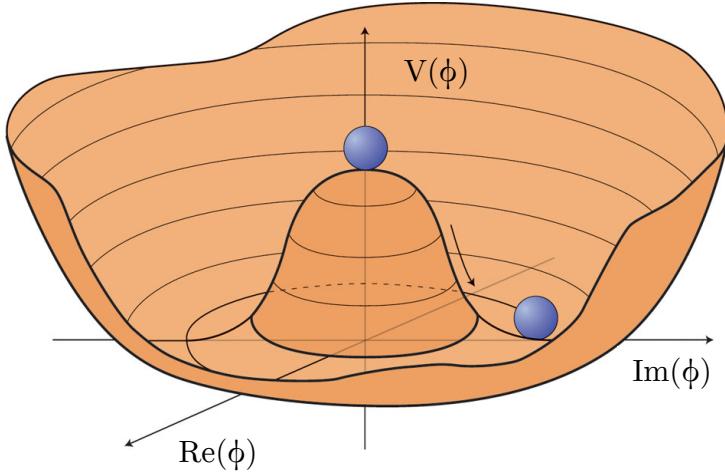


Figure 2.2: Cartoon illustrating the Higgs potential. The Higgs field during spontaneous symmetry breaking can be thought of as the ball, which “rolls” from the unstable point to the stable minima. Reprinted from Reference [3].

electric charge  $Q$ :

$$Q = T^3 + \Upsilon \quad (2.5)$$

where  $\Upsilon$  is the weak hypercharge and  $T^3$  is the isospin. The short-range weak interactions are mediated by the  $W^\pm$  and  $Z$  bosons, and the photon mediates the long-range electromagnetic interactions.

### 2.1.3 Limitations of the standard model

It is undeniable that the standard model has been a huge triumph of modern physics. However, there are known limitations. These fall into two categories:

1. Experimentally-established phenomena which are not incorporated into the standard model
2. Unsettling questions that the standard model cannot answer but which perhaps would be explained by some more fundamental theory

Neutrino masses—an example of the first category—have already been mentioned. The prevalence of matter over antimatter is another observed contradiction to the standard model as it exists today. The only SM source of CP violation is quark mixing, which is not nearly large enough to explain why the universe is more than just a bath of radiation from the annihilation of particles and antiparticles.

The existence of dark matter also falls into the first category. A wealth of evidence, including galaxy cluster velocity dispersions, gravitational lensing studies, and measurements of the Cosmic Microwave Background, indicates that approximately 25% of the energy budget of the universe is contained in dark matter (see Reference [5] for a review). We know that dark matter interacts gravitationally, and stringent limits have been placed on the strength of its other interactions. However, so far there is no conclusive evidence as to the nature of dark matter.

The second category includes questions about why the 19 free parameters in the standard model have the values that they do. For instance, it could be seen as disconcerting that the masses of the particles in Figure 2.1 cover such a large range, from  $m_e = 0.000511$  GeV to  $m_t = 173$  GeV. This has motivated the formation of flavor models that explain the mass structure between the generations of quarks and leptons.

The problem of “fine-tuning” is often discussed in this context. A theory is said to be fine-tuned if several free parameters have to take on precise values and interact in a convenient way in order to explain the observed results. While nothing is inherently wrong with fine-tuning, it is an unsatisfying way to solve a problem. Many people use fine-tuned parameters as sign-posts pointing to the existence of a broader theory in which the coincidences are explained.

The hierarchy problem<sup>1</sup> is a particularly worrisome fine-tuning problem. The

---

<sup>1</sup>Much of the description here and in Section 2.2 is derived from the excellent pedagogical paper “A Supersymmetry Primer” by Stephen P. Martin [58].

squared mass parameter  $\mu^2$  in the Higgs potential receives huge quantum corrections from loop diagrams. As already stated in Equation 2.4,  $\mu^2$  can be written in terms  $v$  and  $\lambda$ , and  $v$  has been measured to be approximately 174 GeV. Some fantastic cancellation must occur to counteract these quantum corrections.

Every particle that couples directly or indirectly to the Higgs field contributes to the quantum corrections to  $\mu^2$ . For a fermion  $f$ , the loop-level contributions come from the Feynman diagram of Figure 2.3a:

$$\Delta\mu^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{\text{UV}}^2 + \dots \quad (2.6)$$

where the relevant term in the Lagrangian is  $-\lambda_f H \bar{f} f$  and  $\Lambda_{\text{UV}}$  is the ultraviolet momentum cutoff of the theory.

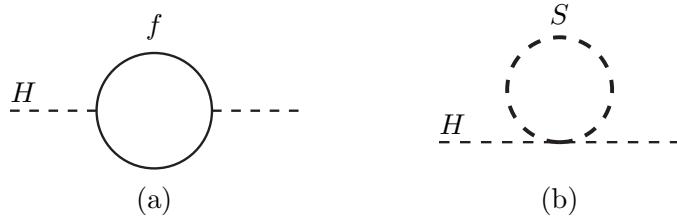


Figure 2.3: One-loop quantum corrections to the Higgs mass parameter  $\mu^2$  from a Dirac fermion  $f$  (left) and a scalar  $S$  (right). Reprinted from Reference [58].

Even if it does not couple directly to the Higgs, any new particle would also contribute to  $\mu^2$  at the two-loop level if it participates in the SM gauge interactions. For example, consider a heavy fermion  $F$  with mass  $m_F$  that only couples indirectly to the Higgs. Example two-loop Feynman diagrams are shown in 2.4. For such a

fermion, the quantum corrections will be given by

$$\Delta\mu^2 = C \left( \frac{g^2}{16\pi^2} \right)^2 [a\Lambda_{\text{UV}}^2 + 24m_F^2 \ln(\Lambda_{\text{UV}}/m_F) + \dots] \quad (2.7)$$

where the constant  $C$  represents several group theory factors,  $a$  depends on the renormalization method, and  $g$  is the gauge coupling [58].



Figure 2.4: Two-loop quantum corrections to the Higgs mass parameter  $\mu^2$  from a fermion  $F$  that couples indirectly to the Higgs through gauge interactions. Reprinted from Reference [58].

The root of the hierarchy problem is the presence of  $\Lambda_{\text{UV}}^2$  in both Equations 2.6 and 2.7. If  $\Lambda_{\text{UV}}^2$  is near the Planck scale ( $M_P = 2 \times 10^{18}$  GeV), then the corrections to  $\mu^2$  are an unbelievable 30 orders of magnitude above its observed value. If we use dimensional regularization, it is possible to get rid of the  $\Lambda_{\text{UV}}^2$  terms in Equations 2.6 and 2.7, but the term proportional to  $m_F^2$  remains [58]. Since it seems incredibly unlikely that there are no new particles to be found in the 16 orders of magnitude between the electroweak and Planck scales, there is a tension between the experimentally-preferred and the theoretically-preferred values.

As will be discussed next, this observation has motivated many proposed extensions to the standard model, including supersymmetry.

## 2.2 Supersymmetry

Supersymmetry (SUSY) is a particularly beautiful theoretical framework that addresses several of the standard model limitations described above. A “supersymmetry” is a symmetry relating bosons and fermions. The supersymmetric operator  $Q$  turns a bosonic state into a fermionic state:

$$Q|\text{boson}\rangle = |\text{fermion}\rangle, \quad Q|\text{fermion}\rangle = |\text{boson}\rangle \quad (2.8)$$

The immediate consequence is that the number of elementary particles is at least double that in the standard model. For each SM boson, there will be a fermion “superpartner” and vice versa. The deeper implications of Equation 2.8, however, have many interesting effects, some of which provide attractive solutions to the issues with the SM listed above.

### 2.2.1 Supersymmetric solution to the hierarchy problem

The key to solving the hierarchy problem with supersymmetry appears if we consider a heavy complex scalar  $S$  with mass  $m_S$ . If the interaction term between  $S$  and  $H$  in the Lagrangian is  $-\lambda_S|H|^2|S|^2$ , then diagrams such as those in Figure 2.3b give rise to the following quantum corrections:

$$\Delta\mu^2 = -\frac{\lambda_S^2}{16\pi^2}[\Lambda_{\text{UV}}^2 - 2m_S^2 \ln(\Lambda_{\text{UV}}/m_S) + \dots] \quad (2.9)$$

Upon close inspection of Equations 2.6 and 2.9, it becomes clear that if there are two complex scalars per fermion that satisfy  $\lambda_S = |\lambda_f|^2$ , then the contributions to  $\mu^2$  that are proportional to  $\Lambda_{\text{UV}}^2$  cancel exactly. Less obvious is whether the contributions from diagrams with more loops also cancel. The good news is that it can be shown that the contributions to  $\mu^2$  cancel to all orders in a supersymmetric theory [41].

Any new particles, even extremely massive ones, would be accompanied by their superpartners and  $\mu^2$  would remain unaffected.

### 2.2.2 Particles in minimal SUSY models

In a supersymmetric theory, the SM particles shown in Tables 2.1 and 2.2 must be arranged into “supermultiplets” containing the SM fermions and bosons and their superpartners. There must be an equal number of bosonic and fermionic degrees of freedom in each supermultiplet ( $n_b$  and  $n_f$ , respectively). The SM particles fall into two types of supermultiplets:

- The standard model fermions belong to **chiral supermultiplets**. Each chiral supermultiplet contains one Weyl fermion ( $n_f = 2$ ) and a complex scalar field ( $n_b = 2$ ).
- The spin-1 gauge bosons belong to **gauge supermultiplets**. Before EWSB, the gauge bosons are massless and have  $n_b = 2$ . Their superpartners are massless spin-1/2 Weyl fermions with  $n_f = 2$ .

The Minimal Supersymmetric Standard Model (MSSM) is the most basic SM extension that incorporates SUSY. Table 2.3 lists the chiral supermultiplets in the MSSM, and the gauge supermultiplets are listed in Table 2.4. Note that all particles in a supermultiplet are in the same representation of  $G_{SM}$ .

There must be two Higgs chiral supermultiplets in the MSSM to avoid gauge anomalies in the electroweak gauge symmetry, one with hypercharge  $\Upsilon = 1/2$  and one with  $\Upsilon = -1/2$ . The  $\Upsilon = 1/2$  chiral supermultiplet  $H_u$  has the Yukawa couplings necessary to give masses to the up-type quarks, and the Yukawa couplings of the  $\Upsilon = -1/2$  chiral supermultiplet  $H_d$  give masses to the down-type quarks and leptons. The 125 GeV standard model Higgs boson is a linear combination of  $H_u^0$  and  $H_d^0$ .

TABLE 2.3  
CHIRAL SUPERMULTIPLETS IN MSSM

Names		Spin 0	Spin 1/2	$SU(3)_C, SU(2)_L, U(1)_Y$
squarks, quarks (3 families)	Q	$(\tilde{u}_L \ \tilde{d}_L)$	$(u_L \ d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
	$\bar{u}$	$\tilde{u}_R^*$	$u_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
	$\bar{d}$	$\tilde{d}_R^*$	$d_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
sleptons, leptons (3 families)	L	$(\tilde{\nu} \ \tilde{e}_L)$	$(\nu \ e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
	$\bar{e}$	$\tilde{e}_R^*$	$e_R^\dagger$	$(\mathbf{1}, \mathbf{1}, 1)$
Higgs, higgsinos	$H_u$	$(H_u^+ \ H_u^0)$	$(\tilde{H}_u^+ \ \tilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, +\frac{1}{2})$
	$H_d$	$(H_d^0 \ H_d^-)$	$(\tilde{H}_d^0 \ \tilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

List of chiral supermultiplets in the MSSM. Reprinted from Reference [58].

The names of the MSSM superpartners are very creatively derived from their SM counterparts. For fermions, the SM names are prefixed with “s” to get the superpartner name: electrons to selectrons, quarks to squarks, etc. For bosons, the suffix “ino” is added:  $W$  to wino, gauge boson to gaugino, etc. The symbols for supersymmetric particles are given simply by adding a tilde to the SM symbol. For example,  $\tilde{e}$  is the symbol for the selectron.

After EWSB, the neutral higgsinos ( $\tilde{H}_u^0$  and  $\tilde{H}_d^0$ ) and the neutral electroweak gauginos ( $\widetilde{W}_0$  and  $\widetilde{B}_0$ ) form four mass eigenstates known as neutralinos  $\tilde{\chi}^0$ . The charged higgsinos ( $\tilde{H}_u^+$  and  $\tilde{H}_d^-$ ) and charged winos ( $\widetilde{W}^\pm$ ) form four mass eigenstates known as charginos  $\tilde{\chi}^\pm$ .

### 2.2.3 $R$ -parity

Phenomenologically, an important feature of many SUSY models is  $R$ -parity conservation.  $R$ -parity is a new symmetry introduced in the MSSM to get rid of inter-

TABLE 2.4  
GAUGE SUPERMULTIPLETS IN THE MSSM

Names	Spin 1/2	Spin 1	$SU(3)_C, SU(2)_L, U(1)_Y$
gluino, gluon	$\tilde{g}$	$g$	(8, 1, 0)
wino, $W$ bosons	$\widetilde{W}^\pm, \widetilde{W}^0$	$W^\pm, W^0$	(1, 3, 0)
bino, $B$ boson	$\widetilde{B}^0$	$B^0$	(1, 1, 0)

List of gauge supermultiplets in the MSSM. Reprinted from Reference [58].

actions that violate lepton number  $L$  or baryon number  $B$ . Experimentally, we know that if  $B$ - and  $L$ -violating processes occur, they must be extremely suppressed. In particular, such processes would mediate proton decay, but we know that the lifetime of the proton must be greater than  $10^{32}$  years [58]. Without  $R$ -parity conservation, however, it is possible to write down terms in the MSSM Lagrangian that would allow the proton to decay to  $e^+\pi^0$  or other lepton + meson combinations with a lifetime of less than one second.

$R$ -parity (also known as matter parity) is a conserved quantum number:

$$P_R = (-1)^{3(B-L)+2s} \quad (2.10)$$

where  $s$  is the spin of the particle. All terms in the Lagrangian are required to satisfy  $\prod P_R = +1$ , where the product is over all fields in the term. The SM particles and the Higgs bosons have  $P_R = +1$ , and all SUSY particles—squarks, sleptons, gauginos, and higgsinos—have  $P_R = -1$ .

Most models of SUSY, including those explored in this analysis, assume  $R$ -parity conservation. One immediate consequence is that each vertex must have an even number of supersymmetric particles (“sparticles”). In a collider experiment this

means that sparticles must be pair-produced. Additionally,  $R$ -parity conservation implies that the lightest supersymmetric particle (LSP) must be absolutely stable. All other sparticles will eventually decay to the LSP. From the cosmologist’s point of view, this means that the LSP, if electrically neutral, is a potential dark matter candidate. From the high energy experimentalist’s point of view, this means that the primary signature of SUSY production in colliders will be an imbalance in the reconstructed momentum in the transverse plane of the detector. This is referred to as missing transverse momentum,  $\vec{p}_T^{\text{miss}}$ . The magnitude of  $\vec{p}_T^{\text{miss}}$  is called the missing transverse energy, or  $E_T^{\text{miss}}$ .

### 2.3 Gauge-mediated supersymmetry breaking

So far no SUSY particles have been observed. If SUSY were an unbroken symmetry, then the masses of the superpartners would be the same as their SM counterparts. The lack of experimental detection means that SUSY must be a broken symmetry. Theorists have come up with many different mechanisms by which SUSY can be broken. The SUSY breaking mechanism and the SUSY breaking scale determine the phenomenology of the model, including the SUSY particle masses and their allowed decay modes.

In gauge-mediated supersymmetry breaking (GMSB) [45–47, 38–40, 59, 12, 2], SUSY is spontaneously broken in a “hidden” sector that is decoupled from the SM interactions. New chiral supermultiplets known as messengers interact with the hidden sector and also interact with the “visible” sector (ie, SM particles and their superpartners) through the normal SM gauge interactions. The messengers are said to communicate between the two sectors and relay the information about the SUSY breaking. From the lack of experimental observation, we know that the masses of the messenger particles must be very high.

One of the nice features of GMSB is that the flavor symmetry of the SM is

preserved, because the gauge interactions are flavor-blind. The messengers provide large contributions to the sparticle masses that are sensitive to the scale of SUSY breaking. Figure 2.5 shows the relevant loop diagrams for gauginos, and Figure 2.6 shows the two-loop diagrams for squarks and sleptons.

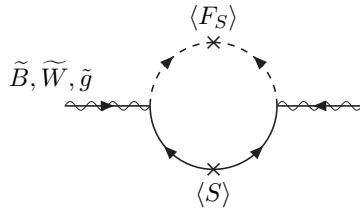


Figure 2.5: Feynman diagram illustrating how the messenger particles in SUSY models with general gauge mediation give rise to mass terms for the MSSM gauginos. Wavy lines superimposed on straight lines represent the gauginos, and  $\langle F_S \rangle$  and  $\langle S \rangle$  represent messenger particles. Reprinted from Reference [58].

GMSB usually refers to the minimal version of the theory, where the hidden sector includes a field  $S$  that is a singlet under  $G_{SM}$ . The more general framework is referred to as general gauge mediation (GGM).

The GGM phenomenology is independent from the exact SUSY-breaking mechanism in the hidden sector. In particular, in all GGM models, the lightest supersymmetric particle is the gravitino  $\tilde{G}$ , the superpartner of the graviton. The graviton is a spin-2 particle, which means the graviton and gravitino can not belong to either the chiral or gauge supermultiplets described in Section 2.2.2. Instead, the gravitino is a spin-3/2 particle.

In GGM, the gravitino must be significantly lighter than all other sparticles, on the order of eV to keV. The gravitino only interacts gravitationally with SM particles and therefore escapes the detector without depositing its energy, resulting in significant

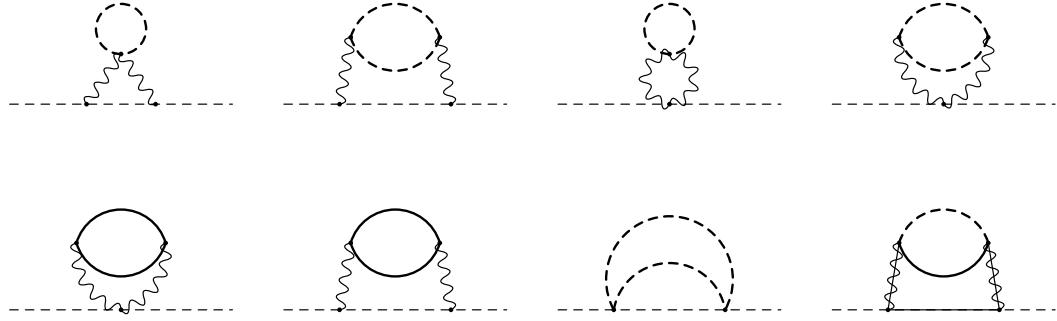


Figure 2.6: Feynman diagrams illustrating how the messenger particles in SUSY models with general gauge mediation give rise to mass terms for the MSSM scalars. Wavy lines superimposed on straight lines represent the MSSM gauginos, dotted lines represent the MSSM scalars, wavy lines represent the SM gauge bosons, heavy dashed lines represent messenger scalars, and solid lines represent messenger fermions. Reprinted from Reference [58].

$E_T^{\text{miss}}$  in the final state.

The next-to-lightest supersymmetric particle (NLSP) is either the lightest chargino or the lightest neutralino. In this analysis, we will only consider models where the NLSP is the neutralino. We further constrain ourselves to models where the neutralino decays promptly. In general, the allowed neutralino decays are  $\tilde{\chi}_1^0 \rightarrow \gamma \tilde{G}$ ,  $\tilde{\chi}_1^0 \rightarrow H \tilde{G}$ , or  $\tilde{\chi}_1^0 \rightarrow Z \tilde{G}$ , but the first case normally dominates [66]. We will assume 100% branching fraction of  $\tilde{\chi}_1^0 \rightarrow \gamma \tilde{G}$  for the interpretation of our results.

If we assume  $R$ -parity conservation, then sparticles must be pair-produced and eventually decay to a gravitino. In a  $p$ - $p$  collider such as the LHC, the dominant production modes for SUSY particles are either gluino pair production or squark pair production. Example decay chains are shown in Figure 2.7. The final state signature is characterized by two photons and significant  $E_T^{\text{miss}}$ .

## 2.4 Experimental bounds on supersymmetry

So far, there has been no experimental evidence for supersymmetry. In the absence of evidence, limits are placed on the allowed sparticle masses and SUSY process cross

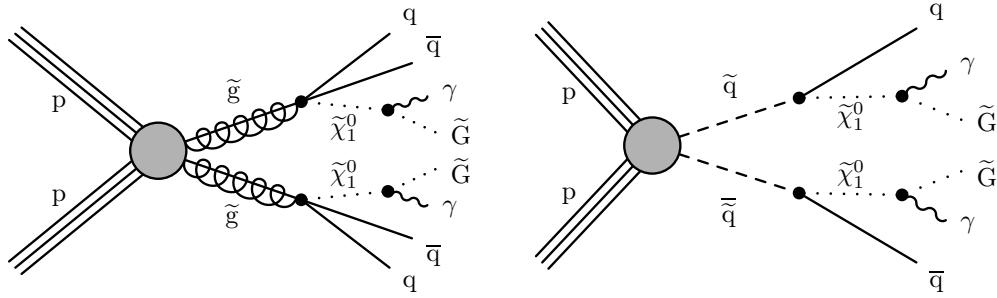
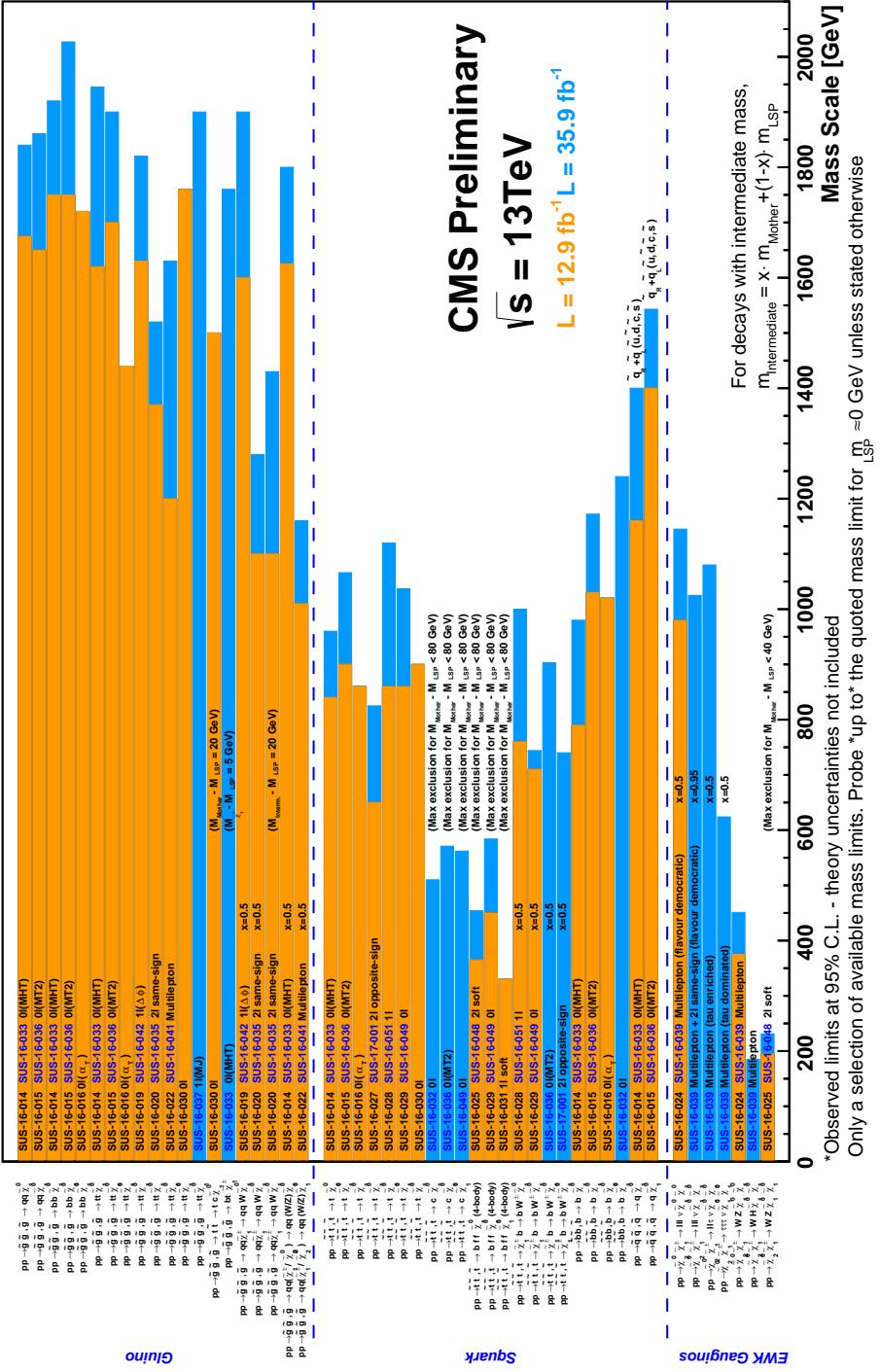


Figure 2.7: Example GMSB processes leading to final states with two photons and  $E_T^{\text{miss}}$  in  $p$ - $p$  collisions. The diagram on the left illustrates the decay chain for gluino pair production, and the diagram on the right illustrates it for squark pair production. In both cases, two neutralinos  $\tilde{\chi}_1^0$  are produced that each subsequently decay to a photon  $\gamma$  and a gravitino  $\tilde{G}$ .

sections. During Run I of the LHC at CERN, protons were collided at a center-of-mass energy  $\sqrt{s} = 8$  TeV, setting a new record for high energy collisions. No signs of physics beyond the standard model were observed. In 2015, the LHC successfully restarted at an even higher center-of-mass-energy,  $\sqrt{s} = 13$  TeV. The increase in energy corresponded to a large jump in discovery potential, but evidence for SUSY has still eluded us.

Figure 2.8 shows the excluded mass regions for EWK gauginos, squarks, and gluinos from a wide variety of  $\sqrt{s} = 13$  TeV CMS analyses [31]. Gluino masses below approximately 1800 GeV and squark masses below approximately 1000 GeV have been excluded at the 95% confidence level. Figure 2.8 does not contain any results interpreted in the context of GGM models. The difference between the orange and blue bars illustrate the effect of nearly tripling the size of the data set, from 12.9  $\text{fb}^{-1}$  to 35.9  $\text{fb}^{-1}$ . See Reference [60] for a full overview of SUSY limits from both colliders and precision measurements.



#### 2.4.1 Exclusion contours for GGM in the diphoton final state

Both CMS and ATLAS have published searches for GGM models in the diphoton final state, both at 8 TeV [8, 24] and at 13 TeV [6, 15]. Figure 2.9 shows the 95% confidence level limits in the gluino versus neutralino and the squark versus neutralino mass planes using  $2.3 \text{ fb}^{-1}$  of 13 TeV data collected with the CMS detector in 2015. Gluino masses below 1.65 TeV and squark masses below 1.37 TeV have been excluded. The simplified models used for the interpretation of results in Figure 2.9 will be described in more detail in Section 9.2.

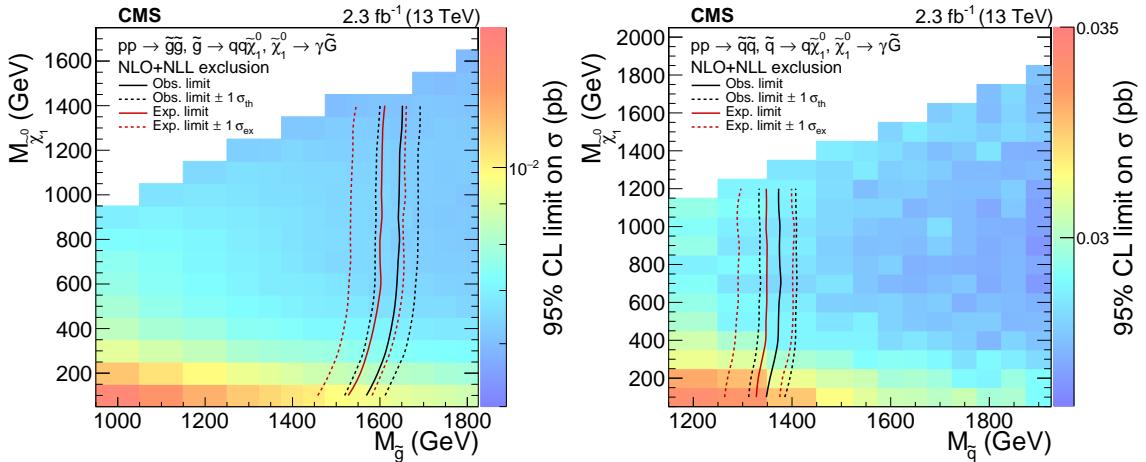


Figure 2.9: The 95% CL upper limits on the gluino (left) and squark (right) pair production cross sections as a function of neutralino versus gluino or squark mass. The contours show the observed and median expected exclusions with their one standard deviation uncertainties. Reprinted from [15].

The 2015 ATLAS results are shown in Figure 2.10. Limits were placed only on the masses of particles in a gluino pair production model. Gluino masses below 1.65 TeV were excluded at a 95% confidence level. The search strategy of the ATLAS analysis

is similar to the CMS strategy. The primary difference is that ATLAS defines their signal regions so that the expected SM background is close to zero. This is the source of the asymmetric expected limits in Figure 2.10. The CMS methodology as applied to the 2016 data set is the main focus of this dissertation and will be described in detail, particularly in Chapter 8.

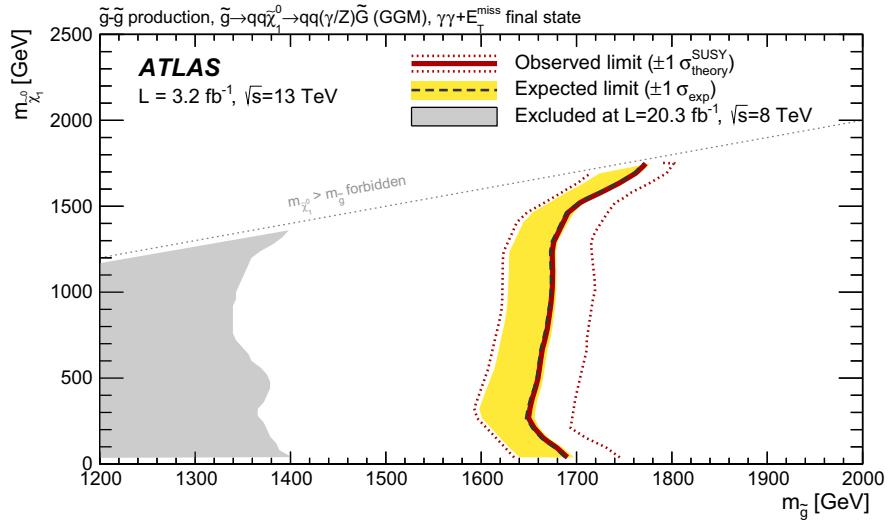


Figure 2.10: 95% confidence level exclusion contours in the neutralino-gluino mass plane derived using  $3.2 \text{ fb}^{-1}$  of data collected with the ATLAS detector in 2015. The expected limits are shown in black, with a yellow band representing  $\pm 1\sigma$ , and the observed limits are shown in red, with dotted red lines representing  $\pm 1\sigma$ . The gray area is the region previously excluded by 8 TeV ATLAS analyses. Reprinted from [6].

## CHAPTER 3

### THE LARGE HADRON COLLIDER

#### 3.1 Introduction

The Large Hadron Collider (LHC) is a particle accelerator at CERN, the European Organization for Nuclear Research. It is the world's largest accelerator and hosts several experiments, including the all-purpose Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS) experiments, the  $b$ -physics experiment LHC beauty (LHCb), and the heavy-ion experiment A Large Ion Collider Experiment (ALICE). The LHC is located 40 to 170 m underground in the countryside outside Geneva, Switzerland on the Switzerland-France border. A full description of the collider can be found in Reference [43].

The LHC consists of two counter-rotating beams of protons that collide at the four interaction points (IP) of the experiments listed above. It can also be used for heavy-ion collisions, in particular  $^{208}\text{Pb}$ - $p$  or  $^{208}\text{Pb}$ - $^{208}\text{Pb}$  collisions. In Run I of the LHC (2010 - 2012), a maximum center-of-mass energy  $\sqrt{s} = 8$  TeV was achieved. After a two-year shutdown, the LHC restarted in 2015 at a center-of-mass energy  $\sqrt{s} = 13$  TeV, close to its design value of  $\sqrt{s} = 14$  TeV. Run II extends to the end of 2018, when another long shutdown will allow upgrades for the High-Luminosity LHC era to begin.

### 3.2 Injector chain

A series of smaller accelerators are required to bring protons from rest to the collision energy of 6.5 TeV. The full CERN accelerator complex is shown in Figure 3.1. The journey of a proton in the LHC begins in a small tank of hydrogen. The hydrogen gets ionized and is accelerated to 50 MeV in the Linear Accelerator 2 (Linac 2). Next it enters the Proton Synchrotron Booster and the Proton Synchrotron (PS), which increase the energy to 1.4 GeV and 25 GeV, respectively. The next step is the Super Proton Synchrotron (SPS), which is nearly 7 km in circumference and accelerates the protons to 450 GeV before injecting them into the LHC.

### 3.3 LHC tunnel and magnets

Due to the prohibitive cost of building a new tunnel, the LHC uses the same tunnel as the one previously occupied by the Large Electron-Positron (LEP) Collider. This placed several constraints on the design of the accelerator. Because of the large synchrotron radiation losses that occur when accelerating light particles such as electrons, the LEP tunnel included eight straight sections where the electrons and positrons could be accelerated. For a hadron collider such as the LHC, however, the losses due to synchrotron radiation are less than 7 keV per rotation. The straight sections of the tunnel mean that the arced sections have a smaller radius of curvature  $R$  than if the tunnel was perfectly circular. This directly affects the maximum attainable momentum  $p$  of a particle with charge  $q$  for a given magnetic field  $B$ :

$$B = \frac{p}{qcR} \quad (3.1)$$

To reach the design energy  $\sqrt{s} = 14$  TeV, the LHC dipoles must be able to operate at a magnetic field  $B = 8.33$  T. This requires the magnets to be superconducting. The LHC includes 1232 dipole magnets that are cooled to 1.9 K using liquid helium. The

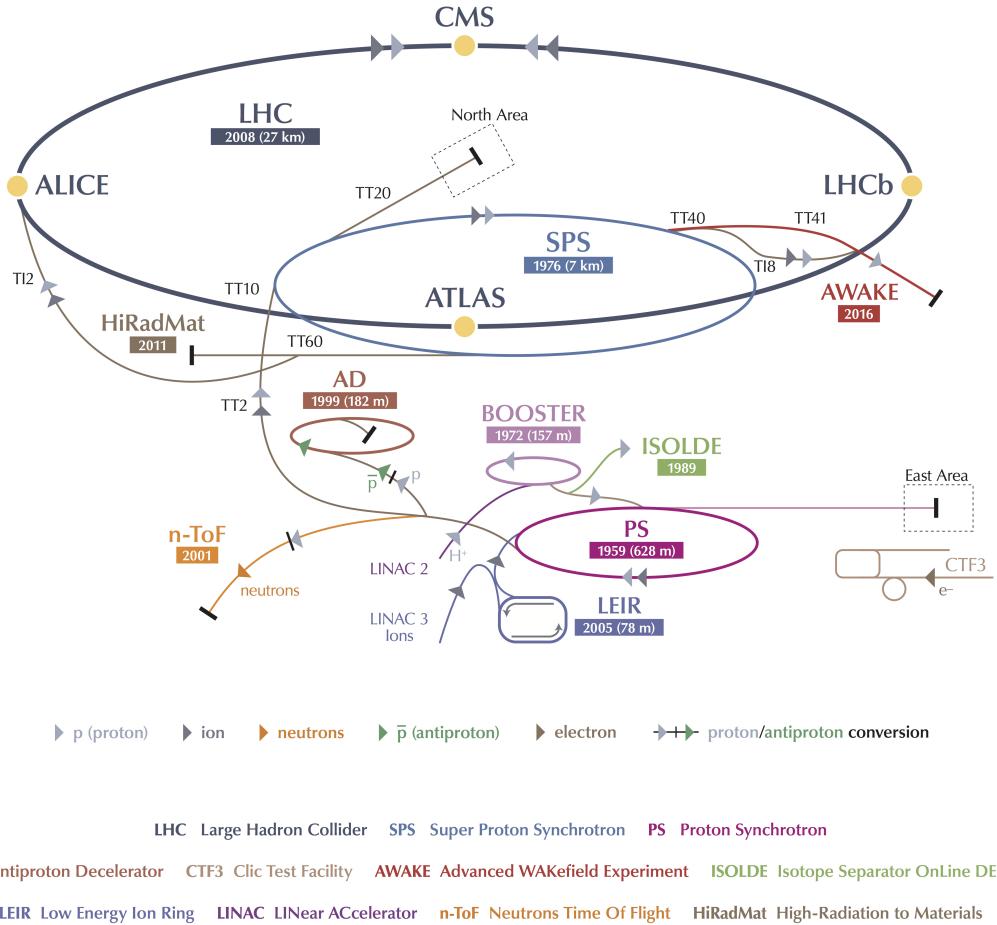


Figure 3.1: Schematic of the CERN accelerator complex. The LHC is shown in dark grey and includes four interaction points: one each for the CMS, ATLAS, LHCb, and ALICE experiments. Prior to being injected into the LHC, the protons are accelerated through the Linear Accelerator 2 (Linac 2), the Proton Synchrotron Booster, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). Reprinted from Reference [14].

magnetic core is made of stacked niobium-titanium (NbTi) filaments of thickness 6-7  $\mu\text{m}$ . The LEP tunnel is too narrow (3.7 m radius) to fit two magnet systems and the associated cryogenics and insulation. Therefore, the LHC magnets use a “twin-bore” design, where both of the beams are embedded in a single magnet system.

The LHC has only one accelerating sector. In the sector, there are eight superconducting RF cavities per beam that operate at a frequency of 400 MHz. The RF

cavities increase the energy of the protons by 485 keV per rotation. The beams consist of “bunches” of protons with a spacing of 25 ns separating consecutive bunches. Each collision between two bunches is referred to as a bunch crossing.

### 3.4 Machine luminosity

The number of events  $N_{\text{event}}$  produced every second for a particular process with cross section  $\sigma_{\text{event}}$  is given by

$$N_{\text{event}} = L\sigma_{\text{event}} \quad (3.2)$$

where  $L$  is the instantaneous luminosity of the collider.

The instantaneous luminosity depends only on machine parameters:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta_*} F \quad (3.3)$$

The variables in the above equation include the number of particles per bunch  $N_b$ , the number of bunches per beam  $n_b$ , the revolution frequency  $f_{\text{rev}}$ , and the relativistic factor  $\gamma_r$ . For the LHC,  $n_b$  is 2808 for the nominal bunch spacing of 25 ns. The number of protons per bunch  $N_b$  is limited to  $1.15 \times 10^{11}$  by the nonlinear beam-beam interactions between protons at each IP.

The normalized transverse beam emittance  $\epsilon_n$  is a measure of the average spread of particles in the beam in position-momentum phase space. In the transverse direction the beam is assumed to have a Gaussian shape with width  $\sigma$ . The beta function is then given by  $\beta = \sigma^2/\epsilon$ , and  $\beta_*$  refers to the value of the beta function at the interaction point. Finally,  $F$  is a geometric factor that depends on the crossing angle at the IP.

The design instantaneous luminosity of the LHC is  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . This goal is only

attainable because the LHC is a  $p$ - $p$  collider, rather than a  $p$ - $\bar{p}$  collider. During the 2016 data-taking period, the LHC met and surpassed its luminosity target, achieving a maximum value of  $1.53 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . This is almost double the maximum value of  $7.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$  that was achieved during Run I. The peak luminosity per day in 2016 is shown in Figure 3.2. The measurement of instantaneous luminosity is described in more detail in Section 4.8.

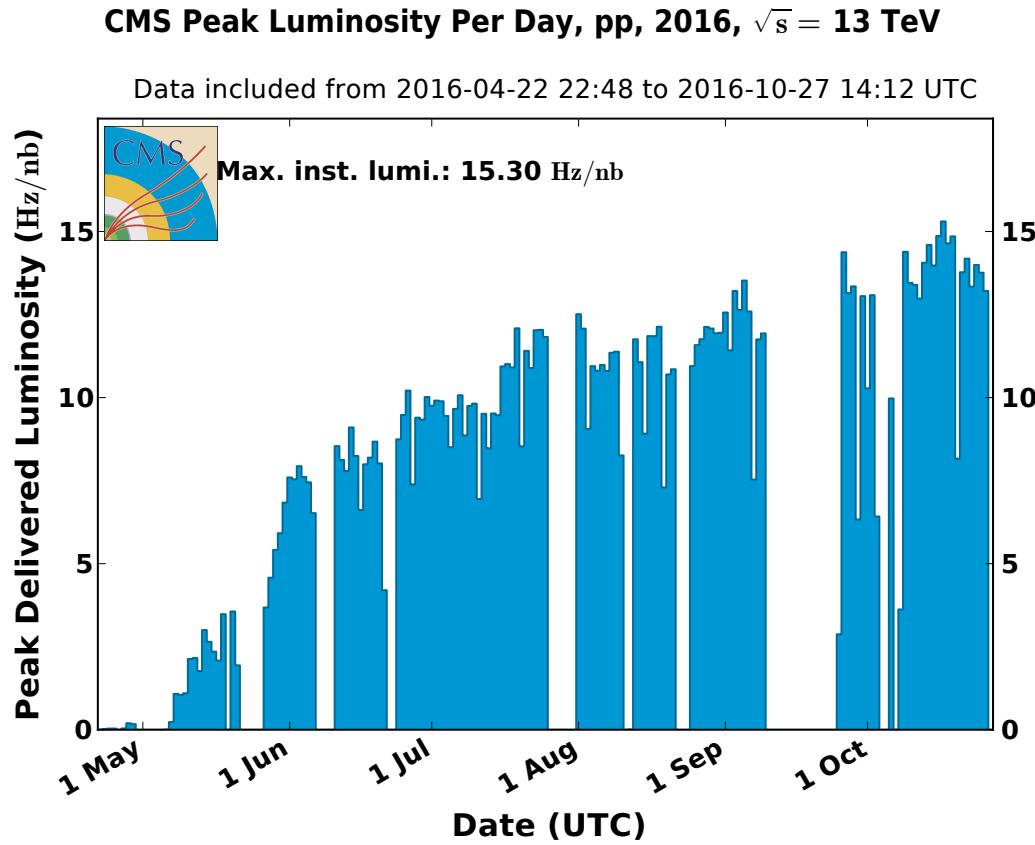


Figure 3.2: Peak instantaneous luminosity per day achieved by the LHC during the 2016 data-taking period. The maximum luminosity was 15.3 Hz/nb, or  $1.53 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ .

## CHAPTER 4

### CMS DETECTOR

The Compact Muon Solenoid (CMS) detector is a multi-purpose detector designed to accurately measure the energy and momentum of all particles produced in proton-proton or heavy ion collisions. Figure 4.1 shows a schematic of the detector as a whole. The CMS detector is 21.6 m long, 14.6 m in diameter, and weighs 12500 t. Moving radially outward from the interaction point, the sub-detectors are the silicon pixel and strip tracker (Section 4.3), the electromagnetic calorimeter (Section 4.4), the hadron calorimeter (Section 4.5), and the muon system (Section 4.7). For a full description of the CMS detector, see Reference [18].

#### 4.1 Coordinate system

The origin of the CMS detector coordinate system is located at the nominal collision point. The  $z$ -axis is oriented along the beam direction, with the positive  $z$ -axis pointing in the counter-clockwise direction when viewing the LHC from above. The  $y$ -axis points vertically upward, and the  $x$ -axis points toward the center of the LHC. The  $xy$ -plane is referred to as the transverse plane.

Due to the nature of particle collisions, however, Cartesian coordinates are often not the most convenient. Because protons are not elementary particles, it is actually the individual quarks or gluons within the proton that interact during the collision. This means that the collision will not be at rest in the lab frame, but will have some non-zero velocity along the  $z$ -axis. To deal with this, it is beneficial to use coordinates that are invariant under boosts in the  $z$ -direction. CMS follows the particle

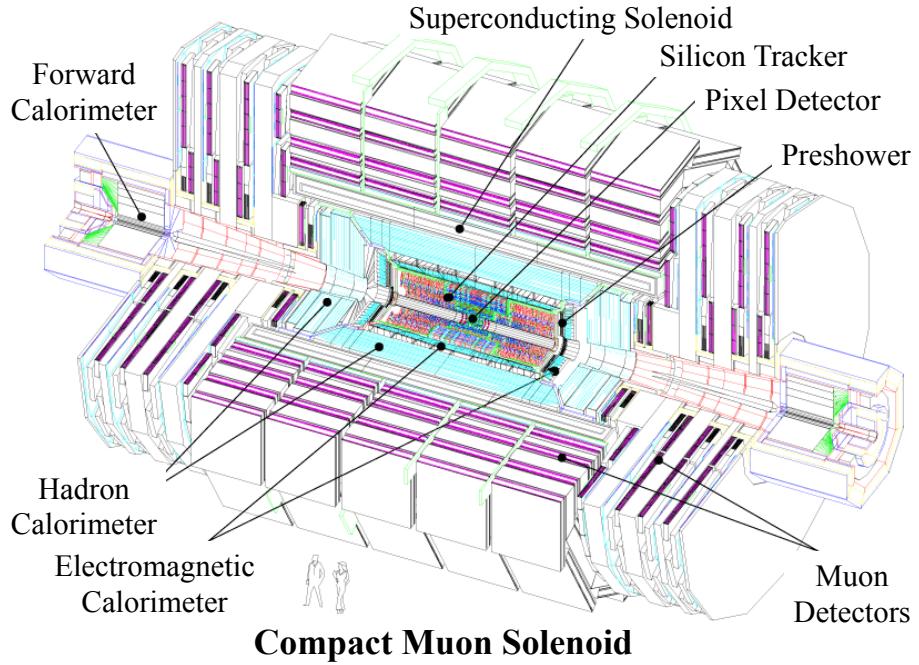


Figure 4.1. Schematic of the CMS detector.

physics convention of describing the position of a particle in terms of its transverse momentum, azimuthal angle, and pseudorapidity. The transverse momentum  $p_T$  is defined as the magnitude of the momentum in the  $xy$ -plane. The azimuthal angle  $\phi$  is defined in the transverse plane, with  $\phi = 0$  corresponding to the positive  $x$ -axis. Finally, the pseudorapidity is defined as  $\eta = -\ln \tan(\theta/2)$ , where the polar angle  $\theta$  is measured from the  $z$ -axis. Distances between particles are generally expressed using the  $\Delta R$  variable, where  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ .

## 4.2 Superconducting solenoid

One of the most important components of the CMS detector is the superconducting solenoid that provides the bending power necessary to precisely measure the momentum of all charged particles produced in the collision. The solenoid is a 4-layer niobium-titanium coil embedded in aluminum and aluminum alloy. The magnet is

located between the calorimeters and the muon system. It is 12.5 m long and has an inner diameter of 6 m. It is capable of producing magnetic fields up to 4 T, although the magnet is generally operated at 3.8 T to prolong its lifetime. At full current, the magnet has a stored energy of 2.6 GJ. A 12,000 ton steel yoke made up of 5 wheels in the barrel and 3 endcap disks serves to return the magnetic flux. The solenoid is suspended in a vacuum cryostat and cooled to 4.5 K with liquid helium. A detailed description of the CMS magnet can be found in Reference [34]. Figure 4.2 shows the calculated magnetic field in a longitudinal slice of the CMS detector.

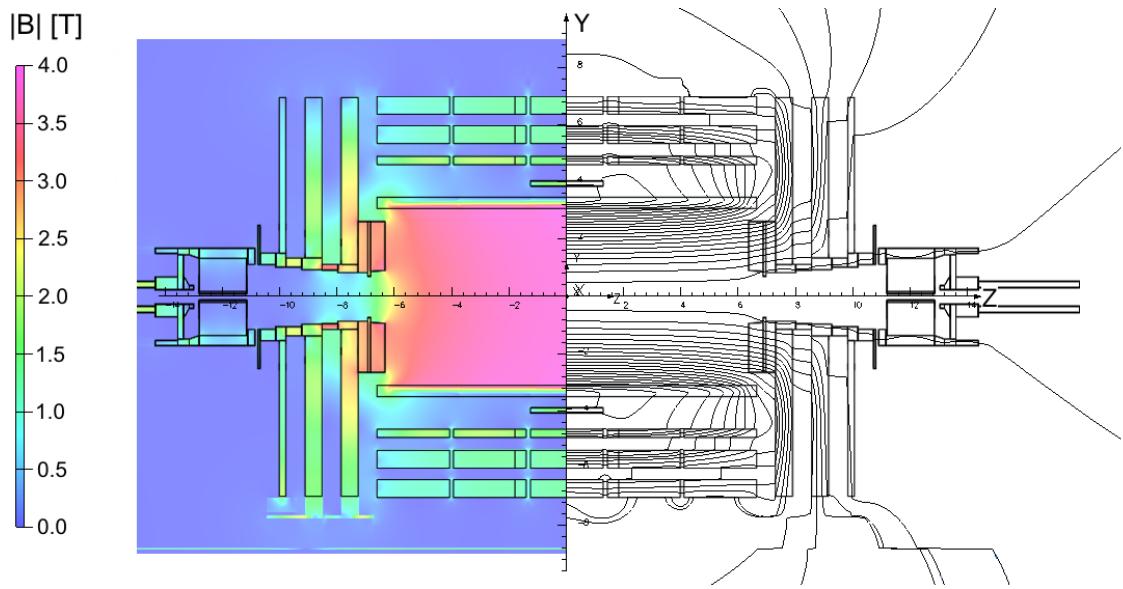


Figure 4.2: Calculated magnetic field  $|\vec{B}|$  for a longitudinal slice of the CMS detector when operated at a central magnetic flux density of 3.8 T. In the right-hand portion of the figure, each magnetic field line represents a magnetic flux step of 6 Wb. Reprinted from Reference [17].

### 4.3 Tracker

The innermost sub-detector is the silicon tracker [36, 37]. The tracker must provide enough information to accurately reconstruct the trajectories of charged particles to a high level of precision. This is accomplished through the use of silicon semiconductors, which rely on the properties of p-n junctions to detect charged particles. A p-n junction is created by bringing a p-type semiconductor into contact with an n-type semiconductor. For p- and n-type semiconductors, the crystal lattice of the semiconductor is doped to either donate or accept extra electrons, respectively. In a p-n junction, the extra electrons in the n-type semiconductor migrate and combine with the electron holes in the p-type semiconductor. This creates a depletion region in the center of the crystal. Charged particles passing through the depletion region will create electron-hole pairs that move towards either end of the junction when an external electric field is applied. The collected charge is proportional to the energy deposited in the detector.

The overall dimensions of the tracker are 5.6 m in length and 1.2 m in radius. The full tracking system is cylindrical in shape and is comprised of a barrel and two endcaps, each of which is split into layers of silicon pixel detectors and layers of silicon micro-strip detectors. In total, there are 48 million  $150 \times 100 \mu\text{m}$  pixels and 9.6 million strips that are between 80 and 180  $\mu\text{m}$  wide. Figure 4.3 shows a schematic drawing of the layout of the tracker subsystems. For charged hadrons with  $p_T$  less than 20 GeV, the  $p_T$  can be measured with a resolution of 1%.

#### 4.3.1 Pixel detectors

Due to the high occupancy of the tracker within 10 cm of the beam pipe, pixel detectors rather than strip detectors are used as the innermost layers of the tracker. Three layers in the barrel and two disks of pixel detectors in the endcap give three high-precision points for every charged particle moving away from the interaction

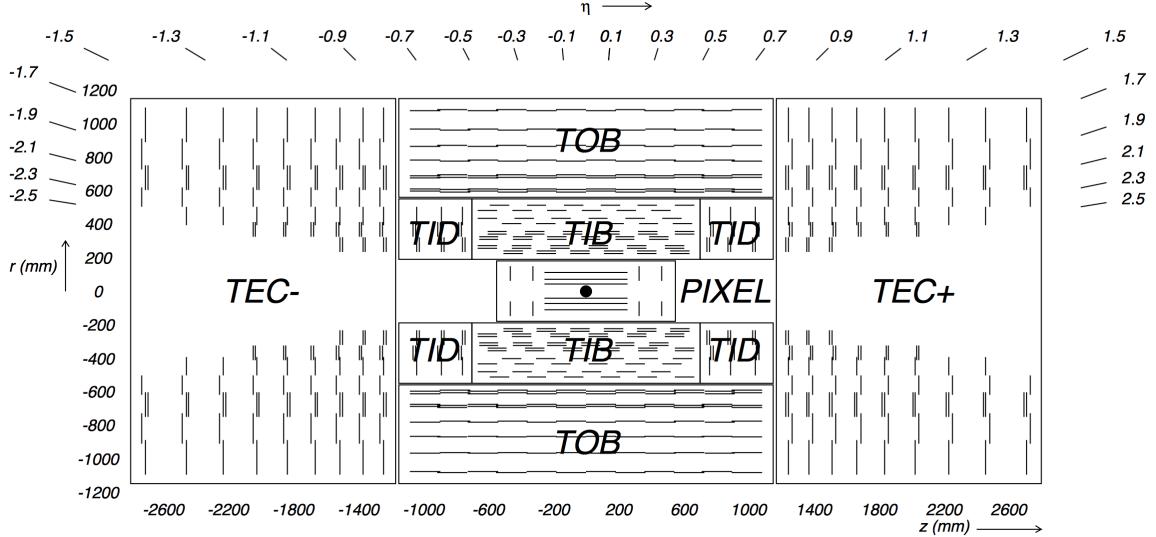


Figure 4.3: Schematic of the CMS tracker system. Detector modules are represented by single lines, and back-to-back modules are represented by double lines. The strip tracker is further divided into the Tracker Inner/Outer Barrel (TIB/TOB), Tracker Inner Disk (TID) and the Tracker Endcap (TEC). Reprinted from Reference [18].

point. The small pixel size of  $150 \times 100 \mu\text{m}$  is critical for accurate secondary vertex reconstruction, for forming seed tracks used for high level triggering (see Chapter 5), and for the reconstruction of charged particles in the event (see Chapter 6). The three barrel layers are 53 cm long and are located at mean radii of 4.4, 7.3 and 10.2 cm. The endcap modules extend radially from 6 to 15 cm and are placed on each side at  $z = \pm 34.5$  and  $z = \pm 46.5$  cm. The endcaps extend the coverage of the sub-detector to  $|\eta| < 2.5$ .

#### 4.3.2 Strip detectors

Moving radially beyond the pixel detectors are the silicon strip trackers. These operate on the same basic principles as the pixel detectors, but each silicon strip is  $10 \text{ cm} \times 80 \mu\text{m}$ , giving a precise position measurement along one dimension only. In the barrel, the strips run parallel to the  $z$ -axis with a pitch, or spacing between strips, between 80 and 183  $\mu\text{m}$ . Strips in the endcaps are aligned radially with a

pitch between 100 and 184  $\mu\text{m}$ . In total there are 10 layers of strip sensors in the barrel and 12 layers in the endcaps.

The tracker has a spatial resolution of 25-50  $\mu\text{m}$  perpendicular to the strip direction. In order to improve the precision of the detector in the direction parallel to the strips, several of the layers are arranged in pairs. By aligning the second layer 100 mrad off from the strips in the first layer, a spatial resolution of 230 to 530  $\mu\text{m}$  can be achieved. These back-to-back modules are represented by double lines in Figure 4.3.

#### 4.4 Electromagnetic calorimeter (ECAL)

Beyond the tracker is the electromagnetic calorimeter (ECAL), the most important sub-detector for this analysis [21, 22]. The ECAL measures the energies deposited by photons and electrons when they are stopped by the detector. For energies above 10 MeV, electrons lose energy primarily through the production of photons in bremsstrahlung. Energetic photons lose energy through the production of  $e^+e^-$  pairs [44]. Therefore, when an energetic electron or photon is incident on one of the crystals of the ECAL, the result is an electromagnetic cascade (“shower”) of photons and electrons with successively lower energies.

The CMS ECAL is designed to contain the full electromagnetic shower from electrons and photons with initial energies as high as a few TeV. It is a homogeneous calorimeter made with 75,848 scintillating lead tungstate  $\text{PbWO}_4$  crystals. The ECAL is divided into a barrel region (EB) covering  $|\eta| < 1.479$  and two endcap regions (EE) covering  $1.479 < |\eta| < 3.0$ . Each region includes a single layer of  $\text{PbWO}_4$  crystals. In the barrel, the crystals have a truncated pyramidal shape with a radial depth of 23 cm. The front face of the crystal has dimensions  $22 \times 22 \text{ mm}^2$ , and the rear face has dimensions  $26 \times 26 \text{ mm}^2$ . In the endcap, the crystals are 22 cm long and have a cross section of  $28.62 \times 28.62 \text{ mm}^2$  on the front face and  $30.0 \times 30.0 \text{ mm}^2$  on the rear face.

Lead tungstate is an inorganic scintillator. The passage of charged particles produce electron-hole pairs in the conduction and valence bands of the material, and light is emitted when the electrons return to the valence band. There are several properties of  $\text{PbWO}_4$  that make it an ideal choice for the scintillating material of the CMS ECAL. Its Molière radius, defined as the radius of a cylinder containing 90% of the shower's energy deposition, is only 2.2 cm, allowing for excellent position resolution and separation between showers. The latter is particularly important when trying to distinguish photons from isolated  $\pi^0 \rightarrow \gamma\gamma$  decays. The fast response time of lead tungstate is also critical. Approximately 80% of the light is emitted within 15 ns [44], making it possible for each particle to be assigned to the correct bunch crossing.

Another positive feature of  $\text{PbWO}_4$  is that it is dense enough ( $8.3 \text{ g/cm}^3$ ) that the ECAL can fit into a relatively compact area. The depth of material that is needed is quantified by looking at the radiation length of the material ( $\chi_0$ ). The radiation length is the average distance an electron needs to travel to reduce its energy by a factor  $1/e$ . The radiation length of  $\text{PbWO}_4$  is 0.89 cm, which means  $24.7 \chi_0$  fit in the 22 cm radial distance of each EB crystal.

Finally, the last characteristic of  $\text{PbWO}_4$  that put it ahead of other inorganic scintillators is its radiation hardness. Nuclear reactions caused by prolonged exposure to severe radiation conditions can cause defects in the crystals of inorganic scintillators. This affects the transparency of the crystals, leading to a degradation in the crystal response. This is shown in Figure 4.4 for the 2011, 2012, 2015, and 2016 data-taking periods. To correct for this effect, each crystal is illuminated with laser light. Part of the light is redirected to a silicon photodiode off the detector for a reference measurement. This correction is important both for long-term damage as illustrated in Figure 4.4, but also for the fast component of the radiation damage that changes the crystal response over the course of a single LHC fill.

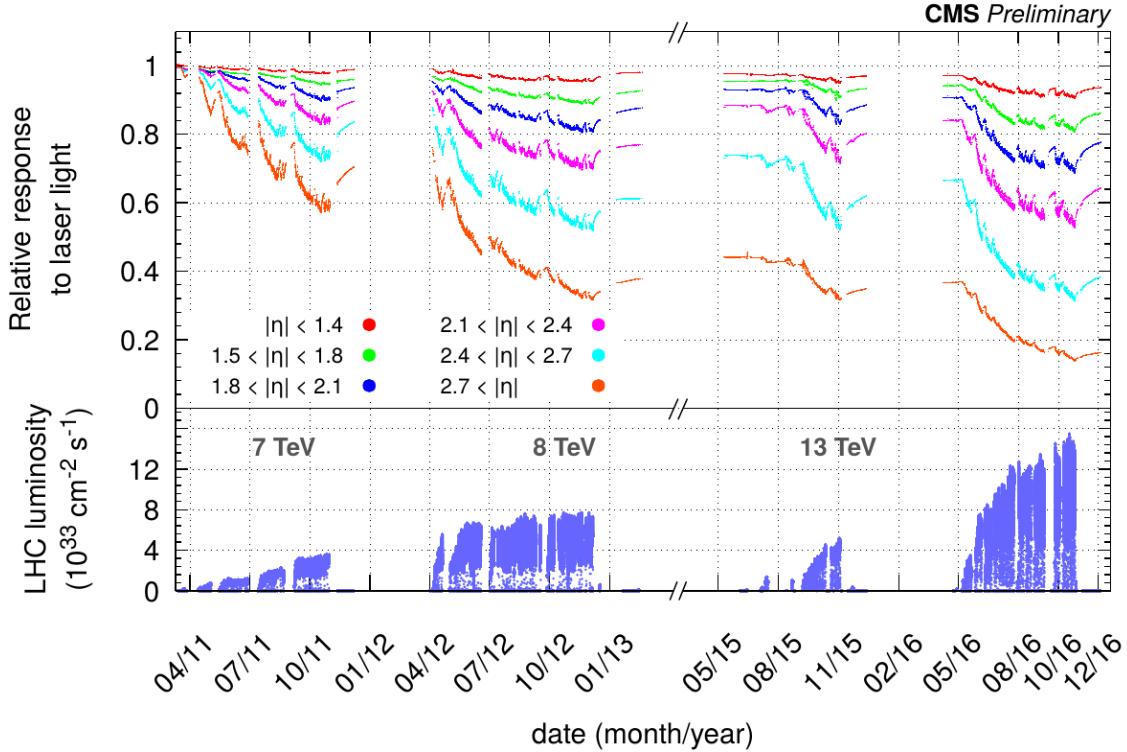


Figure 4.4: Relative response of the ECAL crystals for the 2011 to 2016 data-taking periods. The average observed change in response is up to 10% in the barrel and 50% for  $|\eta| < 2.5$ . The bottom plot shows the instantaneous luminosity delivered during this period. The crystals recovered some, but not all, of their response during Long Shutdown 1 (2013-2014). Reprinted from Reference [19].

The downside to  $\text{PbWO}_4$  is that it has a relatively low light yield of only 2 photoelectrons per MeV of incident energy. This makes the electronics used to collect the signal especially important. The crystal itself acts as an optical waveguide, and the scintillation light is internally reflected until it reaches the photodetectors glued directly onto the rear face of the crystal. Avalanche photodiodes (APD) are used in the barrel, and vacuum phototriodes (VPT) are used in the more radioactive endcap region. Both of these were chosen to be able to operate within the 3.8 T magnetic field, and both increase the gain by a factor of approximately 1000. The signal then goes to a Multi-Gain Preamplifier (MGPA), which dynamically changes the gain based on the energy of the incident particle. Temperature control of the  $\text{PbWO}_4$

crystals and the attached APDs is crucial to maintaining an excellent resolution. The crystal response and APD gain change by approximately 2.2% and 2.4% per  $^{\circ}\text{C}$ , respectively.

The last sections of the ECAL system are two pre-shower detectors located in front of the EE. The pre-shower detector is designed to improve the spatial resolution of the ECAL, so that the two photons from  $\pi^0 \rightarrow \gamma\gamma$  decays can be distinguished. Each pre-shower detector consists of alternating strips of lead and silicon detectors. The lead forces the photons to interact, and the silicon detector measures the electrons and positrons produced in the shower with high granularity. The pre-shower detector achieves a spatial resolution of 2 mm, compared to the 3 cm resolution of the EB and EE.

#### 4.5 Hadron calorimeter (HCAL)

The hadron calorimeter (HCAL) is the next sub-detector after the ECAL. It is a brass sampling calorimeter, with alternating layers of plastic scintillator and brass absorbers. Heavy particles such as hadrons interact with the brass layers, and the scintillation light from the nuclear showers is collected with wavelength-shifting optical fibers embedded in the plastic scintillator. The light is then guided to pixelated hybrid photodiodes and electronics that amplify the signal. Figure 4.5 shows a longitudinal slice of the full HCAL system.

There are 17 scintillator layers in the barrel of the HCAL (HB), which extends radially from 1.77 to 1.95 m and to a pseudorapidity  $|\eta| = 1.3$ . The first and last layers of absorber are made of steel rather than brass for structural reasons. The scintillation layers are divided into tiles with area  $\Delta\phi \times \Delta\eta = 0.87 \times 0.87$ . The tiles are further organized into towers that each get a single readout channel. The outer HCAL (HO) is located outside the solenoid and serves to increase the thickness of the calorimeter.

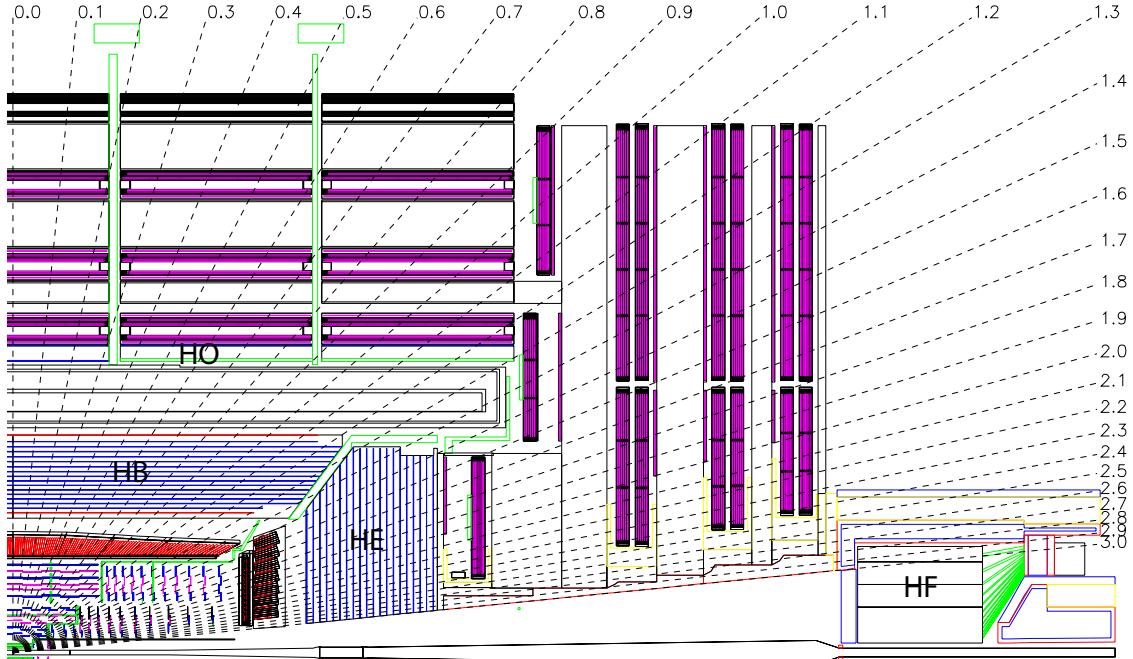


Figure 4.5: Longitudinal slice of the CMS HCAL, including the barrel HCAL (HB), outer HCAL (HO), endcap HCAL (HE), and forward HCAL (HF). The dashed lines represent lines of constant pseudorapidity. Reprinted from Reference [18].

The HCAL endcap (HE) covers the pseudorapidity range  $1.3 < |\eta| < 3$ . There are 17 layers of plastic scintillator and 18 layers of absorber in the HE. For  $|\eta| < 1.6$ , the tile size is the same as those in the barrel, and for  $|\eta| > 1.6$  the tile size is reduced to  $\Delta\phi \times \Delta\eta = 0.17 \times 0.17$ . Depending on the pseudorapidity, there are two or three readout channels per tower.

The last part of the HCAL system is the forward HCAL (HF) that covers the pseudorapidity range from  $3 < |\eta| < 5.2$ . Charged particles are detected via Cherenkov radiation in quartz fibers planted in the steel absorber. The quartz fibers were chosen to be able to withstand the high radiation levels of particles emitted close to the beam pipe. All three segments of the HCAL are calibrated using radioactive sources  $^{136}\text{Cs}$  or  $^{60}\text{Co}$  mounted on the tip of a moving wire. The radioactive sources produce photons at known energies, allowing for an absolute calibration of each scintillator.

Because most of the energy from the nuclear showers is deposited in the absorbers

rather than the scintillation material, the HCAL naturally has a lower energy resolution than the ECAL . In addition, nuclear showers might start before particles even reach the HCAL, or charged particles might deposit energy in the ECAL via bremsstrahlung. The mean distance traveled by a hadronic particle before undergoing an inelastic nuclear interaction is known as the nuclear interaction length. The radial depth of the ECAL corresponds to 1.1 nuclear interaction lengths, and in the HCAL there are 5.82 interaction lengths at  $|\eta| = 0$  and 10.6 interaction lengths at  $|\eta| = 1.3$ . The resolution of the calorimeters is described in more detail below.

At the analysis level, the objects of interest are not individual hadrons, but objects known as jets. A jet is a spray of particles in a narrow cone that is produced by the hadronization of quarks and gluons. The reconstruction of jets is described in detail in Chapter 6.

#### 4.6 Calorimeter performance

In general, the energy resolution  $\sigma/E$  of a calorimeter can be modeled with the following equation where  $\oplus$  is used to represent summing in quadrature [44]:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (4.1)$$

The first term is a stochastic term that takes into account random fluctuations in the amount of deposited energy for a particle with incident energy  $E$ . Homogeneous detectors like the ECAL have an excellent intrinsic resolution and a very small stochastic term. Sampling calorimeters such as the HCAL, on the other hand, have larger stochastic terms because the number of charged particles that hit the active layers varies from shower to shower.

The  $b/E$  term represents electronic noise from the equipment that is used to collect the signal. This term can depend sensitively on the temperature of the elec-

tronics, as already mentioned for the EB APDs. Maximizing the energy yield helps improve this contribution to the energy resolution. The final term is a constant and covers any irregularities in the detector response. This can include variations between ECAL crystals, radiation damage effects, or differences in response due to temperature gradients across the detector.

The resolution of the ECAL as measured using  $Z \rightarrow ee$  events in the first  $2.5 \text{ fb}^{-1}$  of 13 TeV data is shown in Figure 4.6. For the central region of EB, the electron energy resolution is better than  $2\%/\sqrt{E}$ .

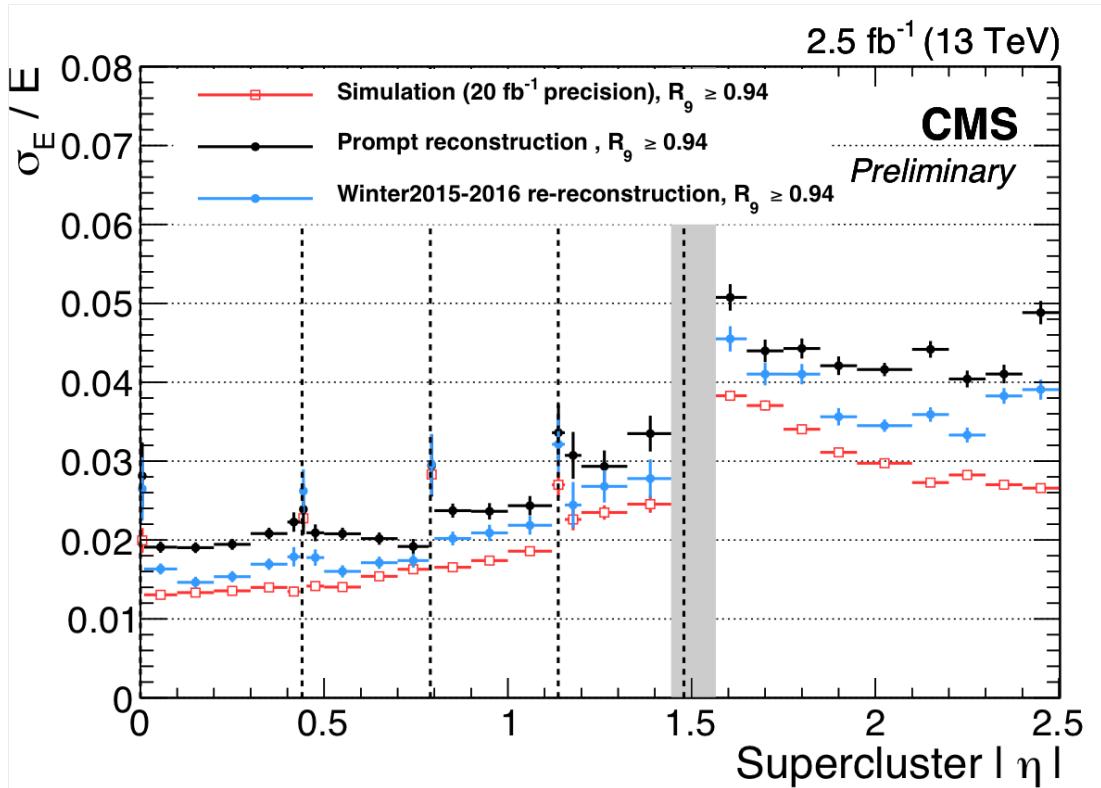


Figure 4.6: Relative electron energy resolution in bins of pseudorapidity. The energy resolution was calculated from an unbinned likelihood fit to  $Z \rightarrow ee$  events. Barrel module boundaries are shown with vertical dashed lines and often correspond to regions where the resolution is somewhat degraded. The gray band at  $|\eta| = 1.5$  represents the boundary between EB and EE. The data correspond to  $2.5 \text{ fb}^{-1}$  collected with the CMS detector in 2015. Reprinted from Reference [20].

The combined energy resolution of the ECAL and HCAL can be expressed as the following [27]:

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E}} \oplus 9\% \quad (4.2)$$

This resolution was measured using a pion test beam. The jet energy resolution after applying the full event reconstruction will be described in more detail in Chapter 6.

#### 4.7 Muon system

Aside from weakly-interacting particles such as neutrinos, muons are the only particles that make it past the HCAL. Muons do not interact via the strong force and are too heavy to be stopped by electromagnetic interactions alone. For this reason, the muon detector system is the outermost layer of CMS. Three different types of detectors are used in the muon system: drift tube (DT) chambers, cathode strip chambers (CSC), and resistive plate chambers (RPC). The muon system is located at a radial distance  $r > 3.5$  m and is embedded in the return yoke for the magnetic flux.

A drift tube consists of a conducting wire held at high positive voltage in the center of a gas cell. When a charged particle passes through the gas, the gas is ionized. The freed electrons are drawn to the positively charged wire, and in turn cause more ionization as they are accelerated. The time between the passage of the initial particle and the resulting electron avalanche can be used to reconstruct the position of the interaction perpendicular to the wire.

Drift tubes are used in the barrel region of the detector and extend to  $|\eta| < 1.2$ . The CMS drift tubes are made from aluminum plates and contain a mixture of 85% Ar and 15% CO<sub>2</sub>. A gold-plated stainless steel anode wire is located at the center of each tube. The wires are held at 3.6 kV and have a thickness of 50  $\mu\text{m}$ . Each drift tube has dimensions 1.3 cm  $\times$  4.2 cm  $\times$  2.4 m. Four layers of drift tubes

make up a superlayer, and a group of 2 or 3 superlayers makes up one muon chamber.

The endcap system uses cathode strip chambers because of their improved performance in high flux areas and non-uniform magnetic fields. Positively-charged wires are aligned perpendicular to negatively-charged copper strips in a gas, which gets ionized by passing muons. Electrons produce a signal in the anode wires and positive ions produce a signal in the cathode strips. This provides two position coordinates for each muon. The spatial resolution in the azimuthal direction is  $150 \mu\text{m}$ . The CSCs extend the pseudorapidity coverage of the muon system to  $|\eta| < 2.4$ .

Finally, resistive plate chambers are used in both the barrel and the endcap for more precise timing measurements. This is important for triggering purposes and to make sure each muon is assigned to the correct bunch crossing. A high voltage is applied to two large plates with a gaseous layer between them. The plates themselves have a high resistance, and the cascade caused by the passage of a muon through the gas is detected via readout strips outside the chamber. The CMS RPCs have a timing resolution of 1 ns.

#### 4.8 Instantaneous luminosity measurement

Five different detectors (luminometers) are used to measure the instantaneous luminosity  $L$  delivered by the LHC [33]: the pixel detector, the barrel drift tubes (DT), the forward hadron calorimeter (HF), and two specialized instruments, the Fast Beam Conditions Monitor (BCM1f) and the Pixel Luminosity Telescope (PLT). The last three use a fast readout system that is separate from the rest of the CMS system, and the pixel detector and DT use the standard data acquisition system.

Van der Meer (VdM) scans [67] are used to set an absolute calibration of each detector. Dedicated LHC running conditions allow the two beams to be scanned step-wise through one another in the two transverse directions. By finding the optimum beam position in the horizontal and vertical planes, the size of the beams at the

collision point can be determined. This in turn is used to calculate the visible cross sections  $\sigma_{vis}$  for each detector. The instantaneous luminosity is then given by the following, where  $R$  is the measured rate for a given luminometer:

$$L = \frac{R}{\sigma_{vis}} \quad (4.3)$$

The integrated luminosity  $\mathcal{L}$  is the integral of all collisions taking place within CMS and represents the total amount of data collected. Figure 4.7 shows the integrated luminosity delivered to CMS by the LHC and the integrated luminosity recorded by CMS during the 2016 data-taking period for 13 TeV  $p$ - $p$  collisions. CMS collected 92.5% of the  $40.76 \text{ fb}^{-1}$  delivered by the LHC. CMS can fail to record events delivered by the LHC if any sub-detector is off or if there is trigger deadtime. The 2016 “golden JSON”, the full list of runs and events that are certified for use in data analysis, corresponds to  $35.92 \text{ fb}^{-1}$ . The total uncertainty on the integrated luminosity is 2.5% [33]. The LHC is on track to deliver  $100 \text{ fb}^{-1}$  over the course of Run II.

### CMS Integrated Luminosity, pp, 2016, $\sqrt{s} = 13$ TeV

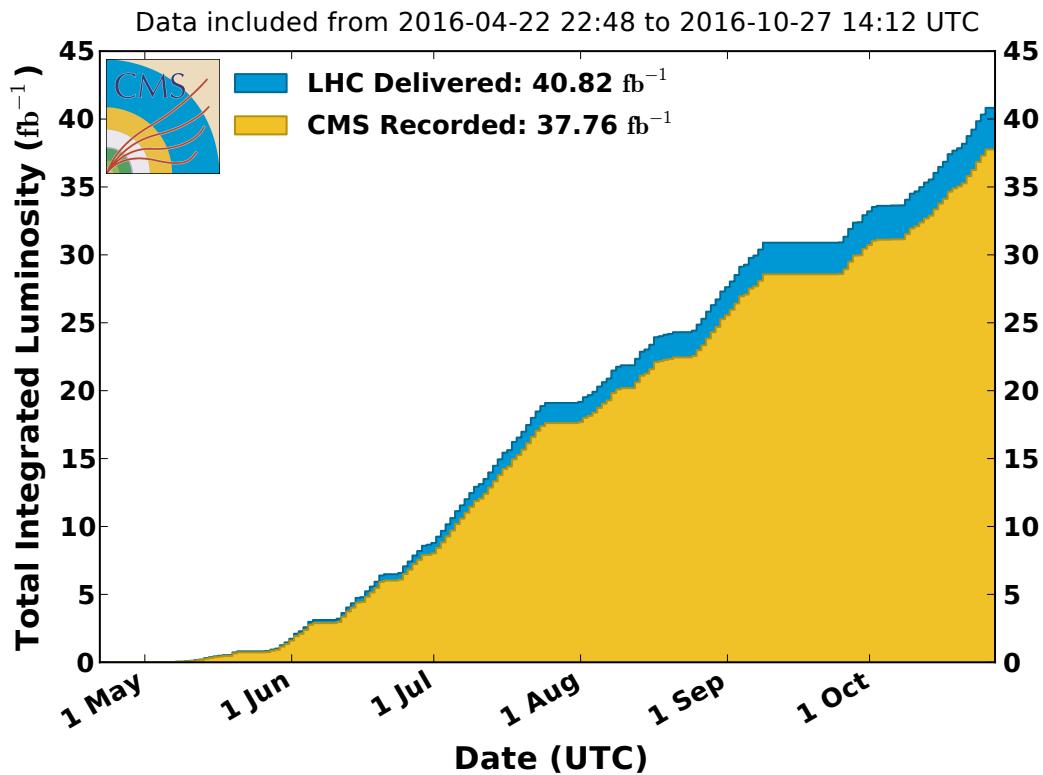


Figure 4.7: Integrated luminosity delivered to CMS by the LHC (blue) and the integrated luminosity recorded by CMS (orange) during  $p\text{-}p$  collisions at  $\sqrt{s} = 13$  TeV in 2016.

## CHAPTER 5

### TRIGGER SYSTEM

#### 5.1 Overview of the CMS trigger

Many more collisions occur within the LHC than can be stored and analyzed. At the four interaction points of the LHC, bunches of protons collide every 25 ns. For every bunch crossing, there are on average 25 “soft-scatter” collisions (pileup). This corresponds to an overall event rate of 1 GHz. Given that each event produces approximately 1 MB of data, it is impossible to store every event or even the majority of events. This has been especially true in recent years, with the LHC continually breaking new records for instantaneous luminosity.

To solve this problem, the CMS detector includes a sophisticated trigger system to determine which events get written to tape and eventually analyzed. The CMS trigger is divided into two steps: the hardware-based Level 1 (L1) trigger, and the software-based High Level Trigger (HLT). The L1 trigger reduces the rate from 1 GHz to approximately 100 kHz, and the HLT makes the final decisions necessary to reduce the rate to a few hundred Hz, the maximum amount that can be written and stored. The L1 trigger and the HLT are both highly configurable, which allows CMS to define and alter the thresholds as needed to suit different running conditions. Both steps of the CMS trigger are described in more detail below.

## 5.2 Level 1 trigger

The L1 trigger performs rough calculations of event parameters using field-programmable gate arrays located in the detector cavern. The L1 trigger only has  $3.8\ \mu\text{s}$  to make a decision about each event in order to reduce the output event rate to 100 kHz [25]. For inputs, the L1 trigger receives data from the ECAL, HCAL, and muon detector subsystems. The L1 trigger is naturally divided into two paths. The calorimeter trigger, shown schematically in Figure 5.1, takes data from the ECAL and HCAL subsystems and builds photon, electron, and jet candidates, as well as overall event quantities such as missing transverse momentum and total hadronic activity. The second half of the L1 trigger is the muon trigger, and its overall structure is shown in Figure 5.2. The output from the muon and calorimeter triggers goes into the Global Trigger (GT) processor [69], which makes the final decision regarding whether or not to accept the event.

### 5.2.1 Calorimeter trigger

As shown in Figure 5.1, the raw inputs for the calorimeter trigger are trigger primitives (TP) from the ECAL and the HCAL. The trigger primitives from the HCAL are divided between the HCAL Barrel and Endcap (HB and HE, respectively), and the forward HCAL (HF). Trigger primitives consist of coarse information about the energy deposits in the calorimeter for every bunch crossing. The ECAL is divided into trigger towers, groups of crystals corresponding to a region of approximately  $0.087 \times 0.087$  in  $\eta$  and  $\phi$ , and a TP is generated by summing the energy from each crystal in the tower. In the HCAL, TPs are generated for each trigger tower by summing the transverse energy in two consecutive time slices.

Layer-1 of the calorimeter trigger collects the TPs from the ECAL and HCAL. Its primary role is to distribute the data to one of nine Layer-2 nodes. Each Layer-2 node receives the full set of TPs for a particular bunch crossing, and identifies photon,

electron, jet, and  $\tau$ -tagged jet candidates using several dynamic clustering and local maxima finding algorithms. The outputs from the Layer-2 nodes are fed into a de-multiplexing (demux) node, which ranks the candidates by transverse momentum and sends the data to the Global Trigger.

### 5.2.2 Muon trigger

Information from all three of the muon detector subsystems—the cathode strip chamber (CSC) detectors, resistive plate chamber (RPC) detectors, and drift tubes (DT)—are combined in the L1 muon trigger to reconstruct muon candidates and their momenta. The muon trigger has three track-finding subsystems that reconstruct muons for different  $|\eta|$  regions: the barrel track finder ( $|\eta| < 0.83$ ), the endcap track finder ( $|\eta| > 1.24$ ), and the overlap track finder ( $0.83 < |\eta| < 1.24$ ).

Trigger primitives from the DT and RPC detectors are first combined in the TwinMUX system into “super-primitives”. This step serves to increase the precision of the position and timing of muon hits by taking advantage of the redundancy between the two subsystems. The barrel track finder uses the super-primitives as inputs for its track finding algorithm, processing twelve  $30^\circ$  wedges in parallel.

The endcap track finder receives information from the CSC detectors, and the overlap track finder uses information from all three subsystems. Both of these track finders use large look-up tables to convert specific hit patterns into  $p_T$  assignments for the muon candidates.

The outputs from all three track finders are sent to the Global Muon Trigger, which ranks the muons by quality and transverse momentum and sends the top 8 muon candidates to the Global Trigger for a final decision on the event.

### 5.2.3 Global Trigger

The reconstructed objects and quantities built by the calorimeter trigger and the muon trigger—photon, electron, jet, and muon candidates, missing energy and total hadronic activity—get sent to the Global Trigger (GT) for final processing. The GT can implement up to 512 different trigger algorithms using these objects [69]. For example, one of the L1 algorithms (or “seeds”) used in this analysis is DoubleEG\_22\_12. This path requires two electron or photon candidates with leading and trailing  $p_T$  greater than 22 and 12 GeV, respectively.

A full set of L1 algorithms is known as the L1 menu. The menu can be adjusted as needed to meet the requirements of the CMS physics program. In particular, the “prescale” of each L1 seed can be adjusted to take full advantage of varying LHC running conditions and instantaneous luminosities. A prescale is an integer value  $N$  used to reduce the rate of a trigger path by only applying the trigger to 1 out of every  $N$  events. For the 2016 data-taking period, the lowest unprescaled DoubleEG seed was DoubleEG\_22\_12, but the menu also included several prescaled algorithms with lower  $p_T$  thresholds.

## 5.3 High Level Trigger

The High Level Trigger (HLT) uses a large farm of commercially available PCs to further reduce the rate to a few hundred Hz. The HLT receives fully-built event data from the CMS data acquisition system (DAQ) and processes only those bunch crossings that have passed the L1 trigger. The 13,000 CPU cores used in the HLT run the CMS software framework referred to as CMSSW, the same framework that is used in the offline analysis.

Approximately 400 HLT “paths” are used to select events of interest. Each HLT path is a single set of criteria for accepting an event if it satisfies a particular physics

signature. The full list of HLT paths is referred to as the HLT menu. The HLT menu is often updated to reflect improvements in the software, updates to the calibrations, changes to the beam condition, or revisions to the physics signatures being sought.

The HLT has a limited amount of time it can spend making a decision on a single bunch crossing. For this reason, each HLT path is “seeded” by one or more L1 algorithms. When an event gets passed to the HLT from the DAQ, the HLT only processes those paths that are seeded by L1 bits that fired. For example, the diphoton HLT path used in this analysis is seeded by a combination of single and double e/ $\gamma$  (EG) L1 algorithms. If none of those particular L1 seeds fired, then the HLT path will not get run on that event.

In addition, significant time is saved at the HLT by running the steps (filters) of each HLT path in order from least CPU intensive to most CPU intensive. If the event fails any of the filters in an HLT path, then the processing is immediately aborted and the remaining filters do not get run. In the diphoton HLT path described below, the most CPU intensive filter is the invariant mass calculation, and therefore it runs only if the event has already satisfied the rest of the path’s requirements.

Based on its general physics signature, each HLT path is assigned to a primary data set (PD). For this analysis, signal events are included in the DoubleEG PD. The SingleElectron and SinglePhoton data sets are also used for object identification studies.

## 5.4 Analysis triggers

The HLT paths used in this analysis are listed in Table 5.1. These triggers were developed for the  $H \rightarrow \gamma\gamma$  search, but also serve our analysis well. Two triggers are used: a primary trigger that requires the diphoton invariant mass to be greater than 90 GeV, and a control trigger that was designed to collect  $Z \rightarrow ee$  events.

TABLE 5.1

## HLT TRIGGER PATHS

<b>Primary Trigger</b>
HLT_Diphoton30_18_R9Id_OR_IsoCaloId_AND_HE_R9Id_Mass90_v*
<b>Control Sample Trigger</b>
HLT_Diphoton30_18_R9Id_OR_IsoCaloId_AND_HE_R9Id_-
DoublePixelSeedMatch_Mass70_v*

## 5.4.1 Trigger requirements

The requirements to pass the various parts of the trigger are listed in Table 5.2. Because we only use photons with  $|\eta| < 1.4442$  in this analysis (see Chapter 7 for full event selection requirements), we list only the trigger requirements in the barrel.

Each of the variables used in the trigger are defined below. See Chapter 7 for more information on how energy deposits in the calorimeter are reconstructed into photon “superclusters”.

- $E_T$ : The transverse energy  $E_T$  of a photon is defined as the magnitude of the projection of the photon momentum on the plane perpendicular to the beam axis.
- $R_9$ : The variable  $R_9$  is a measure of the overall transverse spread of the shower. It is the ratio of the energy deposited in the ECAL inside a 3x3 crystal matrix centered on the most energetic crystal in the supercluster to the supercluster raw energy.
- $\sigma_{i\eta i\eta}$  : The shower width  $\sigma_{i\eta i\eta}$  is the log-fractional energy-weighted spread within the 5x5 crystal matrix centered on the crystal with the largest energy deposit in the

TABLE 5.2  
PRIMARY TRIGGER REQUIREMENTS

Name	Cuts
Diphoton30_18_-	Leading photon $p_T > 30$ GeV Sub-leading photon $p_T > 18$ GeV
R9Id_-	$R_9 > 0.85$
IsoCaloId_-	$\sigma_{in\eta} < 0.015$ ECAL isolation $< (6 + 0.012 \times \text{Photon } E_T)$ Track isolation $< (6 + 0.002 \times \text{Photon } E_T)$
HE_R9Id_-	$H/E < 0.1$ $R_9 > 0.5$
Mass90_-	$m_{\gamma\gamma} > 90$ GeV

supercluster. The symbol  $in\eta$  indicates that the variable is obtained by measuring position by counting crystals.

- ECAL isolation: The ECAL isolation is the sum of all energy deposits in the ECAL within a cone of  $\Delta R < 0.3$  centered on the photon.
- Track isolation: The track isolation is the sum of the energies of tracks in the tracker within a cone of  $\Delta R < 0.3$  centered on the photon.
- H/E: The ratio between the energy deposited in the HCAL tower closest to the supercluster position and the energy deposited to that supercluster in the ECAL is referred to as H/E.

All photons are required to pass the H/E and loose  $R_9$  cuts in  $_R9Id_-$ , and

either the tighter  $R_9$  cuts in `_R9Id_` or the isolation and shape cuts in `_IsoCaloId_`. The leading leg of the filter requires the photon candidate to be matched to an L1 seed. It can be matched to one of several SingleEG and DoubleEG L1 filters, but the largest contribution comes from the lowest unprescaled triggers: namely, SingleEG40 and DoubleEG\_22\_15. Both photons must satisfy the trailing filter, which is unseeded. In addition to the cuts listed above, the invariant mass of the diphoton system is required to be greater than 90 GeV.

The control trigger shares all of the same requirements as the primary trigger, with two exceptions: the invariant mass of the two electromagnetic objects is required to be greater than 55 GeV rather than 90 GeV, and both electromagnetic objects are required to be matched to a pixel seed. A pixel seed is defined as at least two hits in the pixel detectors that are consistent with the location of the energy deposit in the ECAL.

## 5.5 Trigger efficiency

An important input to the analysis is the overall trigger efficiency. Due to the similarity of the ECAL response to electrons and photons, the trigger efficiency can be calculated from  $Z \rightarrow ee$  events in data using the tag-and-probe method. In this method, two electron candidates are required. One serves as the “tag” and is required to pass loose photon identification criteria. The second electron candidate serves as the “probe” and has to satisfy the same selection criteria as our offline photon identification (see Section 7.3). In order to ensure a high purity of electromagnetic objects, the invariant mass of the di-electron system must be between 75 and 105 GeV. For this study, tag-and-probe events were collected by requiring that the tag pass a single electron control trigger, `HLT_Ele27_WPTight_Gsf`.

The efficiency  $\epsilon$  of the HLT path or trigger filter that is being studied is given by the following equation, where  $N_{total}$  is the total number of tag and probe pairs

passing all requirements, and  $N_{pass}$  is the number of tag and probe pairs in which the probe passes the trigger filter.

$$\epsilon_{trig} = N_{pass}/N_{total} \quad (5.1)$$

### 5.5.1 Efficiency of primary trigger

Because our analysis trigger is seeded by the OR of multiple SingleEG and DoubleEG L1 seeds, the names of which changed over the course of the 2016 data-taking period, calculating the L1 efficiency on its own proved to be tricky. Instead, it was simpler to calculate the total efficiency with which photon candidates pass both the L1 seed and the leading leg of the HLT path. This efficiency as a function of photon  $p_T$  is shown in Figure 5.3. The efficiency was fit to an error function to calculate the overall efficiency at the plateau. For photon  $p_T > 40$  GeV, the leading filter is 98.2% efficient.

Tag and probe objects for the trailing leg efficiency must pass the same set of requirements as those used in the leading leg efficiency calculation, with the additional requirement that the tag must pass the leading filter. This requirement arises from the way HLT modules are structured. Filters are processed sequentially, and if an event fails one filter, the subsequent filters are skipped. Figure 5.4 shows the efficiency of the trailing filter as a function of photon  $p_T$ . For  $p_T > 40$  GeV, the trigger is 99.8% efficient.

Finally, we calculated the efficiency of the trigger with respect to the diphoton invariant mass. For this calculation, we required two photons passing our analysis selection criteria, two photons satisfying the trailing leg of the trigger, and one photon passing the leading leg of the trigger. The efficiency was given by the number of diphoton events passing the full HLT path over the total number of diphoton events passing our requirements. The efficiency of the trigger as a function of invariant mass

is shown in Figure 5.5. For  $m_{\gamma\gamma} > 100$  GeV, the trigger is 99.4% efficient.

The efficiency of the trigger as a whole is the product of all three efficiencies. Two factors of the trailing leg efficiency are needed since both photons are required to pass that leg:

$$\epsilon_{trig} = \epsilon_{lead} \times \epsilon_{trail}^2 \times \epsilon_{mass} = 97.2\% \quad (5.2)$$

### 5.5.2 Efficiency of double electron trigger

In the secondary trigger listed in Table 5.1, the pixel seed requirement is applied only to the trailing leg of the trigger. This additional requirement results in a significantly lower overall efficiency for this leg. As shown in Figure 5.6, the trigger is only 90.4% efficient for  $p_T > 40$  GeV. Since all other cuts are the same between the two triggers, however, the leading leg efficiency is the same as that shown in Figure 5.3. This results in an overall trigger efficiency of 79.8% for the control trigger.

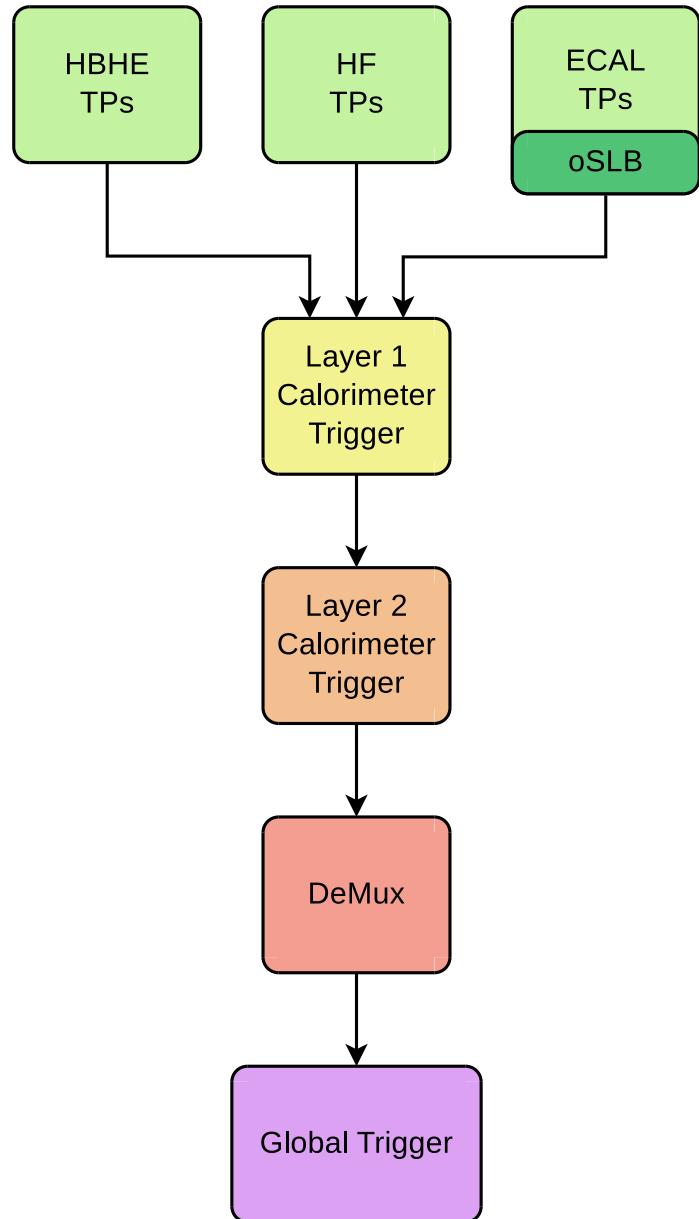


Figure 5.1: Schematic showing the overall structure of the calorimeter half of the L1 trigger. Reprinted from Reference [26].

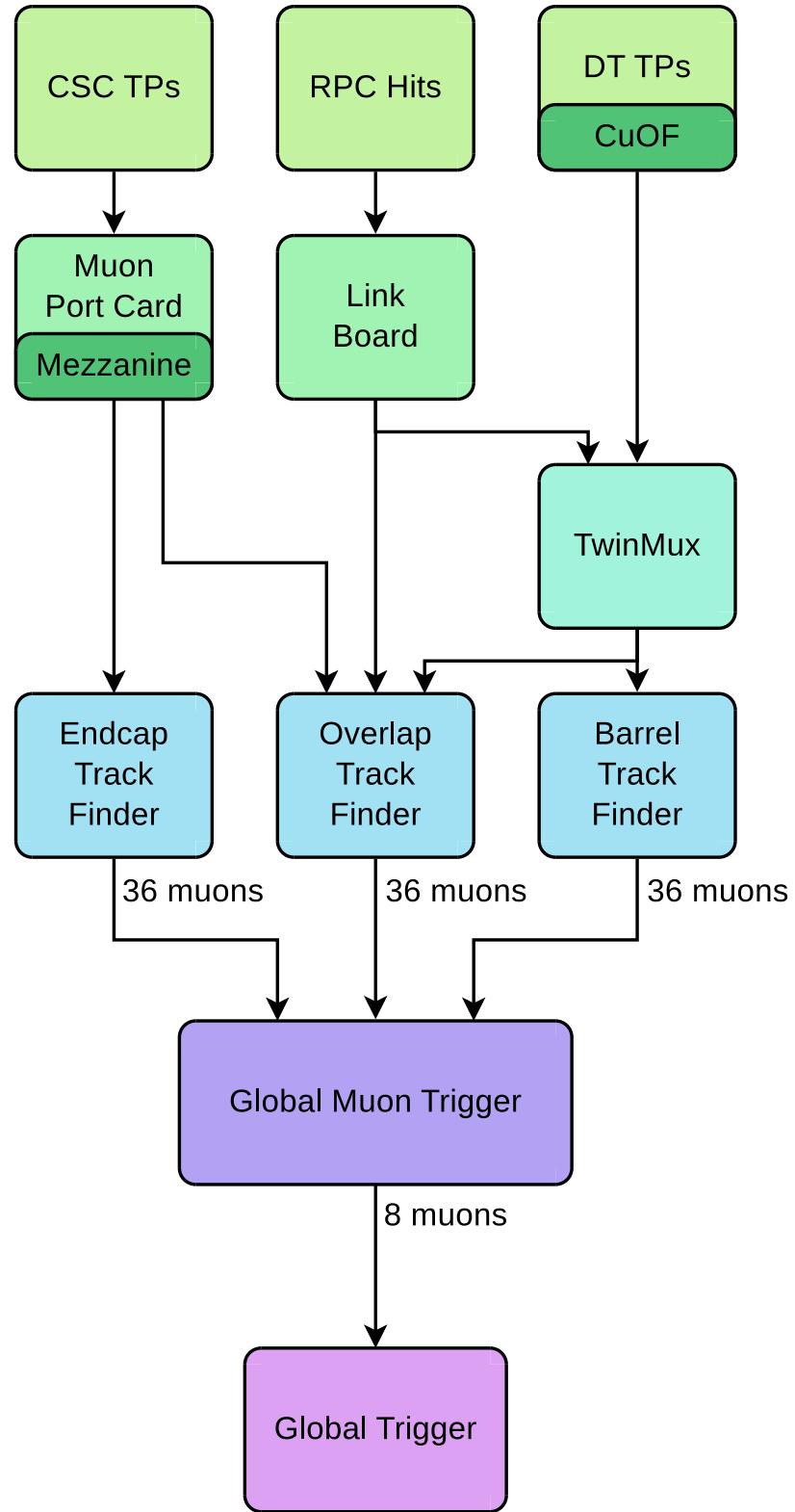


Figure 5.2: Schematic showing the overall structure of the muon half of the L1 trigger.  
Reprinted from Reference [26].

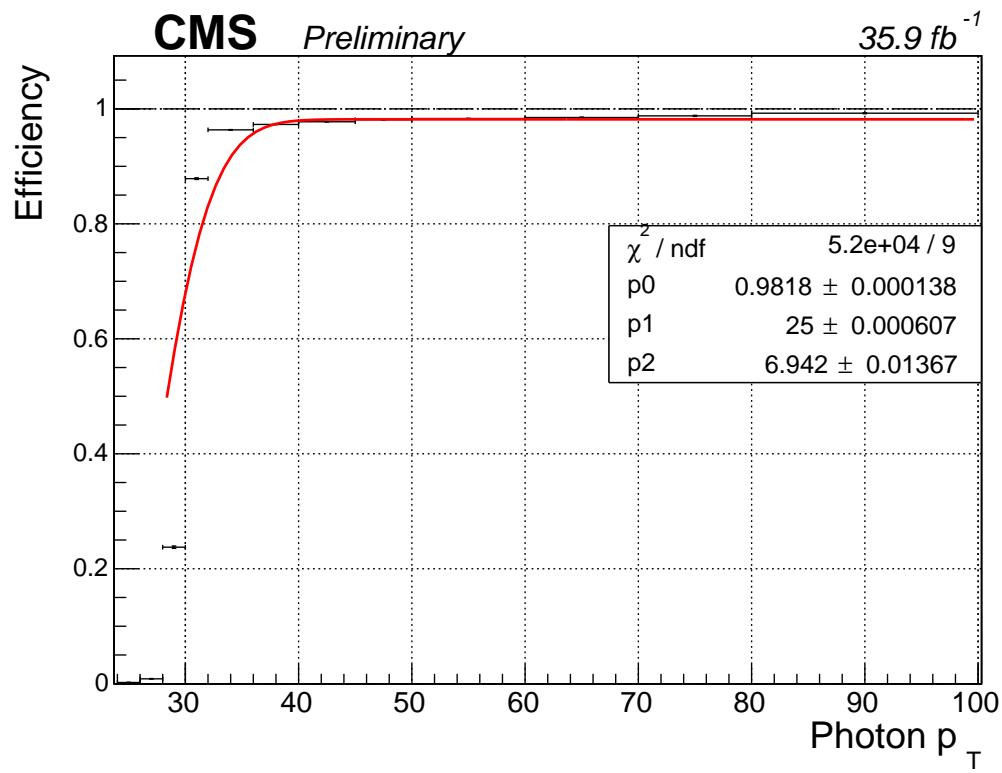


Figure 5.3: Efficiency of the L1 seed and the leading leg of the primary analysis trigger with respect to photon  $p_T$ . For  $p_T > 40$  GeV, the trigger is 98.2% efficient.

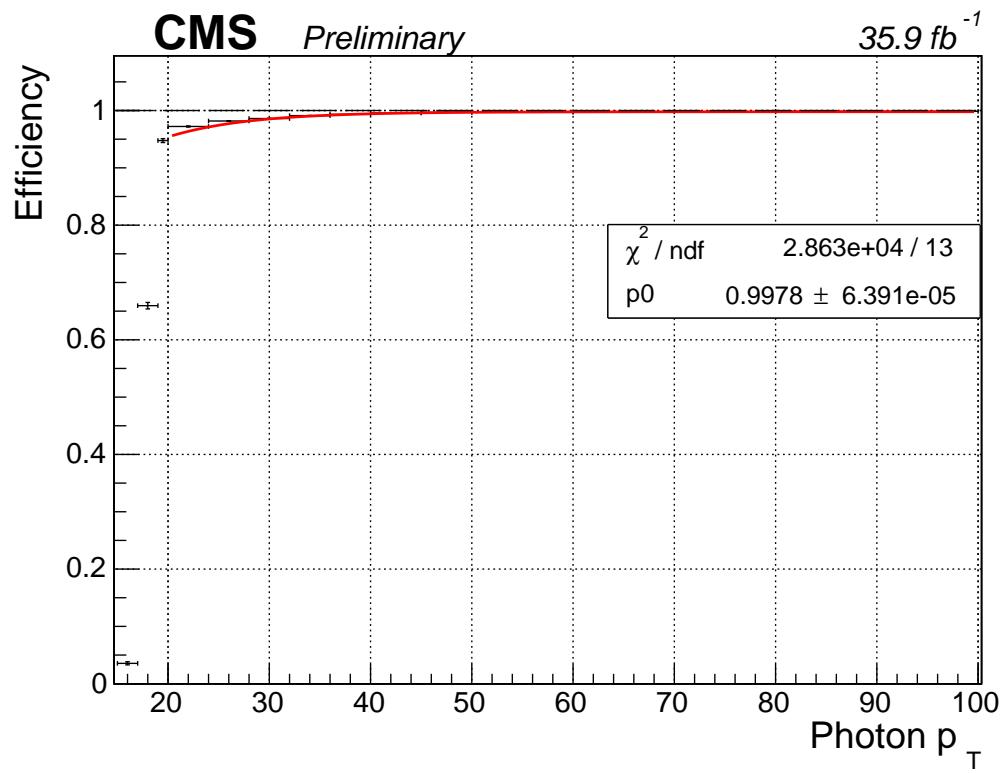


Figure 5.4: Efficiency of the trailing leg of the primary analysis trigger with respect to photon  $p_T$ . For  $p_T > 40$  GeV, the trigger is 99.8% efficient.

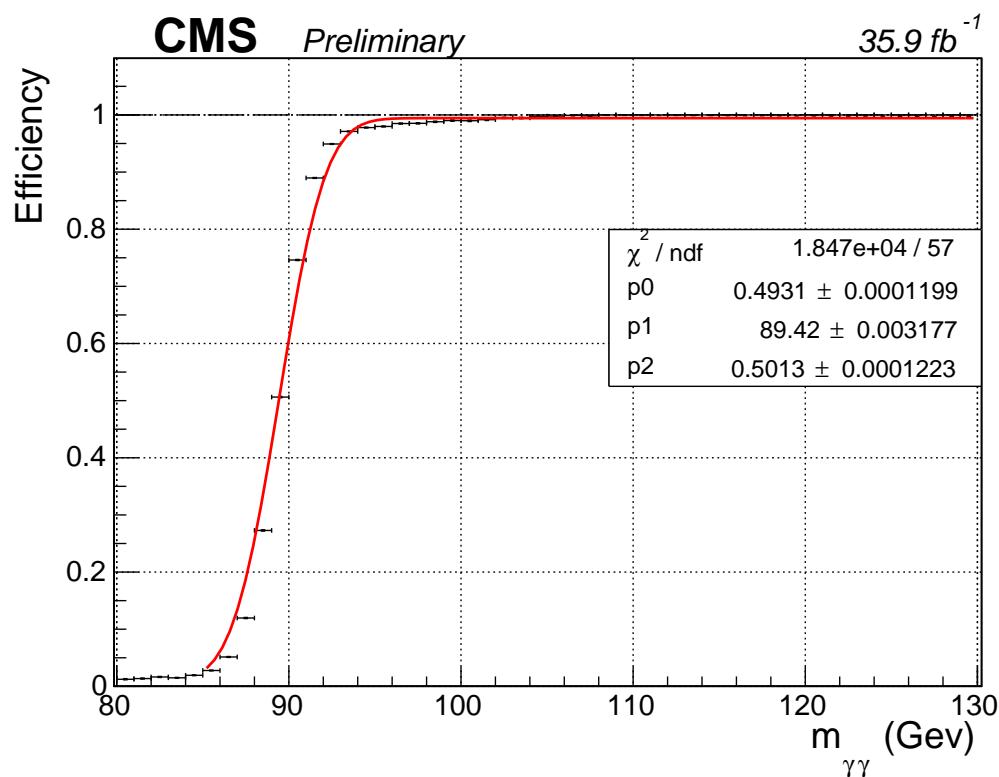


Figure 5.5: Efficiency of the primary analysis trigger with respect to the invariant mass of the diphoton system. For  $m_{\gamma\gamma} > 100 \text{ GeV}$ , the trigger is 99.4% efficient.

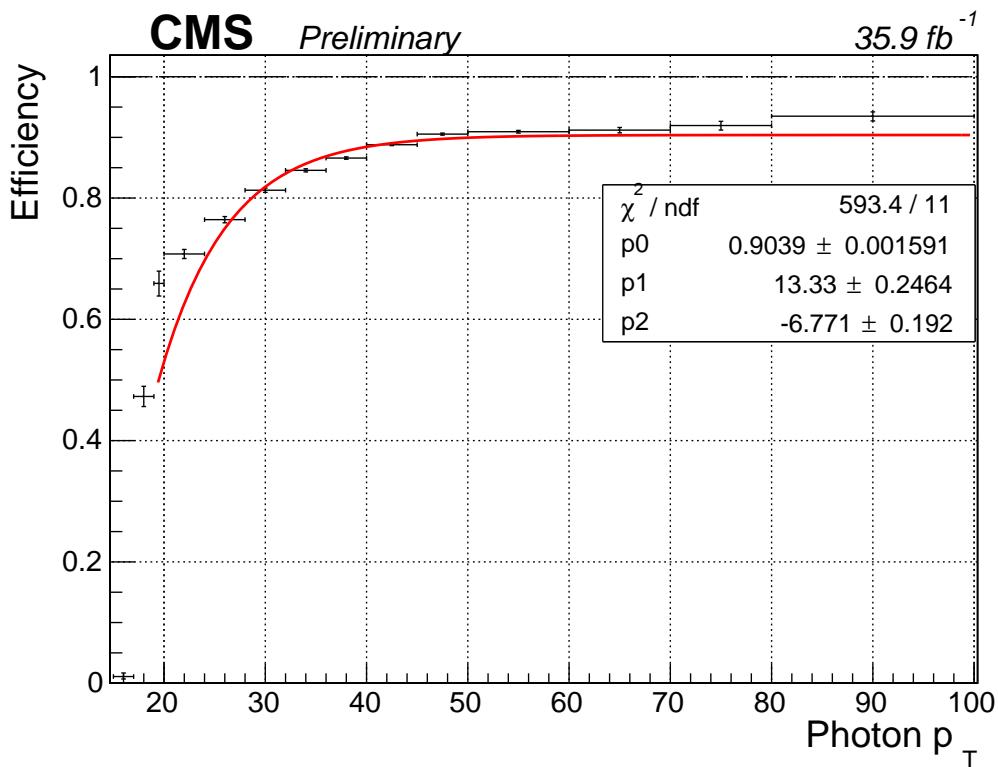


Figure 5.6: Efficiency of the trailing leg of the control trigger with respect to photon  $p_T$ . For  $p_T > 40$  GeV, the trigger is 90.4% efficient. The drop in efficiency with respect to the primary analysis trigger is caused by requiring both electromagnetic objects to be matched to a pixel seed.

## CHAPTER 6

### CMS RECONSTRUCTION

#### 6.1 Overview of the CMS reconstruction process

For each event that passes the trigger, the output from all of the sub-detectors described in Chapter 4 is saved by the data acquisition system (DAQ) and eventually recorded to disk and tape. The data format at this stage is referred to as “RAW” data. It includes information about the response of each detector, but the data are still unprocessed, aside from the minimal processing done to determine whether the event passed the trigger. “Reconstruction” is the general term for algorithms that convert the detector response data into lists of object candidates—muons, photons, electrons, jets, etc.—and event quantities such as missing transverse momentum  $\vec{p}_T^{\text{miss}}$ .

Figure 6.1 shows how each type of particle interacts with the different layers of the CMS detector. Photons, being neutral, do not interact with the tracker and instead deposit their energy in the ECAL. Electrons have a very similar response in the ECAL, but unlike photons, they also leave hits in the tracker. Hadrons deposit their energy in the HCAL, and charged hadrons have an associated track in the tracker. Finally, muons interact with the tracker and the muon system.

Of course, the simplified picture in Figure 6.1 does not cover the full complexity of how particles can interact with the detector. Photons can “convert” and produce an electron-positron pair. Hadrons can interact with the material of the tracker or ECAL and begin to shower before reaching the HCAL. The following sections will describe in more detail how physics objects are reconstructed.

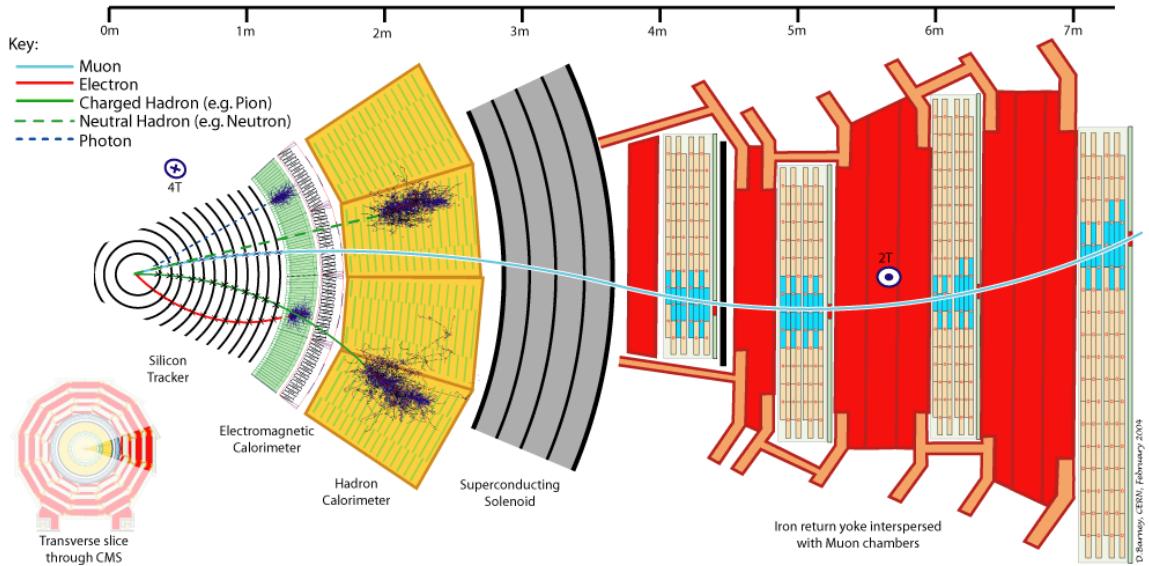


Figure 6.1: Slice of the transverse plane of the CMS detector showing how different particles interact with the detector subsystems. The most important particle for this analysis is the photon (blue dotted line), which interacts only with the ECAL (shown in green). Reprinted from Reference [14].

## 6.2 Particle flow algorithm

CMS uses a “particle flow” (PF) reconstruction algorithm [27]. The PF approach seeks to assign each hit in the tracker and each calorimeter energy deposit (known as a cluster) to a final-state particle. This holistic view maximizes our ability to identify particles by using the full set of information provided by the detector. CMS is the first hadron accelerator experiment to use PF for reconstruction. It is made possible through the excellent spatial resolution of the tracker and calorimeters and the large bending power of the magnet. Both of these features are necessary so that signals from nearby particles do not overlap. For example, the ECAL cluster from a charged particle with  $p_T = 20$  GeV will be deviated by 5 cm from the ECAL cluster of a neutral particle emitted in the same direction, allowing the two particles to be reconstructed independently [27].

The ingredients to the PF algorithm are reconstructed tracks and calorimeter

clusters. The relevant algorithms and methods for each of these elements will be discussed first, before describing how they are assembled to form the final lists of candidate photons, electrons, muons, and jets.

### 6.2.1 Track reconstruction

Tracking in CMS takes place with a combinatorial track finder that is applied in ten iterations. The iterations use successively more complex algorithms with looser requirements on the track properties. Each iteration includes three basic steps:

- Seed generation: The first step is to find a track seed consisting of two or three hits compatible with a charged-particle track.
- Track finding: Pattern recognition is used to identify hits in all layers of the tracker that are compatible with the trajectory implied by the track seed.
- Track fitting: A global  $\chi^2$  fit is performed to determine the final properties of the track, including origin, direction, and transverse momentum.

Because the charged particles can change direction suddenly through interactions with the tracker material, the track finder uses a Kalman filter method [48]. The goal of each iteration is to find tracks with as high of a purity as possible, at the cost of moderate efficiency. The high purity is accomplished through quality cuts on the seeds and the  $\chi^2$  of the track fit. Any hits that are associated with a track are masked for the subsequent iterations. This reduces the combinatorial possibilities, which in turn reduces the likelihood of random hits being built up into a fake tracks.

In addition to particle identification and determination of transverse momentum, tracks are also used to reconstruct vertices in the event. The vertex with the largest value of  $\Sigma p_T^2$  is identified as the primary vertex, associated with the hard-scatter interaction. The other vertices arise from soft-scatter pileup interactions. Thus, the number of reconstructed vertices in an event can be used to approximate the number

of pileup interactions. Two of the later iterations of the track finder are applied with reduced constraints on the origin vertex, in order to reconstruct displaced secondary vertices from the decays of long-lived particles.

### 6.2.2 Electron track reconstruction

Electrons have a higher probability of scattering in the tracker than other charged particles. Therefore, electrons are reconstructed using a separate tracking algorithm that is optimized for multiple scatters. In particular, a Gaussian-sum filter (GSF) is used rather than a Kalman filter because it adapts better in the case of sudden and substantial momentum changes. Two approaches are used: an ECAL-driven approach and a tracker-driven approach. In the ECAL-driven method, ECAL clusters are used to infer the hit locations assuming that the cluster came from either a positron or an electron.

The tracker-driven method is better for electrons in jets and for low- $p_T$  electrons, when the ECAL clusters are overlapping with deposits from other particles or are too spread out to reconstruct fully. In this approach, all tracks with  $p_T$  greater than 2 GeV are used as seeds for an electron. Each track is propagated to the surface of the ECAL and checked to see if it is consistent with the position of an ECAL cluster. The final step of the algorithm uses a boosted decision tree (BDT) to quantify the goodness of fit. This last step serves to reduce the misidentification rate, which is especially important in the context of the PF algorithm to avoid double counting energy.

### 6.2.3 Muon track reconstruction

PF muons are classified into three categories:

- All reconstructed tracks with  $p_T > 0.5$  GeV and  $|\vec{p}| > 2.5$  GeV in the inner tracker are propagated to the muon system, and classified as **tracker muons** if they match

at least one hit in the muon section.

- **Standalone muons** are reconstructed using the muon system only. The process starts with seed tracks in the DT or CSC detectors, and a pattern recognition algorithm identifies DT, CSC, and RPC hits corresponding to the muon trajectory.
- **Global muons** are standalone muons that are consistent with a track in the inner tracker.

Because of the extensive amount of material that muons must traverse before reaching the muon system, the inner tracker gives the best energy resolution for muons up to 200 GeV [27].

#### 6.2.4 Reconstruction of calorimeter clusters

Calorimeter clusters are an important input for the PF algorithm. They are the sole means of detection for photons and neutral hadrons, and they are essential for improving the energy resolution of electrons and high- $p_T$  jets. The reconstruction of a cluster begins with a seed cell (ECAL crystal or HCAL scintillating tile) that corresponds to a local energy maxima above a given threshold. Starting from the seed cell, topological clusters are formed by adding cells that share a side or corner with the cluster and have an energy that is at least twice the noise threshold. For the EB, the seed energy must be greater than 230 MeV, and subsequent crystals must have an energy of at least 80 MeV in order to be added to the topological cluster. Finally, a fit is performed to determine what fraction of the energy in the cell is associated with each PF cluster, where the number of PF clusters is equal to the number of seeds incorporated into the topological cluster.

The ECAL clusters are calibrated for photons and electrons using test beam data and radioactive sources, and residual corrections are derived using simulated single photon events. For hadrons, however, the ECAL must be re-calibrated, since the

response of the ECAL to hadrons is significantly different than its response to electrons. A large simulated sample of neutral hadrons is used to derive a simultaneous calibration for the portions of the hadronic shower’s energy deposited in the ECAL and HCAL.

One complication in the reconstruction of ECAL clusters is the presence of signals known as spikes. These occur when a particle directly ionizes an ECAL avalanche photodiode. The result is a signal whose amplitude is  $10^5$  times higher than the signal from the scintillation light alone. Because these spikes appear in only one or two crystals, they are rejected by considering the ratio of energy deposited in the central crystal to that deposited in the neighboring 4 crystals. In the case where the spike affects two adjacent crystals, the energy deposited in the two crystals is compared to that of their six neighbors. These ratios are referred to as  $E_4/E_1$  and  $E_6/E_2$ , and are required to be less than 5% and 10%, respectively.

#### 6.2.5 Particle identification

A “link” algorithm is used to connect tracks and calorimeter clusters and quantify the likelihood that two elements arose from the same particle. A dedicated procedure is used to link tracks with any ECAL clusters that are consistent with photons from electron bremsstrahlung and to link two tracks compatible with a photon conversion. A group of linked elements is referred to as a PF block.

Particle flow is an iterative procedure. As object candidates are identified within a PF block, their corresponding tracks and clusters are removed from further consideration. This makes PF different from more traditional reconstruction algorithms where clusters or hits could be attributed to more than one particle. The object-identification process is repeated until all tracks and clusters have been assigned to a single PF candidate.

First, muon candidates are identified with two sets of identification criteria: one

for isolated muons and one for muons inside or near jets. Next, electron candidates and photon candidates are reconstructed. Finally, any remaining tracks or calorimeter clusters in the PF block are assigned to a neutral hadron, charged hadron, or photon candidates.

### 6.3 Jet reconstruction

Jets are reconstructed from PF candidates using the anti- $k_T$  algorithm with a distance parameter  $R = 0.4$  [13]. The anti- $k_T$  algorithm is a sequential clustering algorithm where the distance  $d_{ij}$  between two particles  $i$  and  $j$  and the distance  $d_{iB}$  between particle  $i$  and the beam  $B$  are given by the following:

$$d_{ij} = \min(k_{ti}^{2p}, k_{ti}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (6.1)$$

$$d_{iB} = k_{ti}^{2p}$$

The value  $\Delta_{ij}^2$  is equal to  $(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ ,  $k_t$  is the transverse momentum of the particle, and  $R$  is a distance parameter that determines the final radius of the jet.

The clustering algorithm first finds the smallest value of  $d_{ij}$  and  $d_{iB}$  for all particles in the event. If the minimum distance is  $d_{ij}$ , then particles  $i$  and  $j$  are combined into a single entity. If the minimum distance is  $d_{iB}$ , then particle  $i$  is labeled a jet and removed from the list. For the anti- $k_T$  algorithm, the parameter  $p = -1$ . Setting  $p = 0$  corresponds to the inclusive Cambridge/Aachen algorithm, and  $p = 1$  is the inclusive  $k_t$  algorithm. The result of various jet-clustering algorithms is shown in Figure 6.2. One obvious feature of the anti- $k_T$  algorithm is the production of very circular jets compared to jets identified using other clustering algorithms.

One benefit of the anti- $k_T$  algorithm is that it is infrared and collinear (IRC) safe. Infrared safe means that anti- $k_T$  jets are insensitive to nearby soft particles. In the anti- $k_T$  algorithm, soft particles (those with  $k_t \rightarrow 0$ ) have a large  $d_{ij}$  and are therefore

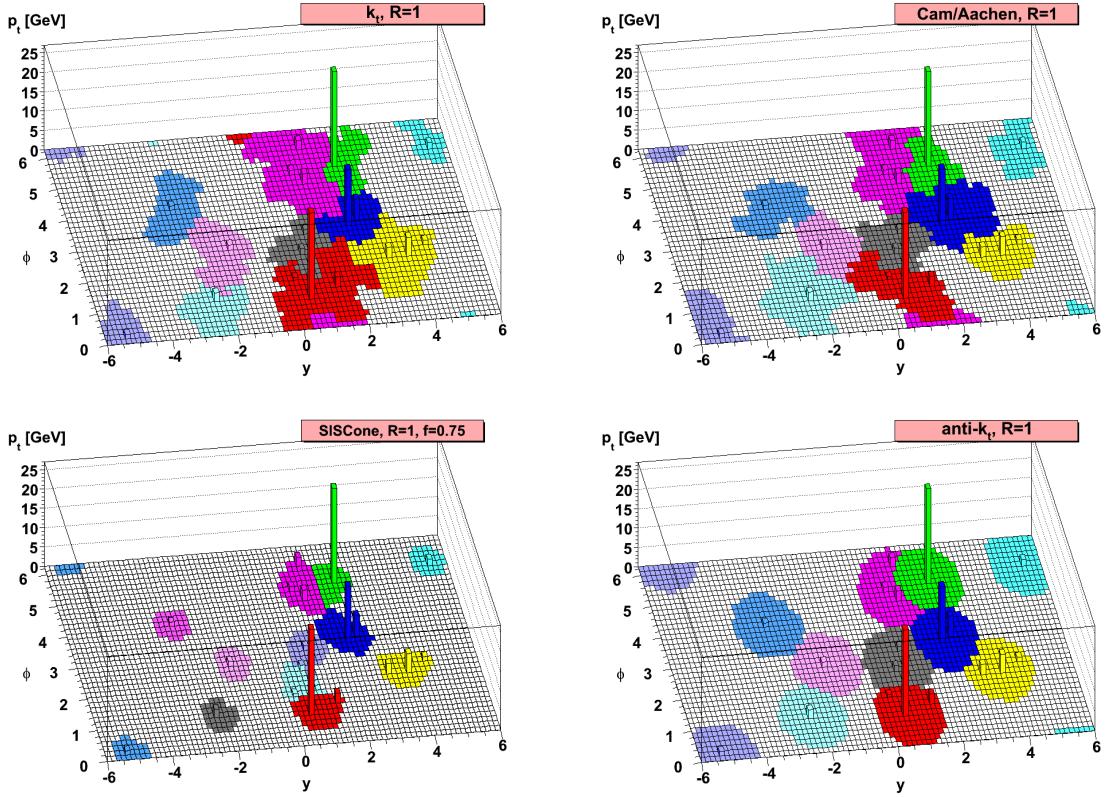


Figure 6.2: Behavior of several jet-clustering algorithms illustrated with a sample parton-level event. CMS uses the anti- $k_T$  algorithm (bottom right) with a distance parameter  $R = 0.4$ . Reprinted from Reference [13].

clustered last. This leaves hard (ie, high  $p_T$ ) jets unaffected. Soft particles could be clustered into many soft jets, but those are simple to remove during the analysis. Being infrared safe is particularly important for the high luminosity environment of the LHC, so that the clustering of the hard jets from the primary interaction is not changed by the many soft particles from pileup interactions.

Collinear safe means that very energetic initial quarks will still get reconstructed as a single jet. In the anti- $k_T$  algorithm, collinear particles will have a small value of  $\Delta_{ij}^2$  and will be clustered together first. Using an IRC safe algorithm is important for comparing theory and experiment, because if jets are not IRC safe, then their cross sections cannot be calculated using perturbation theory.

### 6.3.1 Jet energy corrections

A series of jet energy corrections (JEC) are applied to the jets after clustering to improve the calibration and energy resolution [23]. The  $L1$  correction is a flat correction designed to remove contributions to the jet energy from pileup (see Section 6.5 for a more thorough description of pileup mitigation techniques). The  $L2$  corrections are  $\eta$ -dependent and the  $L3$  corrections are  $p_T$ -dependent. These are calculated in simulation using generator truth information. In data, the correction factors are derived using dijet events and events where a jet is emitted back-to-back with an photon. In the latter case, the superior energy resolution of the ECAL is exploited to provide a handle on the jet energy. The effects of JEC on jet response in QCD simulation is shown in Figure 6.3.

### 6.3.2 Jet energy resolution

After the JEC are applied, resolutions of 10% and 5% are achieved for 100 GeV and 1 TeV jets in the barrel, respectively. Figure 6.4 shows the jet energy resolution as a function of jet  $p_T$  in simulated 8 TeV events.

## 6.4 Missing transverse momentum

After the full set of PF objects have been identified and reconstructed, the missing transverse momentum  $\vec{p}_T^{\text{miss}}$  is given by the following, where the sum is over all PF objects:

$$\vec{p}_T^{\text{miss}} = \sum \vec{p}_T \quad (6.2)$$

The magnitude of this quantity is referred to as the missing transverse energy  $E_T^{\text{miss}}$ . For the calculation of  $E_T^{\text{miss}}$  in this analysis, Type-I corrections are applied, which correspond to propagating the JEC described above to the  $E_T^{\text{miss}}$ .

There are several situations that can arise where particle misidentification or

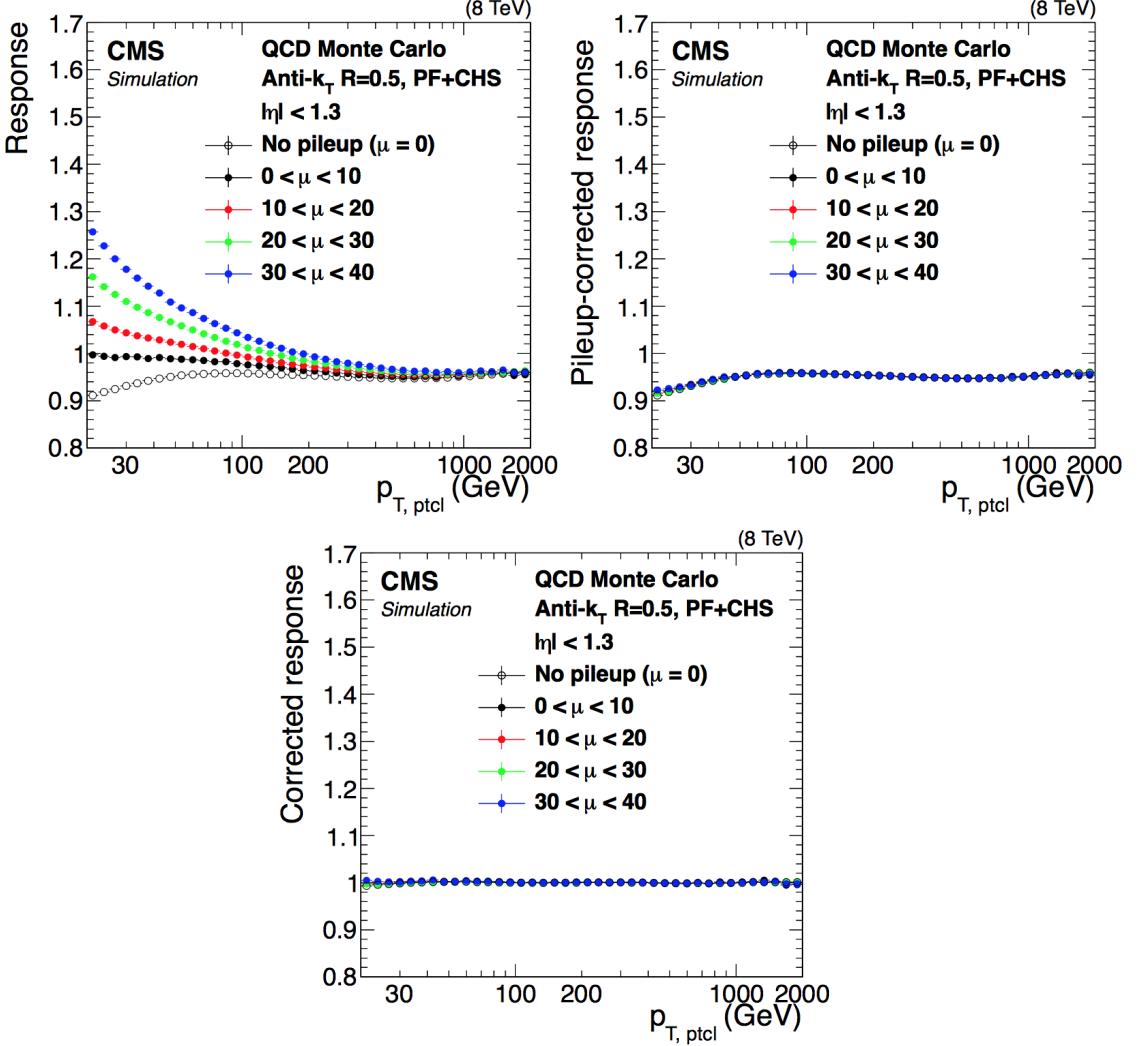


Figure 6.3: Ratio of the average reconstructed jet  $p_T$  to the particle-level jet  $p_T$  ( $p_{T,\text{ptcl}}$ ) in QCD Monte Carlo simulation. The ratio is calculated in bins of  $p_{T,\text{ptcl}}$  before applying any JEC (top left), after applying the  $L1$  pileup corrections (top right), and after all JEC (bottom). The different curves correspond to different numbers of pileup interactions  $\mu$ . Reprinted from Reference [23].

poor reconstruction leads to an artificially large value of  $E_T^{\text{miss}}$ . These are dealt with in post-processing by substituting the reconstructed particles with alternative hypotheses and checking how the  $E_T^{\text{miss}}$  changes. For example, very energetic charged hadrons can occasionally punch through the HCAL and leave tracks in the muon system. In this case, PF will reconstruct a neutral hadron and a muon, leading to

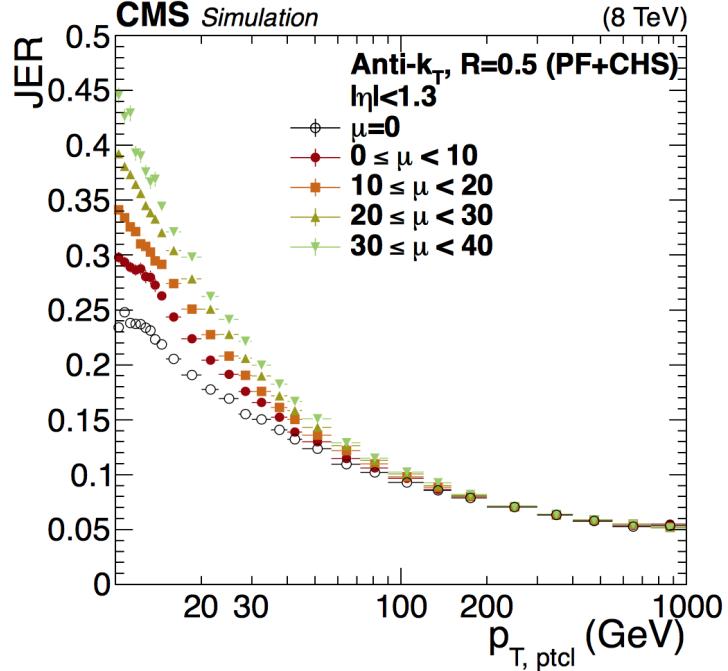


Figure 6.4: Jet energy resolution (JER) in 8 TeV simulation as a function of particle-level jet  $p_T$  ( $p_{T,\text{ptcl}}$ ). PF+CHS refers to particle flow jets with charged hadron subtraction (see Section 6.5). Results are shown for various levels of pileup  $\mu$ . Reprinted from Reference [23].

a large value of  $\vec{p_T}^{\text{miss}}$  pointing in the opposite direction. If replacing the neutral hadron and muon with a charged hadron reduces the  $E_T^{\text{miss}}$  by 50% or more, then that alternative reconstruction is used instead.

In addition to the post-processing corrections just described, a series of  $E_T^{\text{miss}}$  filters are applied. The filters are designed to tag events where the  $E_T^{\text{miss}}$  is poorly reconstructed. At the analysis stage, if an event fails any of the  $E_T^{\text{miss}}$  filters it is removed from consideration. The following  $E_T^{\text{miss}}$  filters applied in this analysis:

- Primary vertex filter
- CSC beam halo filter
- HBHE noise filter
- HBHEiso noise filter

- ECAL trigger primitive filter
- ee badSC noise filter
- Bad PF muon filter
- Bad charged hadron filter
- Duplicate muons filter

## 6.5 Pileup subtraction

The presence of pileup interactions affects many aspects of the reconstruction process, in particular the jet energy resolution, lepton isolation values, and  $E_T^{\text{miss}}$ . Figure 6.5 shows the observed pileup distribution for the 2016 CMS  $p$ - $p$  collisions.

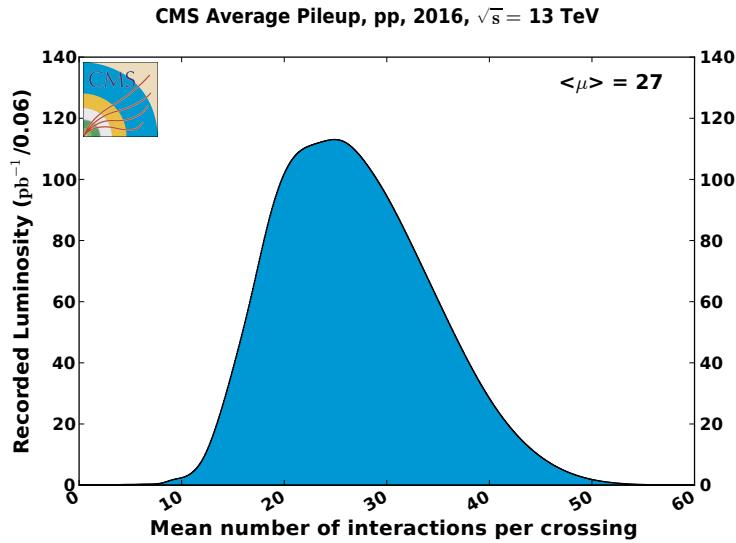


Figure 6.5: Number of pileup interactions per bunch crossing as observed in  $p$ - $p$  collisions with the CMS detector in 2016. On average, 27 pileup interactions were observed per bunch crossing.

Fortunately, there are several ways in which the PF algorithm can be used to mitigate the effects from pileup. Charged-hadron subtraction (CHS), for example, is the procedure of removing charged hadrons from consideration if they are associated with a pileup vertex rather than the hard-scatter vertex. Figure 6.6 shows the composition of jets in data and simulation as a function of the number of pileup interactions  $\mu$ . For an event with  $\mu = 30$ , 10% of the uncorrected jet energy is due to charged hadrons from pileup interactions.

The pileup contributions from photons and neutral hadrons, however, cannot be subtracted on an object-by-object basis, since there is no good way to identify whether these neutral particles were produced in the primary interaction or in a pileup interaction. These effects can only be subtracted from the event on average. For lepton and photon isolation, this is done by calculating  $\rho$ , the average  $p_T$  density due to pileup interactions. Figure 6.7 shows the average value of  $\rho$  for events passing the primary analysis trigger. To calculate  $\rho$ , the  $p_T$  of all PF candidates in a fixed grid in  $\eta$  and  $\phi$  is summed. The value  $\rho$  is defined as the median energy of the grid divided by the area in  $\eta \times \phi$ .

The average contribution to the isolation sum from pileup is then given by  $\rho$  times the effective area  $A_{\text{eff}}$  of the object under consideration. The calculation of  $A_{\text{eff}}$  for photons and electrons is done in a data-driven way by plotting the isolation versus  $\rho$  and fitting the distribution:

$$I_{\text{ave}} = I_0 + A_{\text{eff}}\rho \quad (6.3)$$

where  $I_0$  is an offset value and  $I_{\text{ave}}$  is the average isolation for a given value of  $\rho$ . The values for  $A_{\text{eff}}$  are calculated in bins of  $\eta$ .

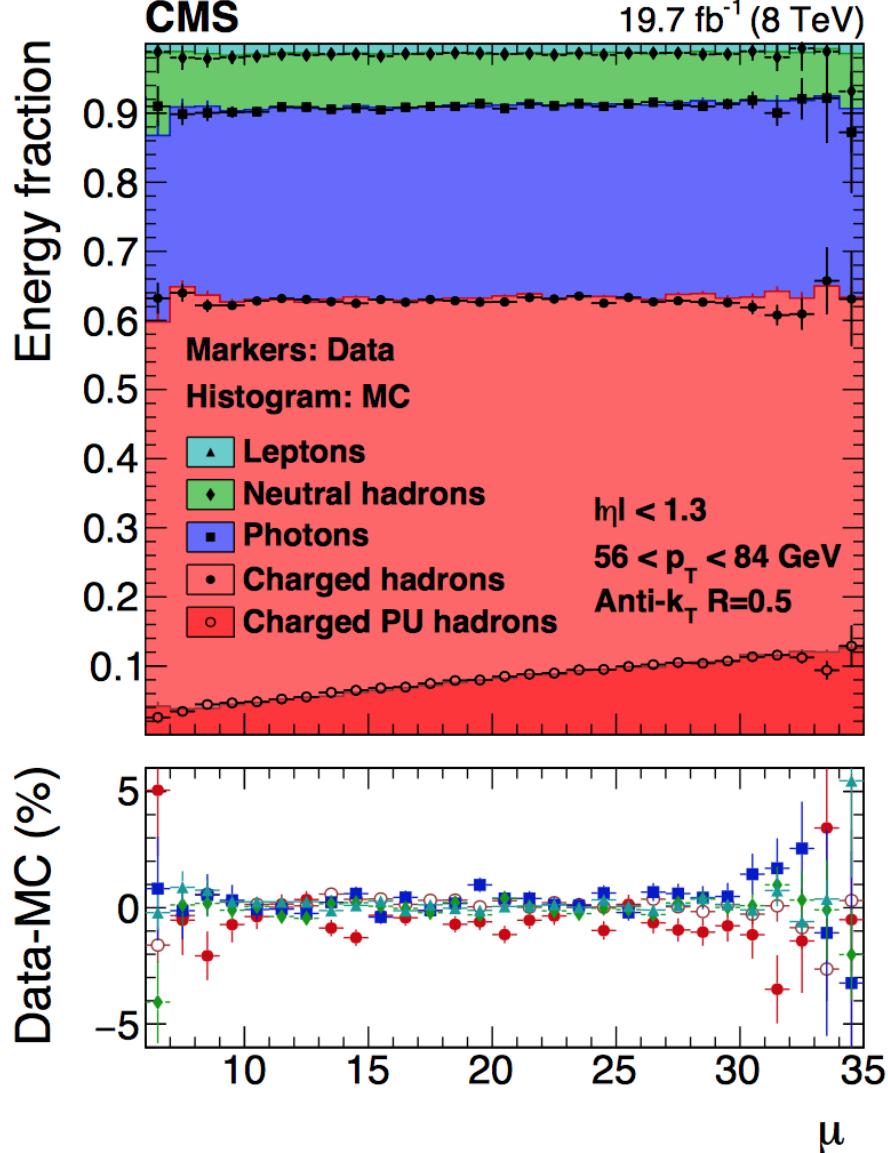


Figure 6.6: Energy composition as a function of the number of pileup (PU) interactions  $\mu$  for jets in data (markers) and simulation (histogram). The data distributions correspond to  $19.7 \text{ fb}^{-1}$  collected at  $\sqrt{s} = 8 \text{ TeV}$  with the CMS detector in 2012. The contribution from charged hadrons from pileup is shown in dark red. The bottom pane shows the agreement between data and simulation for each category. Reprinted from Reference [27].

## 6.6 Photon reconstruction

For defining jets and calculating  $E_{\text{T}}^{\text{miss}}$ , the PF reconstruction of photons and electrons is used. For photons and electrons used to define signal regions or con-

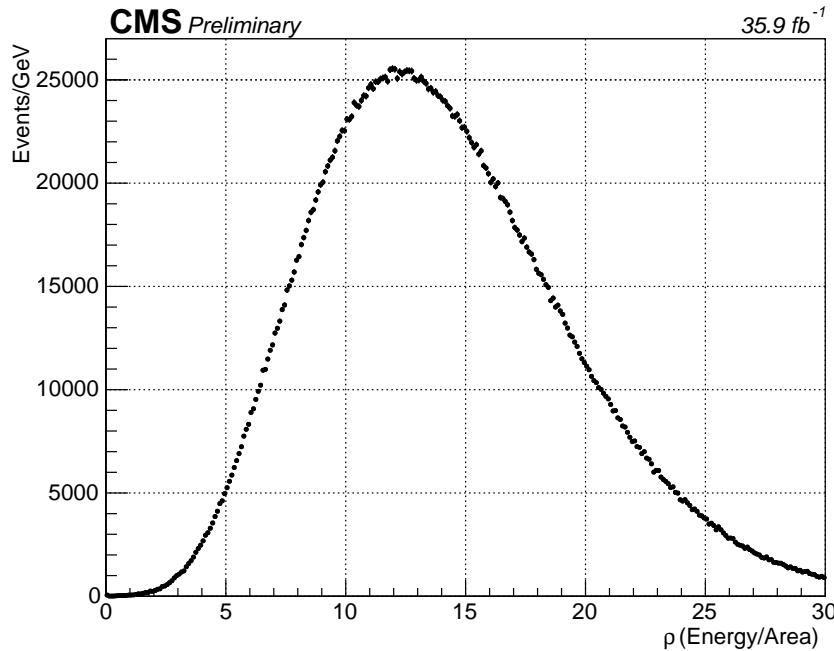


Figure 6.7: Average transverse momentum density  $\rho$  for the 2016  $p$ - $p$  data set.

trol samples, however, a more optimized reconstruction that achieves a better energy resolution is employed [35, 32]. First, an intercalibration is performed between the ECAL crystals to ensure their responses are uniform. The first-order intercalibration is applied using the laser monitoring system described in Section 4.4. Further corrections are derived by looking at the invariant mass distribution in  $\pi_0 \rightarrow \gamma\gamma$  and  $\eta \rightarrow \gamma\gamma$  decays, and by noting that the energy deposited by pileup and the underlying event should be symmetric in  $\phi$ .

In order to achieve an excellent energy resolution, the clustering algorithm must collect the deposits corresponding to all electrons from photon conversions and all photons from electron bremsstrahlung. Similar to the PF cluster reconstruction described above, the reconstruction of photon clusters begins with a local energy maxima or “seed” crystal. In the barrel, clusters have a fixed width of 5 crystals in  $\eta$  centered on the seed. The initial cluster is made from parallel  $5 \times 1$  strips in  $\eta \times \phi$ , and further strips in the  $\phi$  direction (up to a total of 17 strips) are added if the

energy is above a set threshold and lower than the previously added strip. In the endcap, the default cluster size is a  $5 \times 5$  region, and adjacent  $5 \times 5$  regions are added to the supercluster if they are sufficiently energetic. Any overlapping signal in the preshower detectors is simply added to the supercluster energy.

The variable  $R_9$  is defined as the ratio of the energy in a  $3 \times 3$  array of crystals to the energy in the full supercluster. It is used to distinguish between converted and unconverted photons. For photons that have converted, the value of  $R_9$  is lower because the  $e^+e^-$  pair will separate in the  $\phi$  direction due to the strong magnetic field. If  $R_9$  is less than 0.93, then the photon is classified as converted, and the energy of the supercluster is used. If the photon is unconverted (ie,  $R_9 > 0.93$ ), then the energy of the  $5 \times 5$  region is assigned as the photon supercluster energy.

A series of corrections are applied to the supercluster energy to account for variations in the shower containment and any potential losses for photons that convert in the tracker. The shower containment can vary for several reasons, including variations in the longitudinal depth at which the shower passes through the sides of the crystal and is lost in the inter-crystal dead space. The corrections are derived in simulation as a function of  $\eta$ ,  $E_T$ ,  $R_9$ , and the spread of the cluster in  $\phi$ . After the corrections, the photon energy resolution is better than 3% for photons in the barrel [35].

It is worthwhile to note that the photon reconstruction algorithm does not differentiate between showers initiated by an electron and those initiated by a photon. This is important, because it means that  $Z \rightarrow ee$  events can be used as a high-purity sample of  $e/\gamma$  objects to study essential analysis inputs. These include the trigger efficiency (as previously described in Chapter 5) and the photon ID efficiency in data and simulation (see Section 7.5). Chapter 7 will discuss the offline criteria used to further refine the photon selection.

## CHAPTER 7

### DATA SETS AND SAMPLE DEFINITIONS

#### 7.1 Data samples

As already mentioned, this analysis was performed using  $35.9 \text{ fb}^{-1}$  collected with the CMS detector in 2016. More specifically, the data sets listed in Table 7.1 were used. These correspond to the reprocessing of the data that took place in February 2017 to include improved calibrations and performance corrections. The relevant “golden JSON” file was used to specify which events passed all of the offline CMS data quality monitoring requirements:

`Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`

#### 7.2 MC samples

Monte Carlo (MC) simulations are used for several purposes in this analysis. Simulations of the signal processes are used to determine signal efficiencies, and background process simulations are used for validation of the analysis performance and to model the contribution from  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  events.

There are three steps to create simulated events. The first step is the event generation, which simulates the primary scattering given an initial state ( $p$ - $p$  collisions in our case) and a list of final state particles. To do so, a Monte Carlo (MC) method is utilized. The momentum phase-space of the final particles is randomly sampled using a probability distribution proportional to the differential cross section. For our analysis samples, the MADGRAPH5 [4] matrix-element generator was used.

TABLE 7.1  
DATA SAMPLES

/DoubleEG/Run2016B-03Feb2017_ver2-v2/MINIAOD
/DoubleEG/Run2016C-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016D-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016E-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016F-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016G-03Feb2017-v1/MINIAOD
/DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD
/DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD

The second step is to take the list of particles and corresponding momentum 4-vectors from the output of the event generator and perform parton showering. The partons are fragmented until the energy of each is below a set threshold. The new set of partons from the fragmentation are then hadronized. Additionally, the generation of initial state radiation (ISR) occurs during this stage. We used the general-purpose event generator PYTHIA 8 [62] for this step.

Finally, the last step is to simulate the detector response and to convert the list of particles into detector observables. There are two approaches that were used for the detector simulation: full simulation (FullSim) and fast simulation (FastSim). Both rely on the GEANT4 program. In FullSim, the interactions are calculated from first principles using the full geometry of the detector modeled with as much detail as possible. In FastSim, the detector response is parameterized to avoid the time-consuming steps of FullSim. The modeling of the particle interactions with the tracker

and calorimeters is replaced with libraries of particle showers [1].

The simulated tracker hits and calorimeter deposits are passed through the full reconstruction algorithm that was previously described in Chapter 6 with one exception: in FastSim, the time-consuming track reconstruction uses the MC truth information rather than the simulated detector hits. The final simulated output ends up in the same format as the output of the data reconstruction.

The full list of MC samples for this analysis is shown in Table 7.2. The background MC samples are generated with FullSim, and the signal samples are generated with FastSim. The signal simulation will be described in more detail in Section 9.2.

### 7.3 Object definitions

The MINIAOD data format of the data sets in Table 7.1 includes lists of object candidates—photons, electrons, jets, etc.—from the reconstruction algorithms described in Chapter 6. Further refinements, however, are needed in the offline analysis to achieve the desired purities. For this analysis, the primary objects of interest are photons and electrons.

#### 7.3.1 Photon identification

Photons are required to satisfy  $p_T > 40$  GeV in order to be in the region where the trigger is fully efficient. Due to the kinematics of the SUSY models under consideration, the majority of photons in SUSY events will be produced in the central region of the detector. For this reason, and to avoid the added complexity of the high-occupancy endcap, photons must be in the fiducial region of the ECAL barrel,  $|\eta| < 1.4442$ .

We use the medium working point of the cut-based photon ID derived by the CMS  $e/\gamma$  Physics Object Group (EGM POG). The medium working point was tuned to

have an efficiency of 80% and includes the following cuts:

- H/E : The ratio of the energy deposited in the HCAL tower directly behind the ECAL supercluster to the supercluster energy is required to be less than 0.0396.
- $\sigma_{i\eta i\eta}$ : The log-fractional weighted width of the shower in  $i\eta$ -space is required to be less than 0.01022.
- Particle Flow Charged Isolation : The  $p_T$  sum of all PF charged hadrons within a hollow cone of  $0.02 < \Delta R < 0.3$  around the supercluster is required to be less than 0.441 GeV.
- Particle Flow Neutral Isolation : The  $p_T$  sum of all neutral hadrons within a cone of  $\Delta R < 0.3$  around the supercluster is required to be less than  $2.725 + (0.0148 \times p_T^\gamma) + (0.000017 \times (p_T^\gamma)^2)$ .
- Particle Flow Photon Isolation : The  $p_T$  sum of all photons within a cone of  $\Delta R < 0.3$ , excluding a strip in  $\eta$  of 0.015 around the supercluster, is required to be less than  $2.5718 + (0.0047 \times p_T^\gamma)$ .

All of the isolation values are corrected to remove pileup contributions as described in Section 6.5. The corrected value of isolation  $I_{\text{corr.}}$  is given by

$$I_{\text{corr.}} = I_{\text{uncorr.}} - \rho A_{\text{eff}} \quad (7.1)$$

The values of the effective area for each isolation correction are listed in Table 7.3.

In addition to the medium ID criteria, photons must satisfy  $R_9 > 0.5$ . This is imposed to ensure that photons pass the trigger (which includes an  $R_9$  requirement) with a high efficiency. Finally, to distinguish photon candidates from electrons, photons are required to pass a pixel seed veto (PSV). That is, photons must not be

matched to a pixel seed, defined as at least two hits in the pixel detectors consistent with a charged particle trajectory which would arrive at the ECAL.

### 7.3.2 Electron identification

As already mentioned, electrons—particularly electrons from  $Z \rightarrow ee$  decays—are a useful proxy for photons because of the similar response in the ECAL. We define an electron as a candidate object satisfying all of the requirements of Section 7.3.1 but failing the PSV. Using the same ID requirements for electrons and photons enables us to define control regions with electrons without introducing any biases between the control region and the candidate diphoton sample we are trying to model. By reversing the pixel seed veto, we make the electron and photon selections orthogonal and avoid double-counting objects. The requirements for photons and electrons are listed in detail in Table 7.4.

### 7.3.3 Fake identification

In addition to photons and electrons, a third, orthogonal category referred to as “fake” photons is defined. Fakes are primarily electromagnetically-rich jets that have been misidentified as photons. As will be described in Chapter 8, the predicted QCD background is derived using a double fake control sample.

Table 7.5 lists the ID criteria for fakes. The fake definition is taken from a photon ID sideband. Fakes are required to fail either the  $\sigma_{i\eta i\eta}$  or the charged hadron isolation requirement of the photon ID. This ensures that the photon and fake selection criteria are orthogonal. Upper bounds (0.015 and 25 for  $\sigma_{i\eta i\eta}$  and charged hadron isolation, respectively) are placed on the values of both variables to ensure that the fake definition does not differ too much from the nominal photon definition.

In addition, fakes must satisfy  $0.5 < R_9 < 0.9$ . Using the more inclusive cut  $R_9 > 0.5$  resulted in significant contamination from SUSY events in the double

fake control sample. To counteract the effect this cut has on the already-limited fake statistics, the neutral hadron isolation and photon isolation cuts are loosened significantly.

#### 7.3.4 Object cleaning

In order to avoid double counting particles, a set of object cleaning rules are applied. First, because muons are reconstructed with a higher purity than any other particle, any electromagnetic object (photon, electron, or fake) that is within  $\Delta R < 0.3$  of a muon candidate is removed. Second, any photon that overlaps within  $\Delta R < 0.3$  of an electron is removed. Finally, if a fake overlaps with an electron or photon candidate within  $\Delta R < 0.4$ , the fake candidate is removed. The larger  $\Delta R$  separation for fakes is due to the fact that fakes are primarily jets and jets are reconstructed using the anti- $k_T$  algorithm with a distance parameter of 0.4, as described in Section 6.3.

#### 7.3.5 Lepton veto

In addition to the cuts described above, any event that contains additional muons or electrons is vetoed. No additional leptons are present in the SUSY signals of interest, so applying a lepton veto will not hurt our signal sensitivity. More importantly, by vetoing on the presence of additional leptons, our analysis becomes orthogonal to other CMS searches for gauge-mediated supersymmetry breaking with photons in the final state.

### 7.4 Signal region and control samples

After the individual objects are defined and identified, each event is sorted into one of four exclusive categories based on the electromagnetic objects with the highest  $p_T$

in the event. Events with two photons comprise the signal diphoton sample, referred to as the  $\gamma\gamma$  sample. Events with two electrons, one electron and one photon, or two fakes are categorized as  $ee$ ,  $e\gamma$ , or  $ff$  events, respectively. In each case, the two electromagnetic objects are required to be separated by  $\Delta R > 0.6$ .

The  $\gamma\gamma$ ,  $ff$ , and  $e\gamma$  samples are required to pass the primary trigger described in Section 5.4. In order to ensure that the events pass the trigger with a high efficiency, the invariant mass of the two electromagnetic objects is required to be greater than 105 GeV. The  $ee$  sample, on the other hand, is collected using the control trigger listed in Table 5.1 and is required to satisfy  $75 < m_{ee} < 105$  GeV. Chapter 8 will explain in detail how each of these samples is used in this analysis.

## 7.5 Photon ID scale factor

The efficiency of the photon ID is measured via the tag-and-probe method in  $Z \rightarrow ee$  events [35]. The procedure is similar to the one described in Chapter 5 for calculating the trigger efficiency. A sample of  $Z \rightarrow ee$  events is collected using a single electron trigger. One electron is used as the tag, and one electron is used as the probe. The efficiency is given by the number of probes that pass the photon ID over the total number of tag and probe pairs. This value is computed in both data and simulation, and the ratio of the two efficiencies is referred to as the scale factor.

For this analysis, the official scale factors calculated by the EGM POG for Moriond 2017 were used. The scale factors from the EGM POG are calculated in bins of photon  $p_T$  and  $\eta$ . These are shown in Figure 7.1 along with the associated uncertainties [30].

Rather than using the full map of scale factors shown in Figure 7.1, we compute a weighted average over all photons passing our selection criteria in each SUSY signal mass point. The average scale factors and uncertainties are shown in Figure 7.2.

The final value used in the analysis was an average over the mass points shown

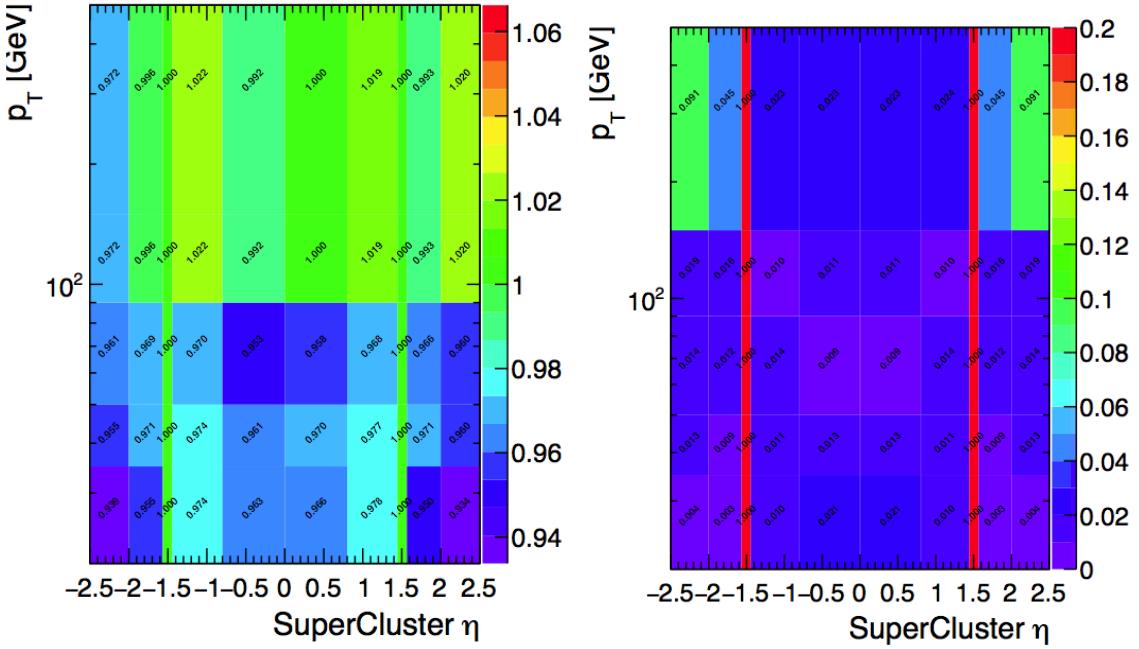


Figure 7.1: Calculated scale factors (left) and uncertainties (right) in bins of photon  $p_T$  and  $\eta$ .

in Figure 7.2:

$$\text{Photon Scale Factor} = 1.002 \pm 0.013 \quad (7.2)$$

### 7.5.1 Scale factor for pixel seed veto

Our prescription for photon identification is nearly identical to that for electron identification, differing only by the presence or absence of a seed track in the pixel detector. The efficiency of the pixel seed veto for photons cannot be determined from the tag-and-probe method described above and must be obtained from photons in  $Z \rightarrow \mu\mu\gamma$  events. Again we use the official scale factor calculated by the EGM POG:

$$\text{Pixel Seed Veto Scale Factor} = 0.998 \pm 0.013 \quad (7.3)$$

Since our candidate sample requires two photons in the final state, two factors of both values are used.

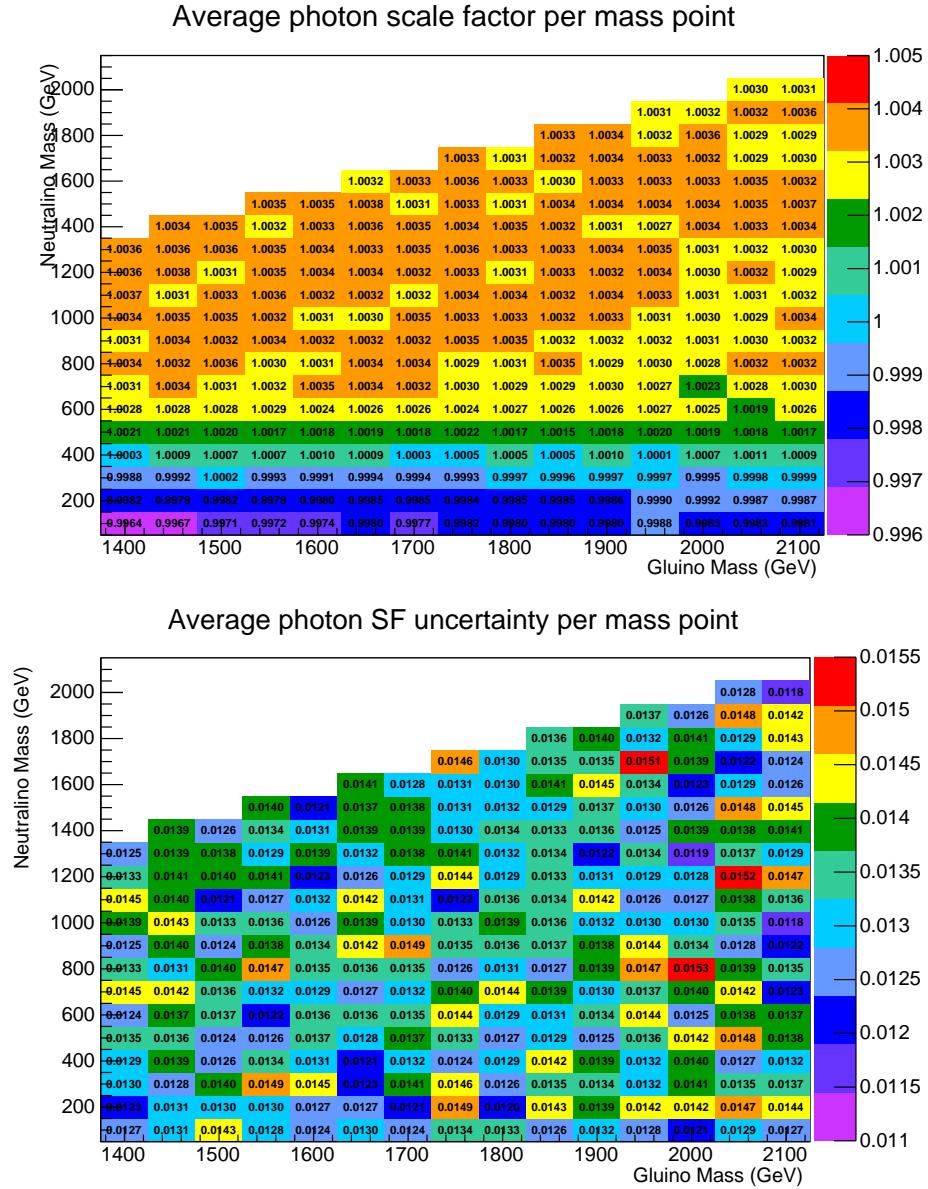


Figure 7.2: Scale factors (top) and uncertainties (bottom) averaged over all photons in each bin in the neutralino versus gluino mass plane.

### 7.5.2 Scale factor for $R_9$ requirement

The photon ID used in this analysis differs from the official POG recipe in one aspect: we apply an  $R_9 > 0.5$  requirement on top of the medium ID due to the presence of an  $R_9$  cut in our analysis trigger. A check was performed to make sure that the  $R_9$  requirement does not change the photon scale factors. The results of this study are shown in Figure 7.3. The data/MC scale factors are all consistent with unity. We therefore choose to simply use the official scale factors described above, rather than applying any additional corrections.

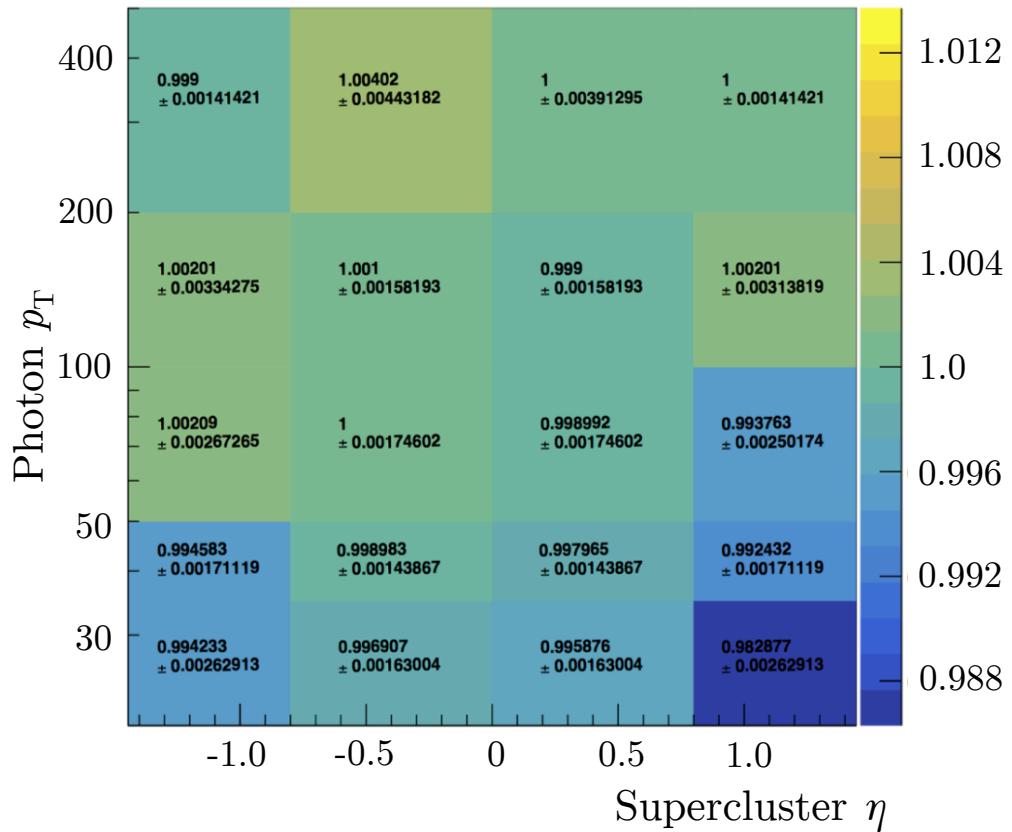


Figure 7.3: Scale factor for applying  $R_9 > 0.5$  on top of the cut-based medium photon ID. The scale factors are consistent with unity in all bins. Therefore, no additional corrections are applied to the scale factors described in Section 7.5.

TABLE 7.2

## MC SAMPLES FOR SIGNAL AND BACKGROUND

SUSY Signal Samples
/SMS-T5Wg_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ RunIISpring16MiniAODv2-PUSpring16Fast_80X_mcRun2_- asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM
/SMS-T5Wg_mGo2150To2500_TuneCUETP8M1_13TeV-madgraphMLM- pythia8/ RunIISpring16MiniAODv2-PUSpring16Fast_80X_mcRun2_- asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM
/SMS-T6Wg_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ RunIISpring16MiniAODv2-PUSpring16Fast_80X_mcRun2_- asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM
/SMS-T6Wg_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ RunIISpring16MiniAODv2-PUSpring16Fast_80X_mcRun2_- asymptotic_2016_miniAODv2_v0-v1/MINIAODSIM
Background MC Samples
/GJet_Pt-40toInf_DoubleEMEnriched_MGG- 80toInf_TuneCUETP8M1_13TeV_Pythia8/ RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_- asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
/ZGGToNuNuGG_5f_TuneCUETP8M1_13TeV-amcatnlo-pythia8/ RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_- asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ RunIISpring16MiniAODv2-PUSpring16_80X_mcRun2_- asymptotic_2016_miniAODv2_v0_ext1-v1/MINIAODSIM

TABLE 7.3  
EFFECTIVE AREAS FOR ISOLATION CORRECTIONS

$ \eta $ Range	Photon Iso	Neutral Hadron Iso	Charged Hadron Iso
$ \eta  < 1.0$	0.120	0.0597	0.0360
$1.0 <  \eta  < 1.479$	0.1107	0.0807	0.0377

TABLE 7.4  
PHOTON AND ELECTRON ID CRITERIA

ID Requirement	Photons	Electrons
Pixel seed veto	Applied	Reversed
$\sigma_{inj\eta}$	$< 0.01022$	
Charged hadron iso.		$< 0.441$
Photon iso.	$< 2.571 + 0.0047 p_T$	
Neutral hadron iso.	$< 2.2725 + 0.0148 p_T + 0.000017 p_T^2$	
$R_9$		$> 0.5$
H/E		$< 0.0396$

Photon and electron identification criteria used to define the signal and control samples for this analysis. The definitions of each of the variables can be found in Section 7.3.1.

TABLE 7.5

## FAKE ID CRITERIA

ID Requirement	Fakes
Pixel seed veto	Applied
$\sigma_{i\eta i\eta}$	$0.01022 < \sigma_{i\eta i\eta} < 0.015$
Charged hadron iso.	$0.441 < \text{Iso} < 25$
Photon iso.	$< 15 + 0.0047 p_T$
Neutral hadron iso.	$< 15 + 0.0148 p_T + 0.000017 p_T^2$
$R_9$	$0.5 < R_9 < 0.9$
H/E	$< 0.0396$

Identification criteria for “fakes.” Fakes come from a photon ID sideband and are primarily jets that have been misidentified as photons. The definitions of each of the variables can be found in Section 7.3.1.

## CHAPTER 8

### DATA ANALYSIS

#### 8.1 Signal region

The signal region for this analysis is made up of events with  $E_T^{\text{miss}} > 100$  GeV that have two photons passing the medium ID requirements listed in Table 7.4. The signal region is divided into six  $E_T^{\text{miss}}$  bins, with bin boundaries at 100, 115, 130, 150, 185, and 250 GeV. The last bin includes all events with  $E_T^{\text{miss}} > 250$  GeV. Figure 8.1 shows an event display of an event in our signal region. The event was recorded on June 12, 2016 and has 144.8 GeV of  $E_T^{\text{miss}}$ .

#### 8.2 Overview of backgrounds

There are several standard model processes that can mimic our signal events. The largest background contribution comes from quantum chromodynamics (QCD) processes. These are primarily multi-jet events, where electromagnetically-rich jets are misidentified as photons, but can also include processes with true photons either from associated photon production or initial-state radiation. In both cases, there is no inherent  $E_T^{\text{miss}}$  in the event. Instead, the apparent  $E_T^{\text{miss}}$  is actually the result of mismeasured hadronic activity. As described in Section 8.3, this background is estimated in an entirely data-driven way using a control region derived from a sideband of our photon ID.

The second-largest background is the electroweak (EWK) background. This background is comprised of  $W\gamma$  or  $W+\text{jet}$  events where  $W \rightarrow e\nu$ . In this case, there is

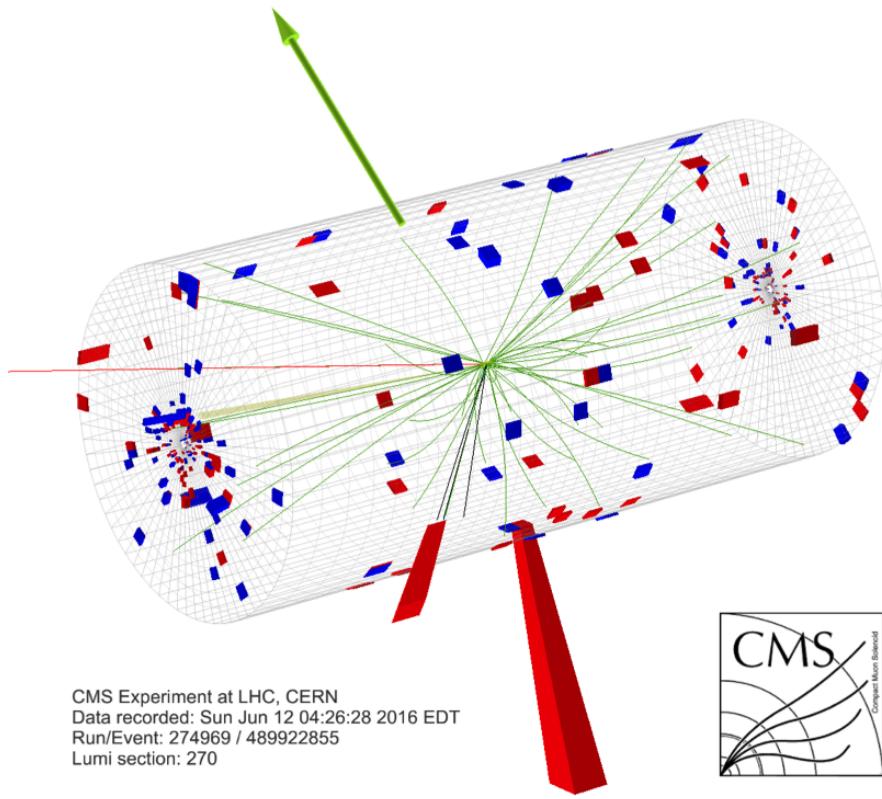


Figure 8.1: Event display of one of the candidate  $\gamma\gamma$  events. The event was recorded by the CMS detector on June 12, 2016. The red and blue blocks represent energy deposits in the ECAL and HCAL, respectively. The size of the blocks are proportional to the amount of deposited energy. The green lines represent reconstructed tracks. The green arrow represents the  $E_T^{\text{miss}}$ . This event has  $E_T^{\text{miss}} = 144.8$  GeV and two identified photons (large red blocks) with  $p_T$  of 151 and 58 GeV.

inherent  $E_T^{\text{miss}}$  from the neutrino, and these events can mimic our signal topography if the electron is misidentified as a photon. By measuring the misidentification rate in data, we can use an  $e\gamma$  control sample to estimate the contribution from the EWK background. The EWK background estimation method is described in detail in Section 8.4.

Finally, there is an irreducible background from  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  events. This background is modeled via simulation and is described in Section 8.5.

### 8.3 QCD background

Due to the large QCD cross section, the most significant background for this analysis comes from QCD events without true  $E_T^{\text{miss}}$  and without two real photons. The observed  $E_T^{\text{miss}}$  is the result of mismeasured hadronic activity, and in most cases the “photons” are misidentified jets with a large electromagnetic component.

To estimate the contribution from the QCD background in our signal region, we use the “fake” object selection that was described in Section 7.3. The fake identification criteria is orthogonal to the nominal photon identification and therefore provides a sideband that can be used as a control region. The  $E_T^{\text{miss}}$  tail of the QCD background is modeled using a “fake-fake” ( $ff$ ) control sample made up of events with two fakes that pass the additional criteria outlined in Section 7.4.

In the 2015 analysis, a double electron ( $ee$ ) control sample was used as the primary estimate, and  $ff$  was used to set an uncertainty on the  $E_T^{\text{miss}}$  shape. The various reasons for not using  $ee$  anymore and the results if we did choose to use  $ee$  are described in Appendix A.

#### 8.3.1 Di-EM $p_T$ reweighting

Because the  $E_T^{\text{miss}}$  in the QCD background stems from poorly measured hadronic activity, it is important that the amount of hadronic activity in the  $ff$  control sample matches that of the  $\gamma\gamma$  events we are trying to model. To account for potential differences between the samples, we define a variable referred to as the “di-EM  $p_T$ ” of an event. Di-EM  $p_T$  is defined as the magnitude of the vector sum of the transverse momentum of the two electromagnetic objects (photons, electrons or fakes):

$$\vec{p}_T^{\text{di-EM}} = \vec{p}_{T1} + \vec{p}_{T2} \quad (8.1)$$

As illustrated in Figure 8.2, the di-EM  $p_T$  variable is used as a measure of the

total hadronic recoil. Because CMS measures the energies of electromagnetic objects with greater precision than the energy of jets, this variable is a more accurate representation of the hadronic recoil than simply adding up the transverse momentum of the jets themselves.

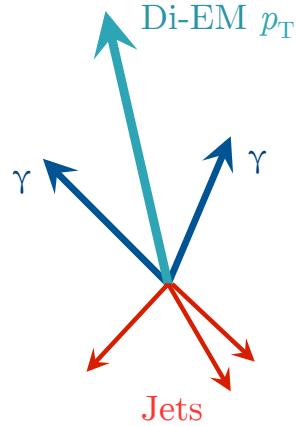


Figure 8.2: The di-EM  $p_T$  vector, shown in light blue, is the vector sum of the  $p_T$  of the two photons in the event, shown in dark blue. The magnitude of the di-EM  $p_T$  vector is used to model the hadronic recoil, shown in red.

The di-EM  $p_T$  distributions for the  $\gamma\gamma$  candidate sample and the  $ff$  control sample are shown in Figure 8.3. The  $ff$  events are reweighted using the  $\gamma\gamma/ff$  ratios displayed in the bottom panel of Figure 8.3. The  $ff$   $E_T^{\text{miss}}$  distribution is normalized to the  $E_T^{\text{miss}} < 50$  GeV region of the  $\gamma\gamma$  sample, where signal contamination is minimal. The unweighted  $ff$  and  $\gamma\gamma$   $E_T^{\text{miss}}$  distributions are shown in Figure 8.4, and Figure 8.5 compares the reweighted  $ff$   $E_T^{\text{miss}}$  distribution to the candidate  $\gamma\gamma$   $E_T^{\text{miss}}$  distribution.

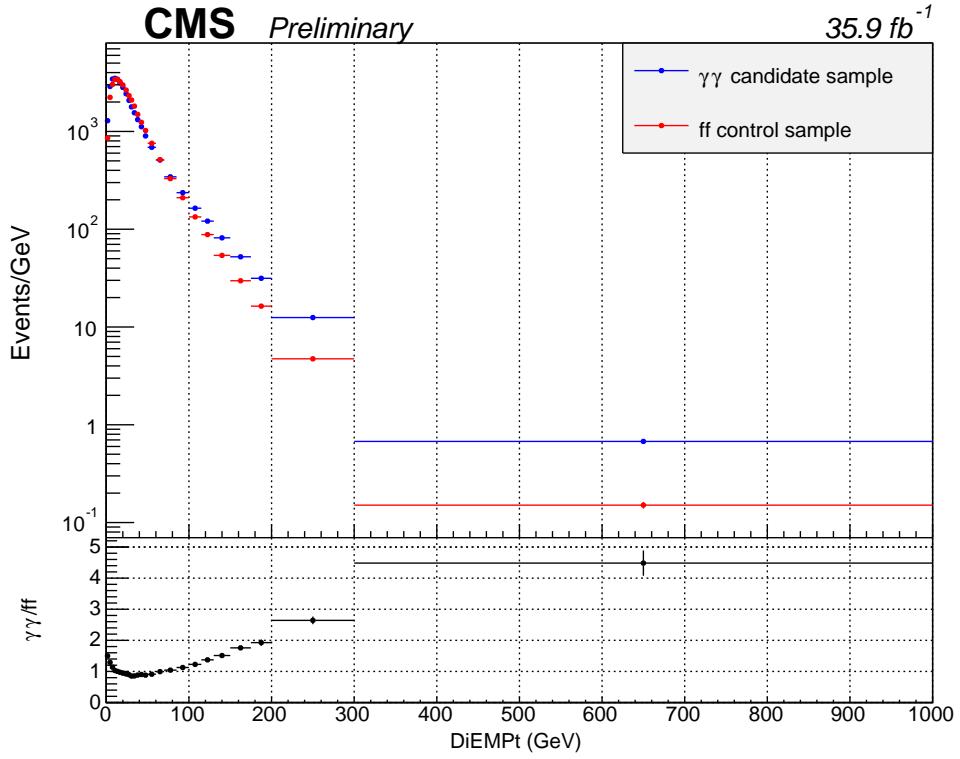


Figure 8.3: Di-EM  $p_T$  distributions of the  $ff$  control sample (red) and the  $\gamma\gamma$  candidate sample (blue). The ratio plot on the bottom shows the  $\gamma\gamma/ff$  ratios that are used to reweight the  $ff$   $E_T^{\text{miss}}$  distribution.

### 8.3.2 Cross check on QCD background

In order to set a systematic uncertainty on the overall  $E_T^{\text{miss}}$  shape predicted using the di-EM  $p_T$  reweighting method, we sought an alternate way to estimate the QCD background. This cross check relies on the assumption that the ratio of  $\gamma\gamma$  events to  $ff$  events should not depend sensitively on  $E_T^{\text{miss}}$ . If this assumption is valid, then we can fit  $\gamma\gamma/ff$  in the low- $E_T^{\text{miss}}$  control region and use it to extrapolate to the high- $E_T^{\text{miss}}$  signal region.

Figure 8.6 shows the ratio of  $\gamma\gamma/ff$  as a function of  $E_T^{\text{miss}}$  in the  $E_T^{\text{miss}} < 100$  GeV control region. The ratio has been fit to an exponential function,  $f(E_T^{\text{miss}}) = p_0 \exp(p_1 E_T^{\text{miss}})$ . The expected number of  $\gamma\gamma$  events in bin  $i$  of the signal region is

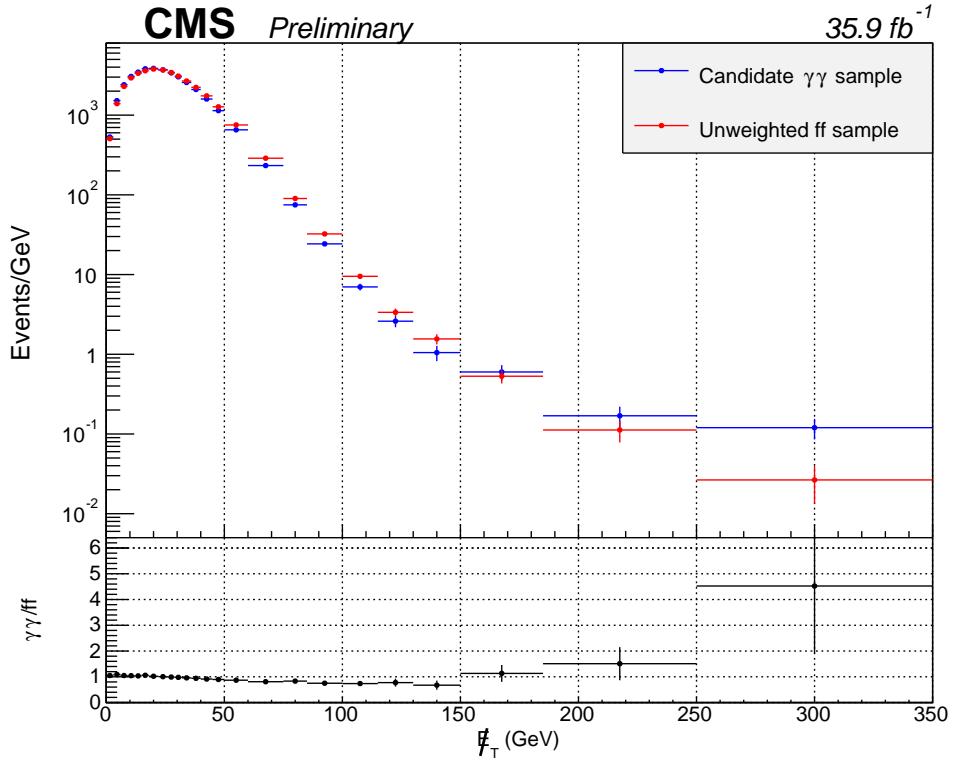


Figure 8.4:  $E_T^{\text{miss}}$  distributions of the unweighted  $ff$  control sample (red) and the  $\gamma\gamma$  candidate sample (blue). The  $ff$  control sample has been normalized to the  $E_T^{\text{miss}} < 50$  GeV region of the  $\gamma\gamma$  candidate sample.

then given by the following equation:

$$N_{\gamma\gamma}^i = f(E_T^{\text{miss}}) \times N_{ff}^i \quad (8.2)$$

Table 8.1 compares the expected QCD background contribution as predicted with the cross check method to that predicted with the di-EM  $p_T$  reweighting method. For the cross check method, the uncertainty arising from the choice of fitting function is estimated by fitting the ratio to a constant function instead. In addition, the cross check uncertainties listed in Table 8.1 include the statistical uncertainty from the limited control sample statistics and the  $1\sigma$  uncertainties from the fit. For the di-EM  $p_T$  reweighting method, the uncertainties include the statistical uncertainty and

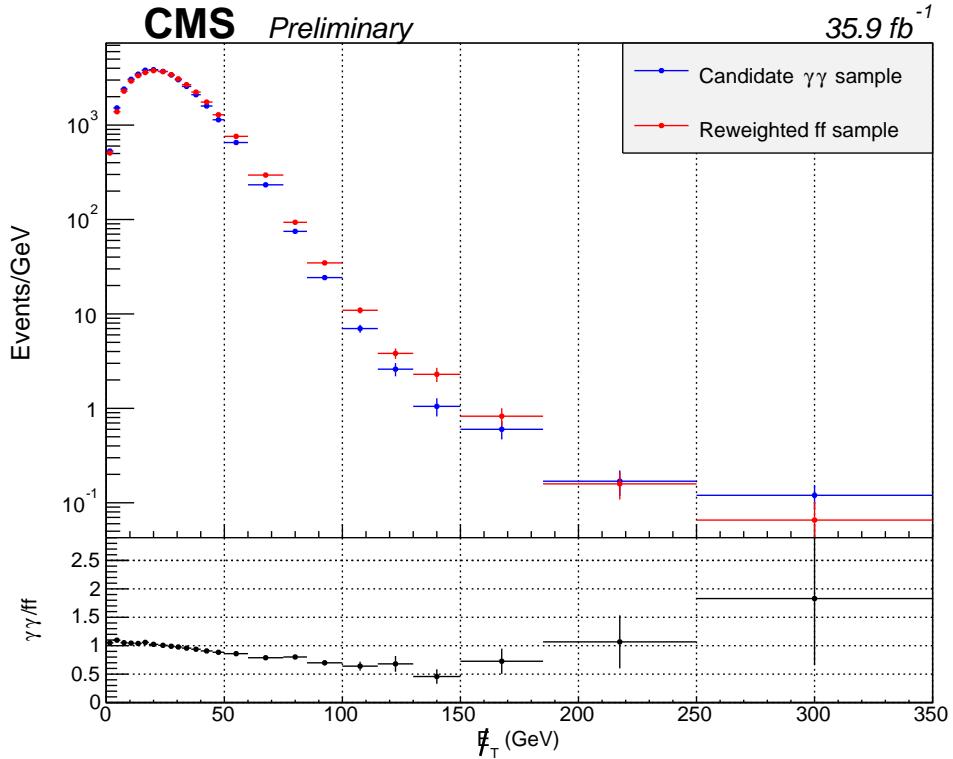


Figure 8.5:  $E_T^{\text{miss}}$  distributions of the di-EM  $p_T$  reweighted  $ff$  control sample (red) and the  $\gamma\gamma$  candidate sample (blue). The  $ff$  control sample has been normalized to the  $E_T^{\text{miss}} < 50$  GeV region of the  $\gamma\gamma$  candidate sample.

the uncertainty from the di-EM  $p_T$  reweighting procedure (the derivation of which is described in detail in Section 8.6.1).

This cross check serves to validate our di-EM  $p_T$  reweighting background estimation method. The difference between the two methods is taken as a systematic uncertainty on the overall  $E_T^{\text{miss}}$  shape.

#### 8.4 Electroweak background

The subdominant background for this search is comprised of  $W\gamma$  and  $W + \text{jet}$  events where  $W \rightarrow e\nu$  and the electron is misidentified as a photon. This background is referred to as the electroweak (EWK) background. Unlike the QCD background,

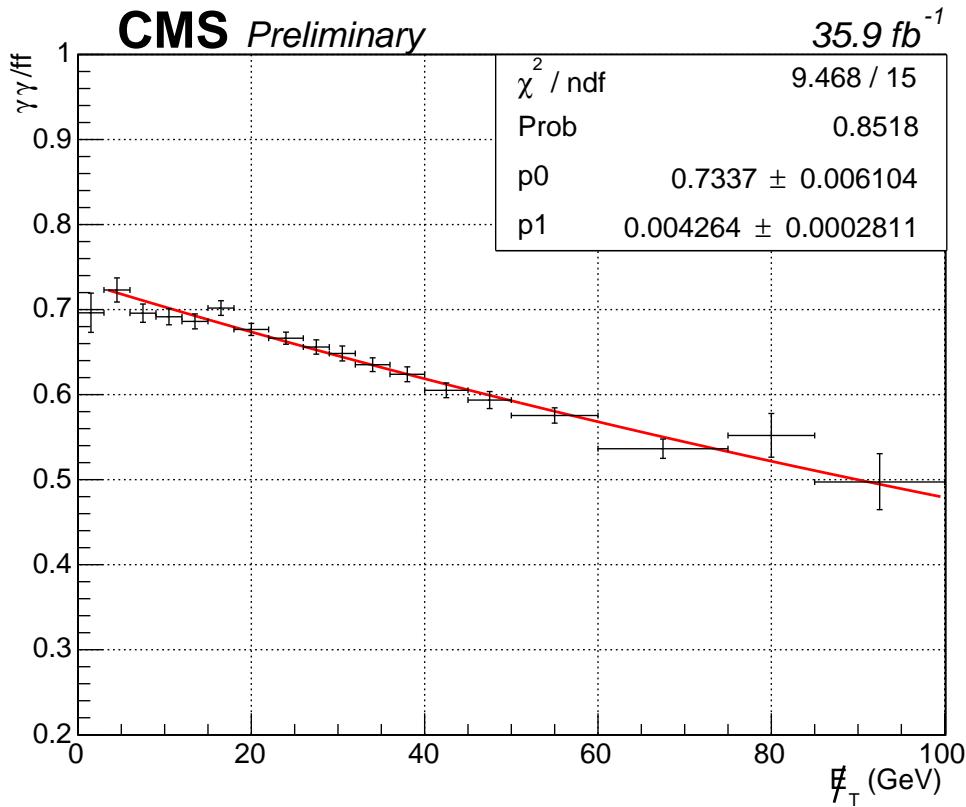


Figure 8.6: Ratio of  $\gamma\gamma$  to  $ff$  events for  $E_T^{\text{miss}} < 100$  GeV. The ratio has been fit to a constant function, which can then be used to extrapolate to the  $E_T^{\text{miss}} > 100$  GeV signal region. This provides an alternative estimation method for the QCD background.

there is inherent  $E_T^{\text{miss}}$  in these events from the escaping neutrino.

To estimate this background, we first calculate the rate at which electrons are misidentified as photons. This is done by comparing the invariant mass peak in a double electron sample ( $ee$ ) with the invariant mass peak in a sample of events with one electron and one photon ( $e\gamma$ ). The composition of the  $e\gamma$  control sample is investigated in Section 8.4.1 and the calculation of the misidentification rate (or “fake rate”) is described in detail in Section 8.4.2. To get the final expected contribution from the EWK background in our signal region, the fake rate is used to calculate a transfer factor (Section 8.4.3). By applying the transfer factor to an  $e\gamma$  control

TABLE 8.1  
 COMPARISON BETWEEN REWEIGHTING METHOD AND RATIO  
 METHOD FOR THE QCD BACKGROUND ESTIMATE

$E_T^{\text{miss}}$ (Gev)	Reweighted ff estimate	Ratio method estimate
100 – 115	$163.8^{+12.0}_{-11.2}$	$99.8^{+40.3}_{-7.8}$
115 – 130	$57.3^{+7.4}_{-6.6}$	$31.7^{+18.2}_{-3.6}$
130 – 150	$45.8^{+7.8}_{-6.7}$	$18.1^{+12.8}_{-2.6}$
150 – 185	$28.9^{+6.6}_{-5.4}$	$9.5^{+8.9}_{-1.8}$
185 – 250	$10.3^{+4.1}_{-3.1}$	$3.0^{+4.3}_{-0.9}$
$\geq 250$	$6.6^{+5.2}_{-3.1}$	$0.8^{+1.9}_{-0.4}$

Comparison of the background estimate using the di-EM  $p_T$  reweighting method and the  $\gamma\gamma/ff$  ratio method. The uncertainties on the di-EM  $p_T$  reweighting method include the statistical uncertainties and the reweighting uncertainty. The uncertainties on the ratio method include the statistical uncertainties, the uncertainties in the fit parameter, and the uncertainty from the choice of fitting function.

sample, we are able to estimate how many of our candidate  $\gamma\gamma$  events are actually events with exactly one photon plus an electron that has been misidentified as a photon (Section 8.4.4).

#### 8.4.1 Composition of $e\gamma$ sample

The  $e\gamma$  control sample is primarily made up of  $\gamma + \text{jets}$  or  $W\gamma$  events. A data versus MC comparison of this control sample is shown in Figure 8.7. The data distribution was fit using the simulated  $\gamma + \text{jet}$  and  $W\gamma$  shapes as templates. From the fit, it was determined that 80% of the  $e\gamma$  events observed in data are from  $\gamma + \text{jets}$  processes, and the remaining 20% are  $W\gamma$  events. In the  $E_T^{\text{miss}} > 100$  GeV signal region, however,  $W\gamma$  events dominate. For  $100 < E_T^{\text{miss}} < 130$  GeV, 12.1% of the total are  $\gamma + \text{jet}$  events. For  $E_T^{\text{miss}} > 130$  GeV,  $\gamma + \text{jet}$  events contribute only 2.6%.

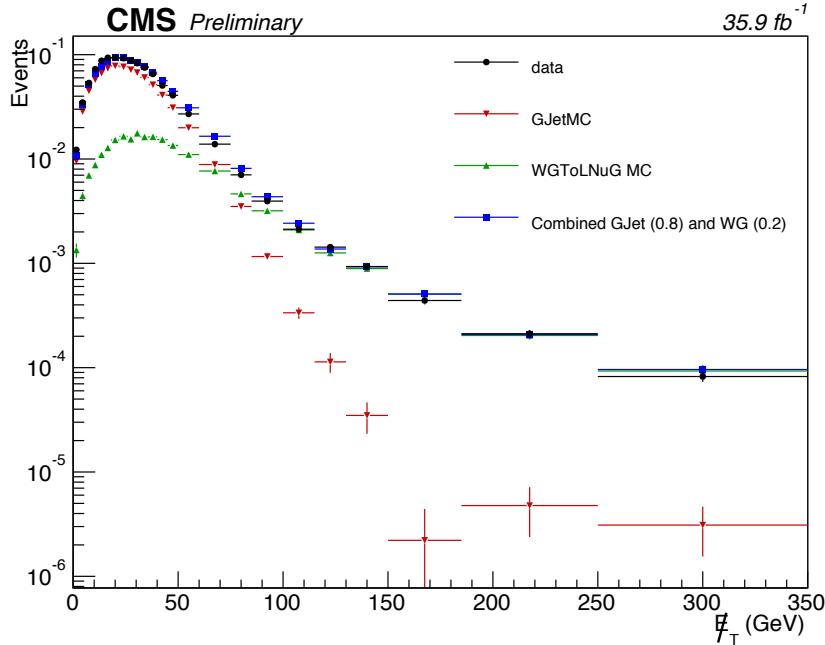


Figure 8.7: Data versus MC comparison of the  $e\gamma$  control sample  $E_T^{\text{miss}}$  distribution. To determine the relative contributions of the  $\gamma + \text{jet}$  and  $W\gamma$  processes, the data distribution was fit using the  $\gamma + \text{jet}$  and  $W\gamma$  shapes as templates. The data are shown in black, and the total MC prediction is shown in blue. The  $\gamma + \text{jet}$  MC (red) was scaled to 80% of the observed events in data, and the  $W\gamma$  MC (green) was scaled to 20% of the data distribution based on the fit results.

#### 8.4.2 Fake rate calculation

A tag-and-probe method is used for the fake rate calculation. Events are selected using a single electron trigger. Electrons passing medium ID criteria and satisfying  $p_T > 30$  GeV and  $|\eta| < 2.1$  are used as tags. The tags are required to be matched to the object firing the single electron trigger within  $\Delta R < 0.2$ . Probes are photon candidates with  $p_T > 40$  GeV that pass all of the identification criteria described in Chapter 7 except for the pixel seed veto. The tag and probe objects are required to be separated by  $\Delta R > 0.3$  and have an invariant mass between 40 and 140 GeV. If there are multiple tag and probe pairs in a single event, all possible combinations are considered. If the probe has a pixel seed, then it is labeled as an electron and falls into the  $ee$  sample. If the probe does not have a pixel seed, however, then it is

labeled as a photon and is included in the  $e\gamma$  sample.

Because processes other than  $Z \rightarrow ee$  decays can be included in this sample, a fit is performed on each invariant mass distribution. The shape of the invariant mass peak in  $Z \rightarrow ee$  MC is used as a signal template, and an electron + muon control region in data is used as the background template. After fitting the  $ee$  and  $e\gamma$  distributions in data, the integrals of the signal shapes between 80 and 100 GeV are referred to as  $N_{ee}^Z$  and  $N_{e\gamma}^Z$ , respectively. Figure 8.8 shows the shapes of the data distributions as well as the background and signal templates with arbitrary normalization. The results after the fit are shown as well.

The maximum of the background shape at 90 GeV arises from the  $p_T$  thresholds of the tag and probe objects. The  $p_T$  distributions tend to be sharply falling, so most probes have a  $p_T$  around 40 GeV and most tags have a  $p_T$  around 30 GeV. Changing the probe  $p_T$  threshold changes the position of the peak in the background template.

The ratio  $R_{\gamma/e} = N_{e\gamma}^Z/N_{ee}^Z$  is calculated from the fits. For the entire spectrum,  $R_{\gamma/e} = 2.63\%$ . The fits are also performed in bins of different kinematic variables to investigate any potential dependencies of  $R_{\gamma/e}$ . These are shown in Figures 8.9 and 8.10.

A 30% systematic uncertainty is assigned to the ratio  $R_{\gamma/e}$  to cover the observed dependencies. This is shown in the red bands in Figures 8.9 and 8.10. The strongest dependency is with respect to the vertex multiplicity. Recall that vertex multiplicity is a proxy for the number of pileup interactions in the event. For a similar analysis that used this fake rate [28],  $R_{\gamma/e}$  was parameterized as a function of the vertex multiplicity, but no significant difference was found between applying the linear fit shown in Figure 8.10 rather than the constant value of 2.63%.

The distinctive features of  $R_{\gamma/e}$  as a function of  $|\eta|$  are due to detector effects. Similar features are seen in data quality monitoring plots showing the occupancy of the ECAL crystals as a function of  $\eta$  and  $\phi$ .

### 8.4.3 Calculating the EWK estimate

The number of observed  $Z \rightarrow ee$  events in the  $e\gamma$  and  $ee$  mass spectra can be expressed as a function of  $f_{e \rightarrow \gamma}$ , the rate at which electrons are misidentified as photons:

$$\begin{aligned} N_{e\gamma}^Z &= f_{e \rightarrow \gamma}(1 - f_{e \rightarrow \gamma})N'_Z \\ N_{ee}^Z &= (1 - f_{e \rightarrow \gamma})^2 N'_Z \end{aligned} \quad (8.3)$$

where  $N'_Z$  is the actual number of  $Z \rightarrow ee$  events.

If we consider a generic sample of  $N'_{e\gamma}$  events with a real photon and a real electron, the number of events  $N_{e\gamma}$  that will actually get reconstructed as having one photon and one electron is given by:

$$N_{e\gamma} = (1 - f_{e \rightarrow \gamma})N'_{e\gamma} \quad (8.4)$$

The background  $N_{\gamma\gamma}$ , the fraction of  $N'_{e\gamma}$  that ends up in the candidate  $\gamma\gamma$  sample, can be written as

$$N_{\gamma\gamma} = f_{e \rightarrow \gamma}N'_{e\gamma} = N_{e\gamma} \frac{f_{e \rightarrow \gamma}}{1 - f_{e \rightarrow \gamma}} \quad (8.5)$$

By looking at Equation 8.3, it can be seen that this last factor is precisely the ratio  $R_{\gamma/e} = N_{e\gamma}^Z/N_{ee}^Z = 2.63\%$  that was discussed above. Thus, the EWK background prediction is simply given by  $R_{\gamma/e}$  times the number of observed  $e\gamma$  events.

### 8.4.4 EWK results

Table 8.2 shows the final EWK estimate in the six signal region bins. There are two sources of uncertainty on the EWK estimate: the statistical uncertainty from the limited number of  $e\gamma$  events observed in data and the 30% systematic uncertainty from the calculation of the fake rate. Both of these are included in the uncertainties listed in Table 8.2 .

TABLE 8.2  
EWK BACKGROUND ESTIMATE

$E_T^{\text{miss}}$ (GeV)	Expected number of EWK events
100 – 115	$13.7 \pm 4.2$
115 – 130	$9.0 \pm 2.7$
130 – 150	$7.4 \pm 2.3$
150 – 185	$6.1 \pm 1.9$
185 – 250	$5.8 \pm 1.8$
$\geq 250$	$3.3 \pm 1.0$

### 8.5 Irreducible background

In addition to the QCD and EWK backgrounds, there is a small irreducible background from  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  events. These events have inherent  $E_T^{\text{miss}}$ , and are therefore not included in the QCD background. These events also have two real photons, meaning they are not included in the EWK background either. We model this background using MC simulation and assign a 50% uncertainty to the estimate to cover any potential mis-modeling. The predicted background from  $Z\gamma\gamma$  events is shown in Table 8.3.

### 8.6 Systematic uncertainties

The systematic uncertainties fall into two main categories: those associated with one of the background estimates and those associated with the expected signal yields.

TABLE 8.3  
  
 $Z\gamma\gamma$  BACKGROUND ESTIMATE

$E_T^{\text{miss}}$ (GeV)	Expected number of $Z\gamma\gamma$ events
100 – 115	$1.3 \pm 0.6$
115 – 130	$1.1 \pm 0.6$
130 – 150	$1.1 \pm 0.6$
150 – 185	$1.3 \pm 0.7$
185 – 250	$1.3 \pm 0.6$
$\geq 250$	$1.1 \pm 0.6$

#### 8.6.1 Uncertainties on background estimates

The largest uncertainties on the total background estimate come from the QCD prediction method. These are summarized in Table 8.4 for each signal bin. All uncertainties are expressed as numbers of events.

The first uncertainty on the QCD estimate comes from the di-EM  $p_T$  reweighting procedure. Recall that to correct for differences in hadronic activity in the events, the  $E_T^{\text{miss}}$  distribution of the  $ff$  sample is reweighted by the  $\gamma\gamma/ff$  di-EM  $p_T$  ratio, so that its di-EM  $p_T$  spectrum matches that of the double photon candidate sample. To propagate the uncertainty from the di-EM  $p_T$  reweighting, we generate one thousand different di-EM  $p_T$  ratios. In each case, the value of the ratio in the bin is obtained by varying the nominal di-EM  $p_T$  ratio using a Gaussian distribution with  $\sigma$  equal to the statistical uncertainty of that bin. We then reweight the  $E_T^{\text{miss}}$  distributions of the  $ff$  control sample with the 1000 generated di-EM  $p_T$  ratios. Final uncertainties are determined from the fluctuations in each  $E_T^{\text{miss}}$  bin by finding the range that includes 68% of the new  $E_T^{\text{miss}}$  values.

TABLE 8.4  
UNCERTAINTIES ON THE QCD BACKGROUND ESTIMATE

$E_T^{\text{miss}}$ (GeV)	QCD Estimate	Stat. Uncert.	Shape Uncert.	Di-EM $p_T$ Uncert.
100 – 115	163.8	+12.0, -11.2	±64.1	±1.3
115 – 130	57.3	+7.4, -6.6	±25.6	±0.4
130 – 150	45.8	+7.70, -6.7	±27.7	±0.9
150 – 185	28.9	+6.6, -5.4	±19.3	±0.5
185 – 250	10.3	+4.1, -3.1	±7.3	±0.1
$\geq 250$	6.6	+5.2, -3.1	±5.8	±0.3

Another systematic uncertainty on the QCD background estimation comes from the uncertainty in the overall  $E_T^{\text{miss}}$  shape. To estimate this uncertainty, we use the ratio method described in Section 8.3.2 as a cross check. By fitting the ratio of  $\gamma\gamma$  to  $ff$  at low  $E_T^{\text{miss}}$  to a constant, the expected number of  $\gamma\gamma$  events at high  $E_T^{\text{miss}}$  can be extrapolated from the observed number of  $ff$  events. This procedure gives us a second estimate for the QCD background in each bin of  $E_T^{\text{miss}}$ . The difference between the two estimates is taken as a symmetric systematic uncertainty and is shown in the second-to-last column of Table 8.4.

Finally, it is worth noting that the dominant source of uncertainty on the QCD prediction comes from the limited  $ff$  statistics. Improving these statistics, perhaps through the use of a high- $p_T$  trigger with looser ID requirements, would be an effective way to improve future iterations of this analysis.

As described above, there are two uncertainties on the EWK background: the statistical uncertainty from the limited statistics in the control region and the 30%

uncertainty on the  $e \rightarrow \gamma$  fake rate. For the  $Z\gamma\gamma$  background, the only uncertainty considered is a 50% uncertainty to cover any potential mismodeling of the  $E_T^{\text{miss}}$  shape or uncertainty on the  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  cross section.

### 8.6.2 Other sources of systematic uncertainties

In addition to the systematic uncertainties arising from the background estimation techniques described above, there are other systematic uncertainties that affect the final analysis sensitivity. These uncertainties are summarized in Table 8.5. Ranges of uncertainty arise when there are different values of the uncertainty for different signal points.

The first is the uncertainty in the parton distribution functions (PDF) and the variation of cross section ratios (K factors) between leading order PDF's and next-to-leading order PDF's. The PDF uncertainties are taken from the NNPDF3.0 variations [9, 11].

Other uncertainties on the signal include finite MC statistics and the photon data/MC scale factor described in Section 7.5. There is also a 2.5% uncertainty on the integrated luminosity of the data sample [33].

The last source of uncertainty comes from propagating the jet energy scale uncertainty described in Section 6.3.1 to the  $E_T^{\text{miss}}$ . The expected signal yields are recalculated using the  $1\sigma$  fluctuations of  $E_T^{\text{miss}}$  if the jet energy scale is varied by its uncertainty. The difference between the yields with and without the jet energy scale changes is taken as a systematic uncertainty.

TABLE 8.5  
UNCERTAINTIES ON SIGNAL YIELD

Systematic Uncertainty	[%]
PDF error on cross section	19 - 35
Finite MC statistics	0 - 16
Photon Data/MC scale factor	2.4
Integrated luminosity	2.3
Jet energy scale	0 - 30

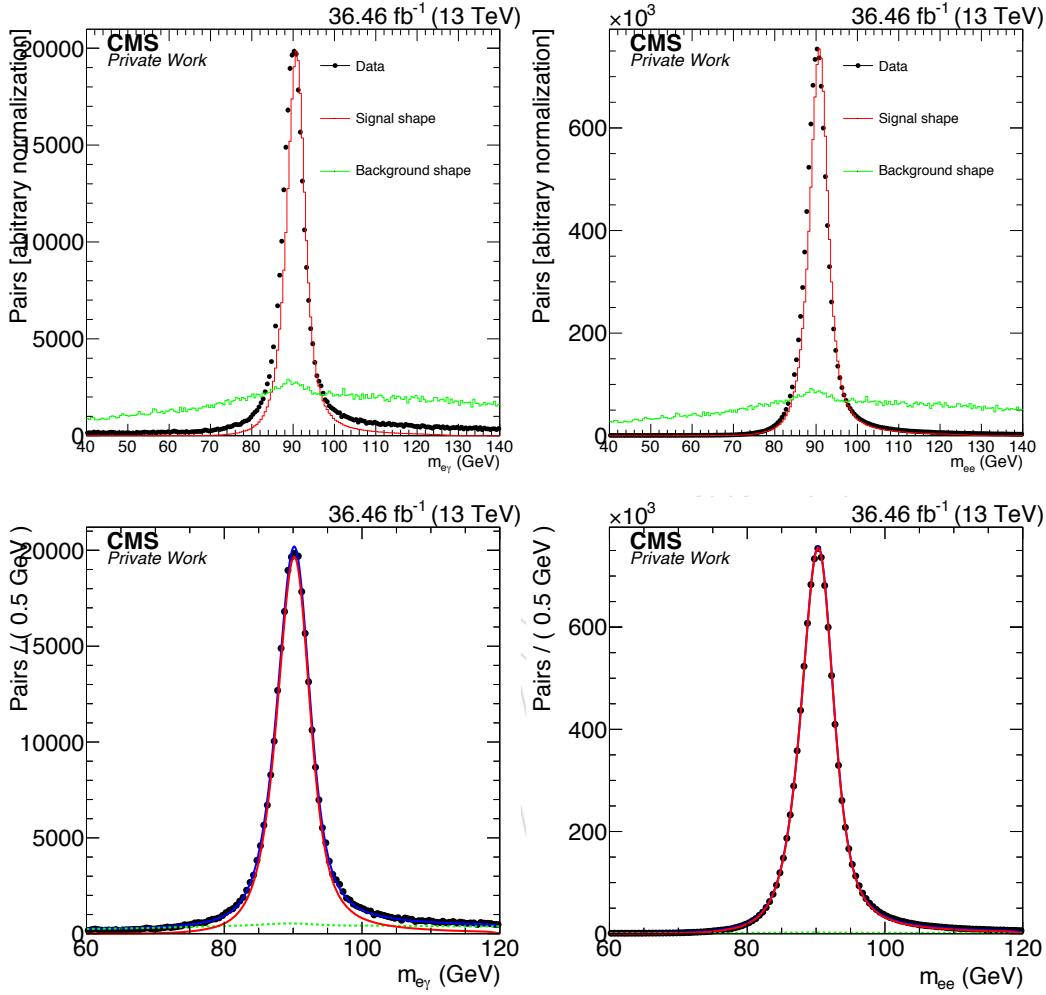


Figure 8.8: Invariant mass distributions for the  $e\gamma$  (left) and  $ee$  (right) samples. Each plot includes the data (black markers), the signal template (red), and the background template (green). The top plots demonstrate the shape of each distribution with arbitrary normalization, and the bottom plots show the distributions after the fits are performed. Reprinted from Reference [56].

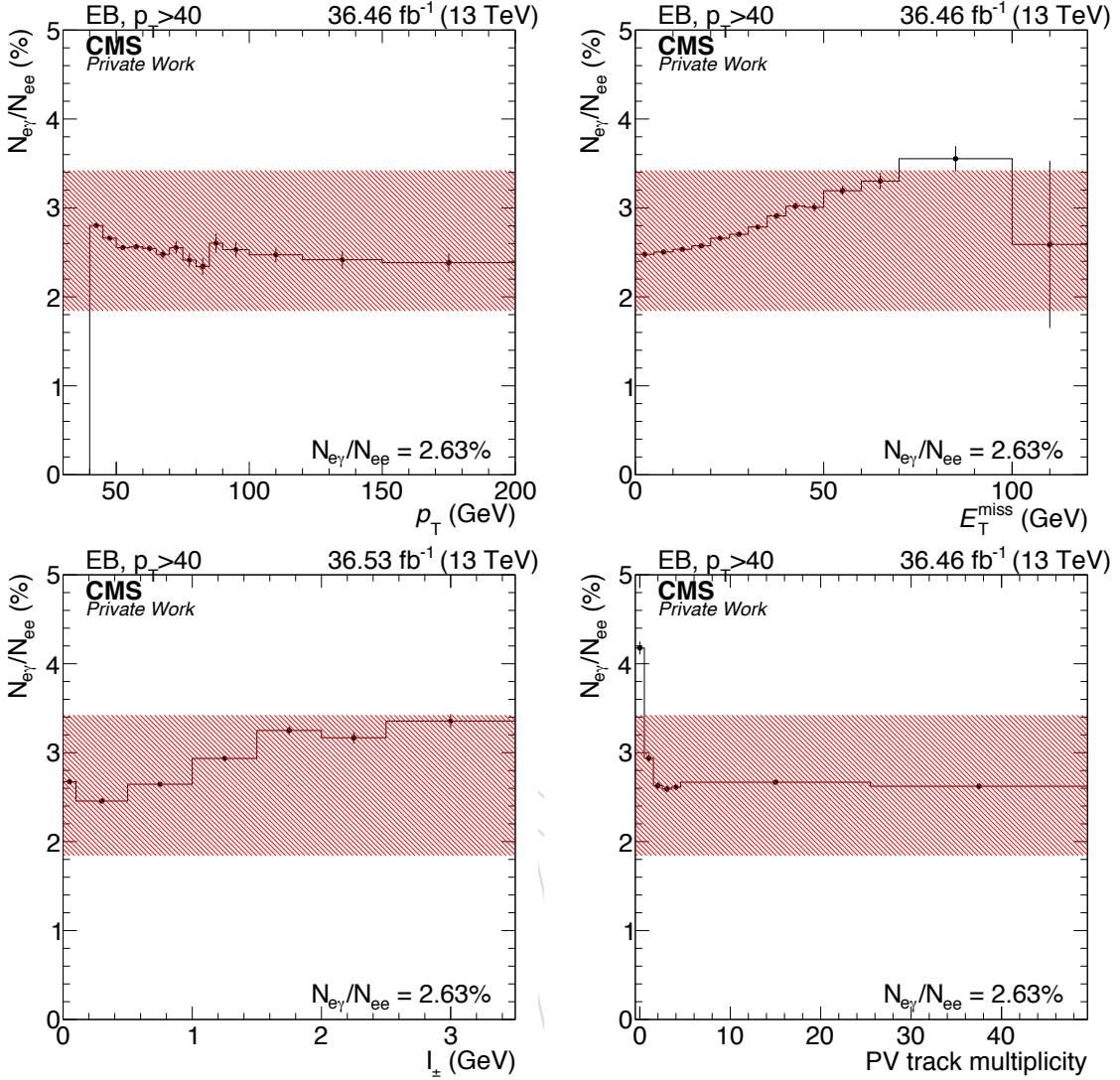


Figure 8.9: Value of  $R_{\gamma/e}$  for probes in the barrel as a function of various kinematic variables: probe  $p_T$ ,  $E_T^{\text{miss}}$ , isolation  $I_{\pm}$ , and track multiplicity of the primary vertex (PV). The red band corresponds to a 30% uncertainty on  $R_{\gamma/e}$ . Reprinted from Reference [56].

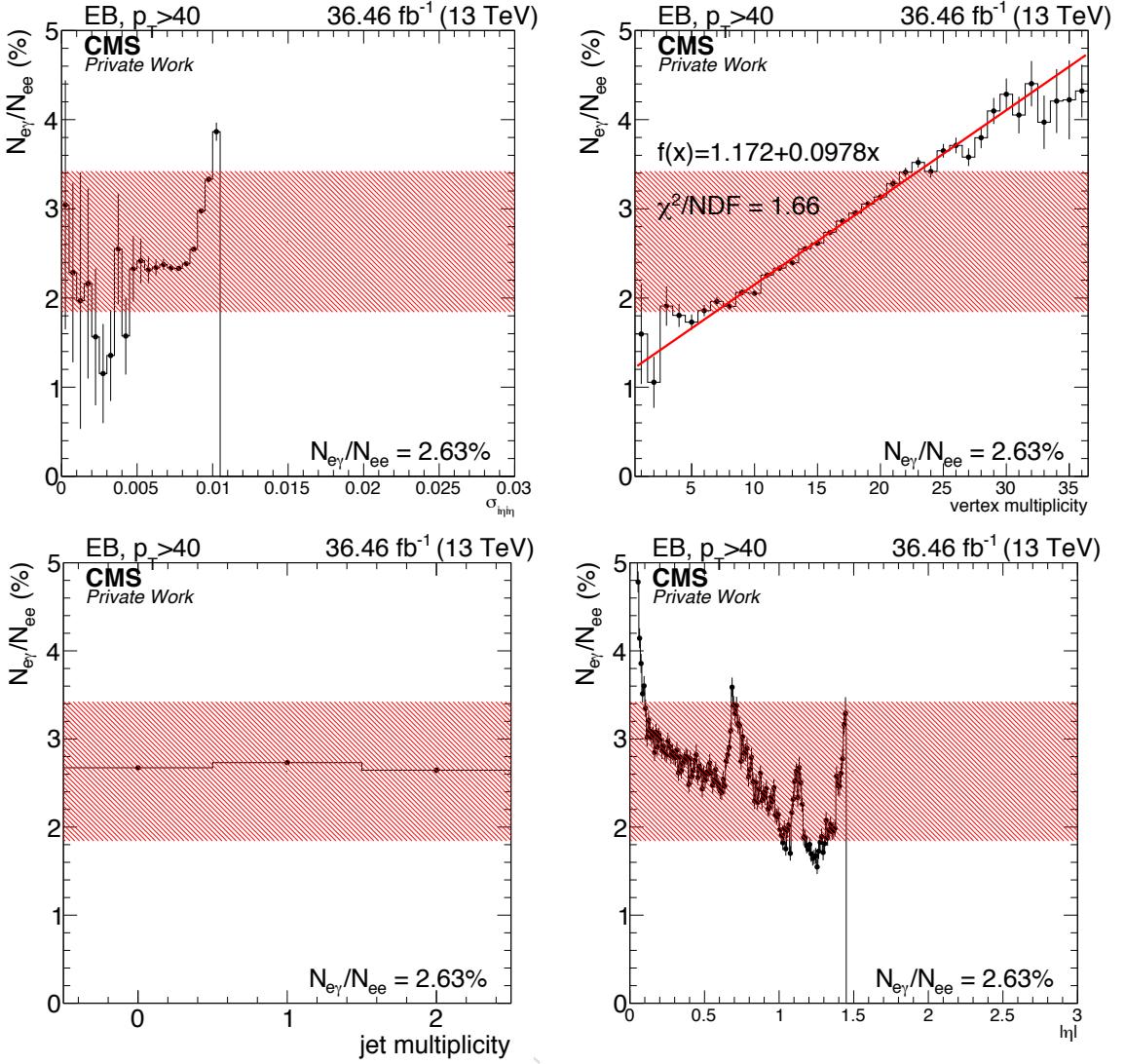


Figure 8.10: Value of  $R_{\gamma/e}$  for probes in the barrel as a function of various kinematic variables:  $\sigma_{i\eta i\eta}$  of the probe, vertex multiplicity, jet multiplicity, and probe  $|\eta|$ . The red band corresponds to a 30% uncertainty on  $R_{\gamma/e}$ . Reprinted from Reference [56].

## CHAPTER 9

### RESULTS AND INTERPRETATIONS

#### 9.1 Prediction versus observation

After deriving estimates for all standard model backgrounds (summarized in Table 9.1), we compare the number of observed events to the expected number of background events. The  $E_T^{\text{miss}}$  distributions for the full background prediction and the blinded data are shown in Figure 9.1. Two benchmark signal models are also shown, corresponding to the gluino pair production simplified model with gluino mass equal to 1700 GeV or 2000 GeV.

Table 9.2 shows the total expected and observed numbers of events for each bin in the signal region. Within the uncertainties, no excess is observed with respect to the Standard Model prediction.

#### 9.2 Simplified models

Two simplified models are used in the interpretation of the results. The T5gg simplified model assumes gluino ( $\tilde{g}$ ) pair production and the T6gg model assumes squark ( $\tilde{q}$ ) pair production. Example decay chains for both models are shown in Figure 9.2.

In both models, the lightest supersymmetric particle (LSP) is the gravitino,  $\tilde{G}$ , which is taken to be nearly massless. The next-to-lightest supersymmetric particle

TABLE 9.1  
FINAL BACKGROUND ESTIMATE

$E_T^{\text{miss}}$ (GeV)	QCD	EWK	$Z\gamma\gamma$	Total
100 – 115	$163.8^{+65.2}_{-65.1}$	$13.7 \pm 4.2$	$1.3 \pm 0.6$	$178.8^{+65.3}_{-65.2}$
115 – 130	$57.3^{+26.7}_{-26.4}$	$9.0 \pm 2.7$	$1.1 \pm 0.6$	$67.4^{+26.7}_{-26.5}$
130 – 150	$45.8^{+28.8}_{-28.5}$	$7.4 \pm 2.3$	$1.1 \pm 0.6$	$54.3^{+28.9}_{-28.6}$
150 – 185	$28.9^{+20.4}_{-20.0}$	$6.1 \pm 1.9$	$1.3 \pm 0.7$	$36.3^{+20.5}_{-20.2}$
185 – 250	$10.3^{+8.4}_{-7.9}$	$5.8 \pm 1.8$	$1.3 \pm 0.6$	$17.4^{+8.6}_{-8.1}$
$\geq 250$	$6.6^{+7.8}_{-6.6}$	$3.3 \pm 1.0$	$1.1 \pm 0.6$	$11.0^{+7.9}_{-6.7}$

(NLSP) is the neutralino,  $\tilde{\chi}_1^0$ . The models assume 100% branching fractions for  $\tilde{\chi}_1^0 \rightarrow \tilde{G}\gamma$  and  $\tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$  and  $\tilde{q} \rightarrow q\tilde{\chi}_1^0$ .

In order to study the expected SUSY signal distributions, two signal Monte Carlo scans were produced. The T5gg scan was produced in bins of gluino mass and neutralino mass, and the T6gg scan was produced in bins of squark mass and neutralino mass. The leading-order event generator `MADGRAPH5_aMC@NLO` [4] is used to simulate the signal samples, which were generated with either two gluinos or two squarks and up to two additional partons in the matrix element calculation. The parton showering, hadronization, multiple-parton interactions, and the underlying event were described by the `PYTHIA 8` [62] program with the CUETP8M1 generator tune. The detector response is simulated using CMS fast simulation [1].

A total of 40,000 events were produced for each bin, except for bins with gluino or squark masses above 2.0 TeV, where only 20,000 events were produced per bin. For gluino masses from 1,400 to 2,500 GeV, events were generated in bins of 50 GeV. In the T6gg scan, the squark masses ranged from 1,400 GeV to 2,050 GeV in bins

TABLE 9.2  
EXPECTED AND OBSERVED EVENTS IN THE SIGNAL REGION

$E_T^{\text{miss}}$ (GeV)	Total exp.	T5gg (1700,1000)	T5gg (2000,1000)	Observed
100 – 115	$178.8^{+65.3}_{-65.2}$	$0.3 \pm 0.1$	$0.1 \pm 0.1$	105
115 – 130	$67.4^{+26.7}_{-26.5}$	$0.2 \pm 0.1$	$0.1 \pm 0.1$	39
130 – 150	$54.3^{+28.9}_{-28.6}$	$0.4 \pm 0.1$	$0.1 \pm 0.1$	21
150 – 185	$36.3^{+20.5}_{-20.2}$	$0.9 \pm 0.1$	$0.2 \pm 0.1$	21
185 – 250	$17.4^{+8.6}_{-8.1}$	$2.0 \pm 0.2$	$0.4 \pm 0.1$	11
$\geq 250$	$11.0^{+7.9}_{-6.7}$	$31.0 \pm 1.2$	$6.8 \pm 0.3$	12

of 50 GeV. The neutralino masses ranged from 10 GeV up to the mass of the gluino or squark and were binned in 100 GeV segments. Finer binning was used in the compressed region where  $M_{\tilde{\chi}_1^0}$  is within 300 GeV of  $M_{\tilde{g}}$  or  $M_{\tilde{q}}$ , and in the region with low  $M_{\tilde{\chi}_1^0}$ . These mass ranges were selected to overlap and expand upon the mass ranges excluded by previous searches [6, 15].

The parton distribution functions are obtained from NNPDF3.0 [9]. Figure 9.3 shows the cross sections as a function of squark and gluino mass. The cross sections are calculated at next-to-leading order (NLO) accuracy including the resummation of soft gluon emission at next-to-leading-logarithmic (NLL) accuracy [57, 10], with all the unconsidered sparticles assumed to be heavy and decoupled. The uncertainties on the cross sections are calculated as described in Reference [11].

### 9.3 Signal acceptance and efficiency

The generated SUSY signal events undergo the same event selection described in Chapter 7. For every mass point, the number of events expected in each of the

six signal region bins is calculated. In addition, the overall acceptance  $\times$  efficiency ( $A \times \epsilon$ ), defined as the number of events passing the full  $\gamma\gamma$  selection divided by the total number of generated events, is calculated across the full 2D mass plane. Figures 9.4 and 9.5 show the  $A \times \epsilon$  distributions for the T5gg and T6gg simplified model frameworks.

At low neutralino mass, the  $A \times \epsilon$  decreases because the photons are softer and more likely to fail the  $p_T > 40$  GeV cut or the  $m_{\gamma\gamma} > 105$  GeV cut. There is also more jet energy at low neutralino masses, which increases the chance that photons will overlap with a jet and fail the isolation requirements.

## 9.4 Limits and interpretations

As shown in Table 9.2, the observed number of events is consistent with (or less than) the total number of expected background events in each bin in the signal region. In the absence of a signal, these results can be used to set limits on the allowed squark, gluino, and neutralino masses in the T5gg and T6gg simplified models.

### 9.4.1 Statistical analysis

The probability  $P$  of observing  $n$  events if we expect  $s$  signal events and  $b$  background events in a given bin or channel can be expressed using a Poisson distribution:

$$P(n|\mu, \boldsymbol{\theta}) = \frac{(\mu s(\boldsymbol{\theta}) + b(\boldsymbol{\theta}))^n e^{-\mu s(\boldsymbol{\theta}) - b(\boldsymbol{\theta})}}{n!} \quad (9.1)$$

The expected number of signal events is parameterized with a *signal strength modifier*  $\mu$ , where  $\mu = 1$  corresponds to the nominal cross section. The systematic and statistical uncertainties described in Chapter 8 are treated as nuisance parameters, and their effects on  $s$  and  $b$  are encoded in the parameter set  $\boldsymbol{\theta}$ .

Each nuisance parameter  $\theta_i$  must have an associated probability distribution  $p_i(\theta)$ .

The statistical uncertainties are modeled with gamma distributions, where  $x$  is the expected event yield,  $\alpha$  is the original number of events in the control sample, and  $\beta$  is the transfer factor,  $\beta = x/\alpha$ :

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (9.2)$$

The statistical uncertainties include those on the observed number of events in the signal region in data, uncertainties from the limited number of events in the control regions used for the data-driven background estimation methods, and the MC statistical uncertainties in the SUSY signal MC and the MC used to model the  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  background.

All of the systematic uncertainties described in Chapter 8 are constrained with log-normal probability distributions. The peak of the probability distribution is set to the nominal background estimate for that bin, and the size of the systematic uncertainty sets the width of the distribution.

To combine several channels or bins—in our case, the six signal bins for  $E_T^{\text{miss}} > 100 \text{ GeV}$ —the probabilities are combined into a *likelihood function*  $\mathcal{L}$ . The likelihood is the probability of having a signal strength  $\mu$  and nuisance parameters  $\boldsymbol{\theta}$  given the observed number of events in each bin,  $\mathbf{n} = \{n_1, n_2, \dots, n_i\}$ :

$$\mathcal{L}(\mu, \boldsymbol{\theta} | \mathbf{n}) = \prod_i P(n_i | \mu, \boldsymbol{\theta}) \prod_j p_j(\boldsymbol{\theta}) \quad (9.3)$$

Maximizing  $\mathcal{L}$  gives us the best fit values for the signal strength and nuisance parameters, referred to as  $\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$ , respectively.

To describe how compatible a given value of  $\mu$  is with the observed data, we use the following test statistic:

$$t_\mu = -2 \ln \frac{\mathcal{L}(\mu, \tilde{\boldsymbol{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad (9.4)$$

where  $\tilde{\boldsymbol{\theta}}$  is the value of the nuisance parameters that maximizes  $\mathcal{L}(\mu)$ . Larger values of  $t_\mu$  represent increased tension between  $\mu$  and the observed data.

The probability of getting a value for  $t_\mu$  larger than what was actually observed is known as the  $p$ -value,  $p_\mu$ . If  $p_{\mu=1}$  is less than 0.05, then we say that we have excluded the signal hypothesis at a 95% *confidence level*. In order to claim a discovery, the convention in high energy physics is to require  $p_{\mu=0} < 2.87 \times 10^{-7}$ , which corresponds to  $5\sigma$  in a Gaussian distribution.

In this analysis, we use the standard  $CL_s$  method [55, 61] to determine 95% confidence level intervals for the cross section times branching fraction in each mass bin. The  $CL_s$  method avoids over-estimating the limits in the case of downward fluctuations in data by normalizing the  $p$ -value for the signal plus background hypothesis ( $CL_{s+b} = p_{\mu=1}$ ) to the probability of the background-only hypothesis ( $CL_b = p_{\mu=0}$ ):

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (9.5)$$

If  $1 - CL_s \leq 0.95$ , then we have excluded the signal hypothesis at a 95% confidence level.

#### 9.4.2 T5gg and T6gg limits

The final cross section limits and exclusion contours are shown in Figure 9.6 for the T5gg simplified model and Figure 9.7 for the T6gg simplified model. In both models, the cross section is independent of the mass of the neutralino and depends only on the mass of the gluino or squark. Therefore, it is not surprising that the exclusion contours do not depend sensitively on the mass of the neutralino.

In the bulk of the distribution, gluino masses below 1.87 TeV and squark masses below 1.60 TeV are excluded at a 95% confidence level. The expected exclusions were 1.93 TeV for gluino pair production and 1.65 TeV for squark pair production.

This is a large jump in sensitivity compared to the 2015 ATLAS and CMS results, which excluded masses less than 1.65 TeV and 1.37 TeV for gluinos and squarks, respectively.

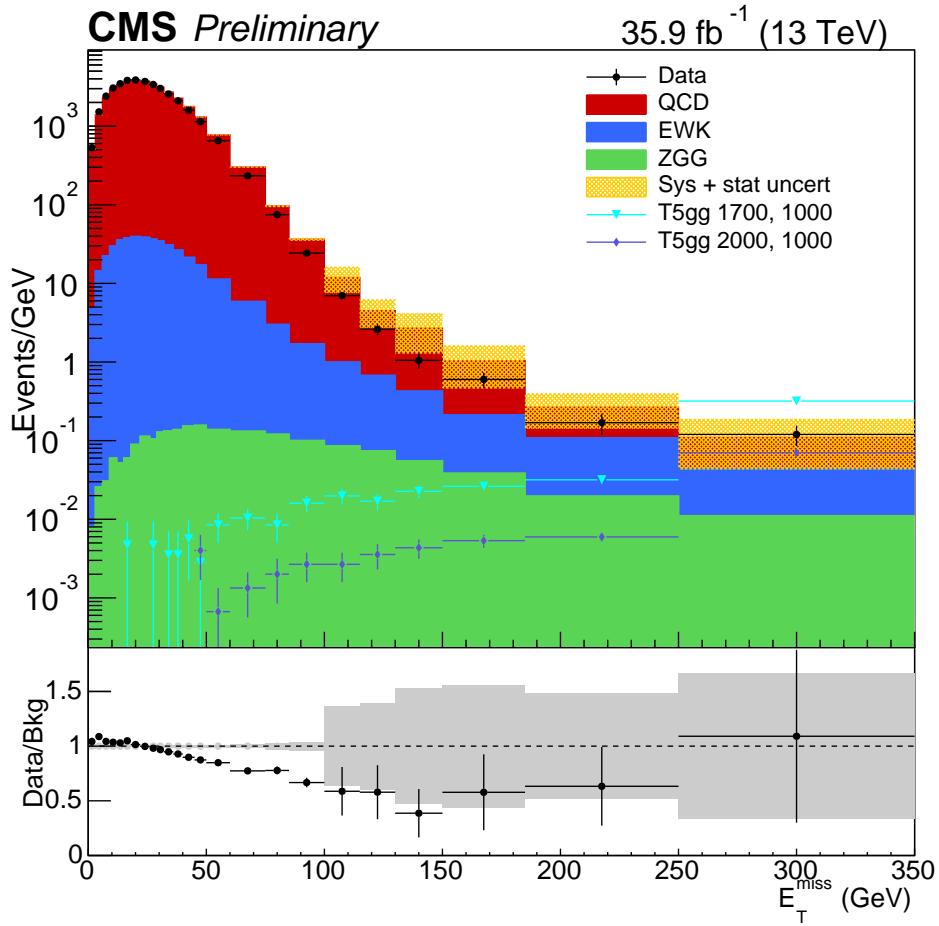


Figure 9.1:  $E_T^{\text{miss}}$  distributions for the full background estimation and the observed data. The black points represent the observed  $E_T^{\text{miss}}$  distribution. The QCD background (red), the EWK background (blue), and the  $Z\gamma\gamma$  background (green) are displayed as a stacked histogram. The combined uncertainty on the background estimation is shown in yellow. The simulated  $E_T^{\text{miss}}$  distributions of two benchmark T5gg mass points are displayed as well: the  $M_{\tilde{g}} = 1700$  GeV,  $M_{\tilde{\chi}_1^0} = 1000$  GeV mass point is shown in cyan and the  $M_{\tilde{g}} = 2000$  GeV,  $M_{\tilde{\chi}_1^0} = 1000$  GeV signal is shown in purple.

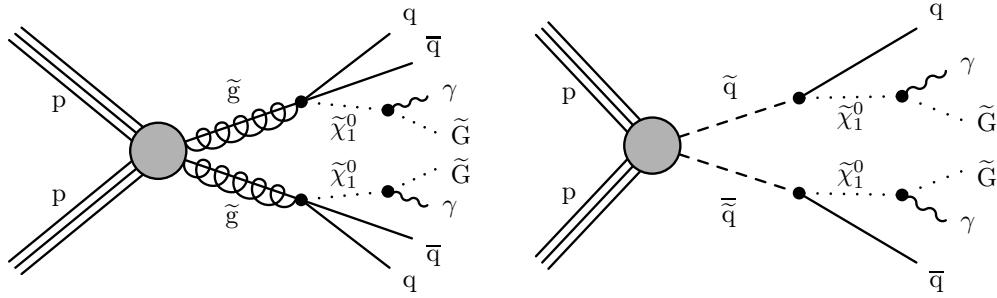


Figure 9.2: Diagrams showing the production of signal events in the collision of two protons  $p$ . In gluino  $\tilde{g}$  pair production in the T5gg simplified model (left), the gluino decays to an antiquark  $\bar{q}$ , quark  $q$ , and neutralino  $\tilde{\chi}_1^0$ . In squark  $\tilde{q}$  pair production in the T6gg simplified model (right), the squark decays to a quark and a neutralino. In both cases, the neutralino subsequently decays to a photon  $\gamma$  and a gravitino  $\tilde{G}$ .

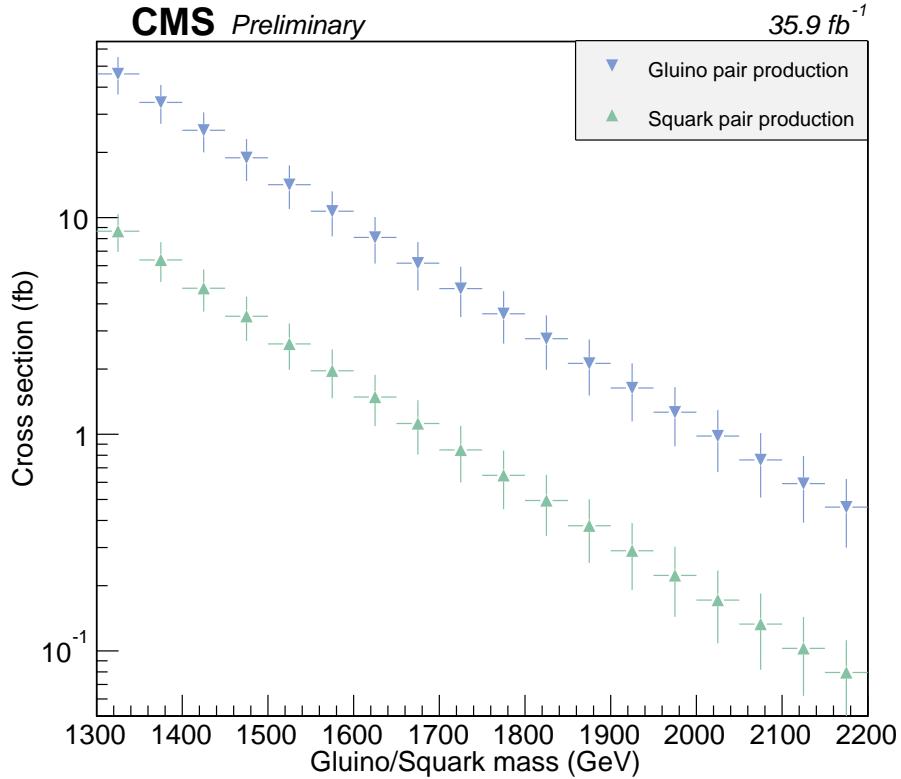


Figure 9.3: NLO+NLL cross sections for the simplified gluino pair production (blue) and squark pair production (green) models used in this analysis.

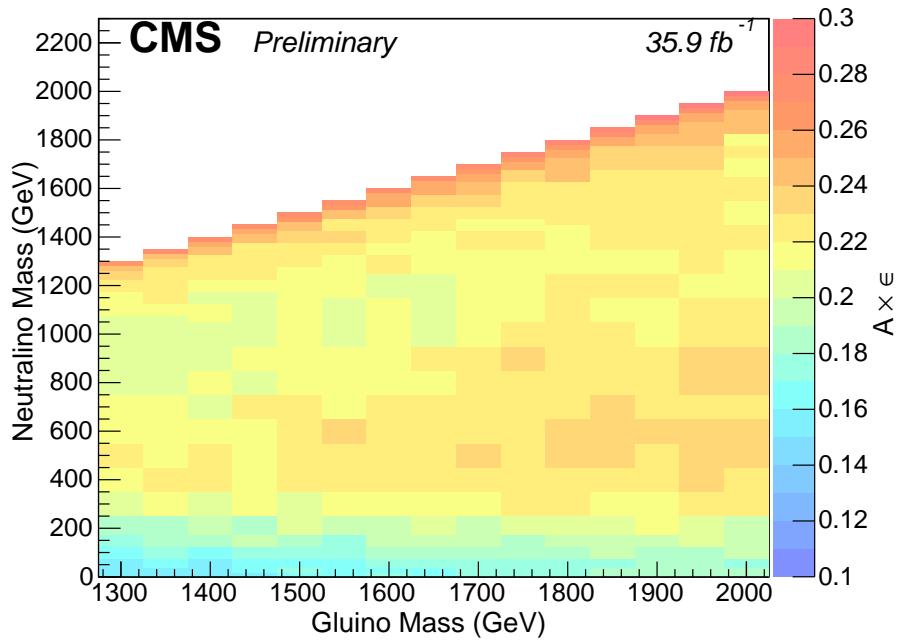


Figure 9.4: Acceptance  $\times$  efficiency for the T5gg simplified model.

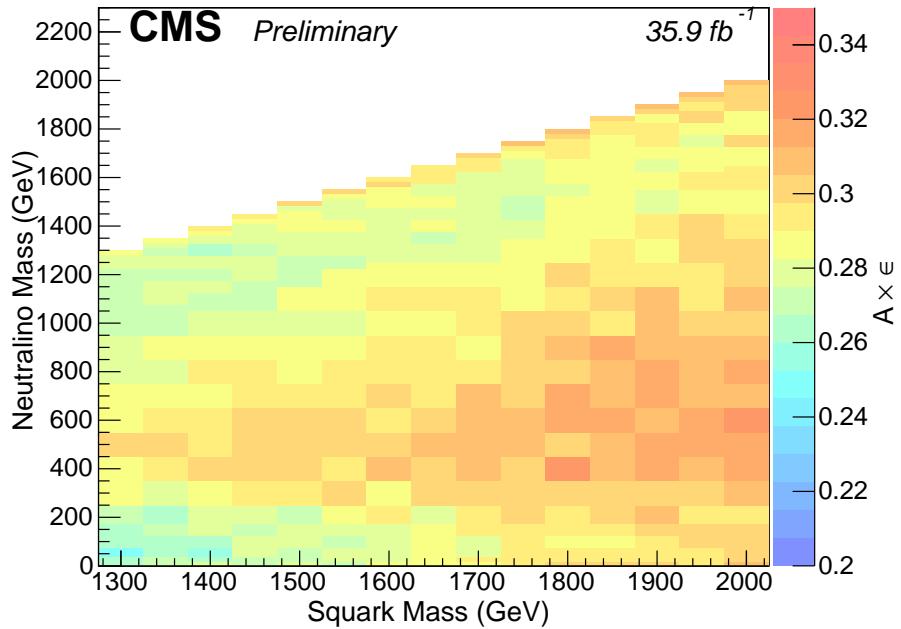


Figure 9.5: Acceptance  $\times$  efficiency for the T6gg simplified model.

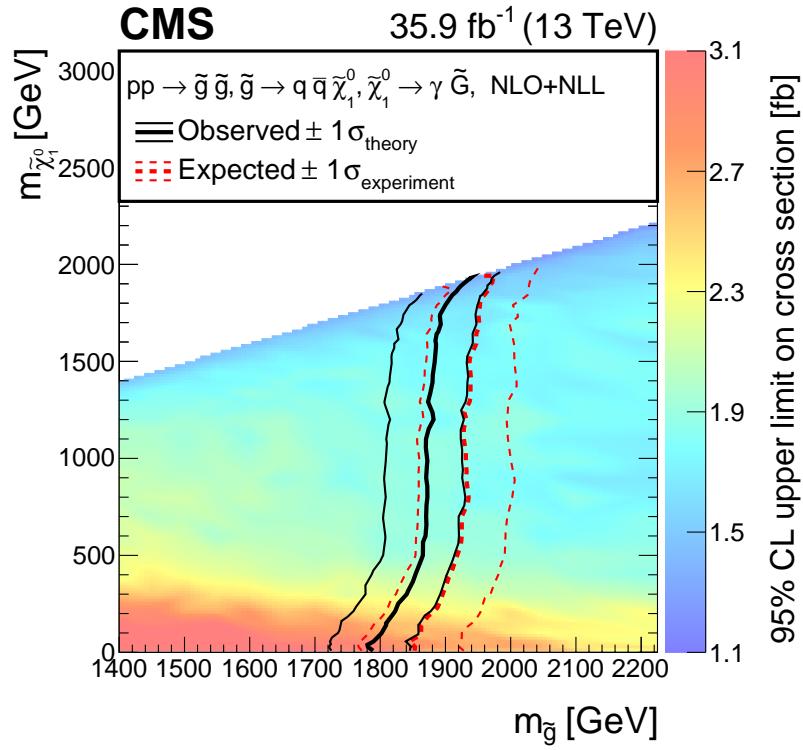


Figure 9.6: The 95% confidence level upper limits on the gluino pair production cross sections as a function of gluino and neutralino masses. The contours show the observed and median expected exclusions assuming the NLO + NLL cross sections, with their one standard deviation uncertainties.

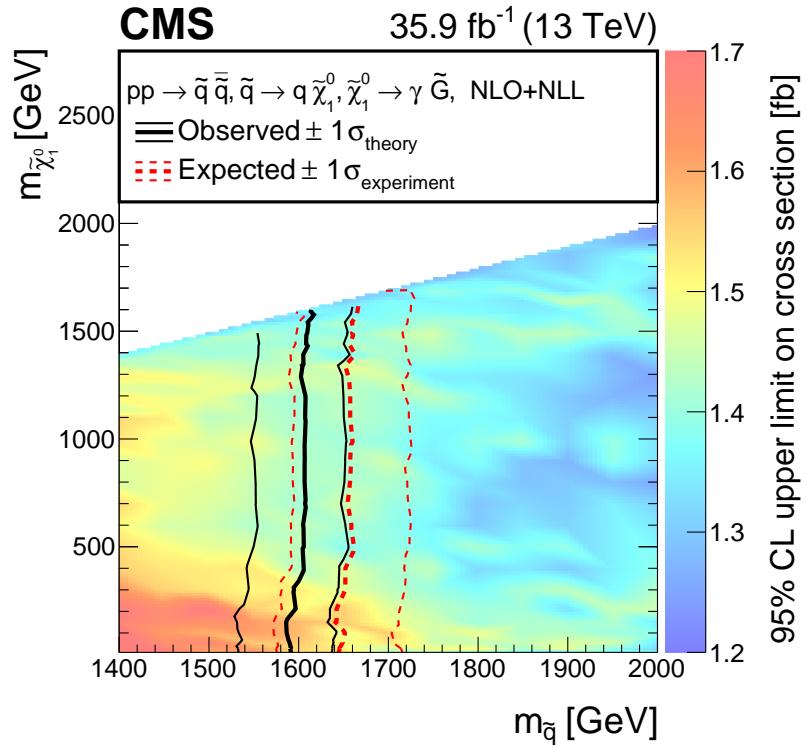


Figure 9.7: The 95% confidence level upper limits on the squark pair production cross sections as a function of squark and neutralino masses. The contours show the observed and median expected exclusions assuming the NLO + NLL cross sections, with their one standard deviation uncertainties.

## CHAPTER 10

### SUMMARY AND FUTURE PLANS

#### 10.1 Future improvements

There are a variety of ways in which this analysis could be improved in future versions. Looking at Table 8.4, it is obvious that our uncertainties are dominated by the limited statistics of the  $ff$  control sample. To acquire more statistics, it would be necessary to move to a different trigger, since our fake definition must be at least as tight as the online trigger requirements. Looser ID requirements often come at the cost of higher photon  $p_T$  thresholds. Because we have already pushed to high gluino and squark masses, however, tighter  $p_T$  thresholds would probably still be feasible.

The largest potential improvement would come from using more than one discriminating variable. Looking at the Feynman diagrams of Figure 9.2, one striking feature is the number of expected jets in the final states of both models. We currently do not use any jet information in the definition of our signal region. By using both  $E_T^{\text{miss}}$  and a hadronic variable such as  $H_T$  (the scalar  $p_T$  sum of all jets in the event), we could achieve a better separation of signal and background. This was studied at length using the 2016 data, but we faced issues with correlations between variables. Our efforts are summarized in Appendix B.

#### 10.2 GMSB combination paper

Efforts are ongoing to perform a combined analysis in the diphoton, photon plus lepton, and inclusive single photon channels [28, 29] to set limits on more realistic

GMSB models. The simplified models used in this analysis assumed 100% branching fraction of  $\tilde{\chi}_1^0 \rightarrow \gamma \tilde{G}$ , but in more complete GMSB models,  $\tilde{\chi}_1^0$  could also decay to a Higgs boson or a  $Z$  boson. Additionally, in many GMSB models the lightest chargino  $\tilde{\chi}_0^\pm$  is almost as light as the lightest neutralino. We refer to this as a co-NLSP. In such cases, we can end up with  $W^\pm$  bosons in the final state as well.

Combining the various photonic final states is not as easy as simply adding more channels to the likelihood of Equation 9.3. The first reason for this is that the signal regions of each analysis are not exclusive. A high- $E_T^{\text{miss}}$  event with two photons and a lepton would be counted in both the diphoton channel and the photon plus lepton channel. The second complication is that there are correlations in the background estimation methods of the different channels. For instance, one of the backgrounds of the photon plus lepton analysis includes events where an electron is misidentified as a photon. This is similar to our EWK background and is modeled in an equivalent way, leading to correlations in our uncertainties. These correlations need to be quantified and understood in order to set accurate limits.

Both of these difficulties make it impossible for an analyzer outside CMS to perform a combination based on the information in the published papers alone. Studies are currently being performed to optimize which channels should lay claim to the overlapping regions of phase space and to quantify the correlations between the methods. The goal is to have results approved in time for the 2018 summer conferences.

### 10.3 Conclusions

In this dissertation, a search for new physics in events with two photons and missing transverse momentum was described. The analysis was performed with 35.9  $\text{fb}^{-1}$  of data taken with the CMS detector in 2016. Fully data-driven background estimations were developed for the two primary backgrounds. For the QCD background, the background prediction was taken from a double fake control sample that

had been reweighted by the  $\gamma\gamma/ff$  di-EM  $p_T$  distribution to correct for differences in hadronic activity. For the EWK background, we first calculated the rate at which electrons were misidentified as photons by looking at the invariant mass spectrum in an  $ee$  sample and an  $e\gamma$  sample. This rate was then applied to an  $e\gamma$  control sample satisfying the same criteria as the candidate  $\gamma\gamma$  sample. Additionally, there is a small background from  $Z\gamma\gamma \rightarrow \nu\nu\gamma\gamma$  events that was modeled in simulation.

No evidence for physics beyond the standard model was observed, and limits were placed on the masses of supersymmetric particles in gauge-mediated supersymmetry breaking (GMSB) models. Results were interpreted in the context of two simplified models, one assuming gluino pair production and the other assuming squark pair production. Gluino masses below 1.87 TeV and squark masses below 1.60 TeV are excluded at a 95% confidence level. This is an improvement of 220 GeV compared to the 2015 CMS results.

## APPENDIX A

### ALTERNATE CONTROL SAMPLE FOR QCD BACKGROUND ESTIMATION

In the analysis performed using the 2015 CMS data set [15], a sample of  $Z \rightarrow ee$  events (referred to as the  $ee$  sample) was used as the primary QCD background estimate, and the  $ff$  control sample was used to set a systematic uncertainty on the shape of the  $E_T^{\text{miss}}$  tail. The idea was that the  $ee$  and  $ff$  control samples represent 100% and 0% photon purity, respectively. By using both, we could derive a bound on any possible sensitivity of the  $E_T^{\text{miss}}$  spectrum to the object purity.

#### A.1 Problems with using $ee$

There are several reasons we decided not to use the  $ee$  control sample in the 2016 analysis. One reason was uncertainty on the efficiency of the control trigger. As mentioned in Chapter 5, the primary analysis trigger cannot be used for  $Z \rightarrow ee$  events due to the  $m_{\gamma\gamma} > 90$  GeV invariant mass requirement. Instead, the  $ee$  sample uses the control trigger listed in Table 5.1. As discussed in Section 5.5.2, the control trigger is only 79.8% efficient because it requires both electromagnetic objects to be matched to a pixel seed. There is a relatively large uncertainty on this efficiency because of its slow turn-on. Our offline photon  $p_T$  threshold of 40 GeV is barely on the plateau, and the trigger continues to get more efficient as  $p_T$  increases.

Another reason why we prefer  $ff$  over  $ee$  is that the kinematics of  $ff$  are closer to the kinematics of our SUSY signal events. In  $ee$  events, the two electromagnetic objects are obviously correlated, since they come from the decay of the  $Z$  boson. In  $ff$  and SUSY diphoton events, on the other hand, the photons and fakes are

uncorrelated. The result is that the  $ee$  and  $\gamma\gamma$   $E_T^{\text{miss}}$  distributions start to diverge once the statistical uncertainties go down.

The most significant reason for not using  $ee$  was actually not the trigger or kinematics, but large contributions from processes with real  $E_T^{\text{miss}}$  such as  $t\bar{t}$  and  $ZZ \rightarrow ee\nu\nu$ . The premise of the QCD background estimation method is that we select control samples that do not have inherent  $E_T^{\text{miss}}$ . For the case of  $ee$ , however, there are processes with inherent  $E_T^{\text{miss}}$  that contaminate the sample. In  $t\bar{t}$  events where both tops decay leptonically, there can be final states with two electrons and multiple neutrinos. Similarly,  $ZZ \rightarrow ee\nu\nu$  events will fall into our  $ee$  category. These contributions are modeled with MC and subtracted from the  $ee$  distribution in data. The MC samples used are listed in Table A.1.

TABLE A.1

BACKGROUND MC SAMPLES FOR EE SAMPLE

/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/ RunIISummer16MiniAODv2-PUMoriond17_ 80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
/ZZTo2L2Nu_13TeV_powheg_pythia8/ RunIISummer16MiniAODv2-PUMoriond17_ 80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM

Table A.2 shows the percent contribution from these processes in the signal region. In some bins, 95% of the observed  $ee$  events can be attributed to these real- $E_T^{\text{miss}}$  processes. As you can see in Figure A.1, very few  $ee$  events remain in the tail after

performing this subtraction.

TABLE A.2  
CONTRIBUTIONS TO EE SAMPLE

$E_T^{\text{miss}}$ bin (GeV)	$t\bar{t}$ Contribution	$ZZ$ Contribution
100 – 115	56.1%	4.4%
115 – 130	82.8%	8.54%
130 – 150	75.8%	9.9%
150 – 185	80.8%	14.6%
185 – 250	50.9%	16.8%
> 250	33.4%	27.7%

For the subtraction of  $ZZ \rightarrow ee\nu\nu$  events, we consider uncertainties from MC statistics and jet energy scale corrections. Both uncertainties are less than a 3% effect on the final  $ee$  estimate in the signal region.

The systematic uncertainties associated with the subtraction of the  $t\bar{t}$  contribution to the  $ee$  control sample are much more significant. Because we are subtracting a large majority of the original  $ee$  events at high  $E_T^{\text{miss}}$ , the relative uncertainties on the expected number of  $ee$  events get inflated. In the signal region, the limited MC  $t\bar{t}$  statistics give rise to an uncertainty of up to 140%. Uncertainties from the parton distribution function and overall scale of the cross section lead to uncertainties on the final  $ee$  estimate up to 73%. Jet energy scale corrections applied to the  $t\bar{t}$  sample lead to uncertainties up to 60%, and the effect of reweighting by the top  $p_T$  to correct

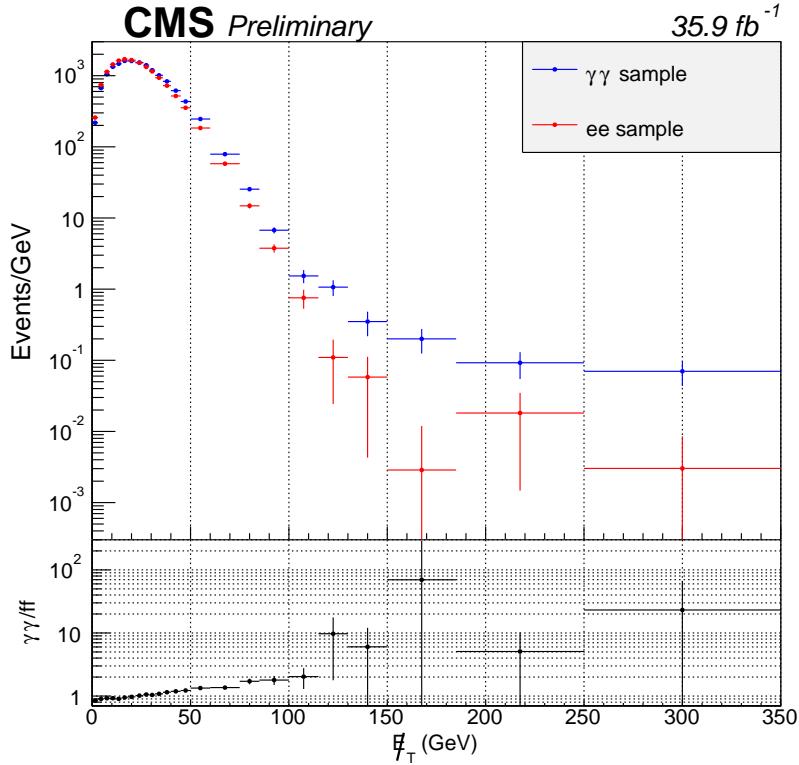


Figure A.1: Unweighted  $E_T^{\text{miss}}$  distributions of the  $ee$  control sample and the  $\gamma\gamma$  candidate sample. The  $ee$  distribution has been normalized to the  $E_T^{\text{miss}} < 50 \text{ GeV}$  region of the candidate sample. Contributions from  $t\bar{t}$  and  $ZZ$  events have been subtracted.

for differences between MC and data is up to 70%.

The large corrections and associated uncertainties to make the  $ee$  sample representative of the events we were trying to model was the primary reason why we ultimately decided not to use the  $ee$  control sample for the QCD background estimate.

## A.2 Results using ee sample

As was done for the  $ff$  sample, any difference in hadronic activity between the control and candidate samples needs to be corrected. Figure A.2 shows the  $ee$  and  $\gamma\gamma$  di-EM  $p_T$  distributions. The  $E_T^{\text{miss}}$  distributions after reweighting are shown in Figure A.3. It is clear from the plot that even after di-EM  $p_T$  reweighting, the  $ee$

sample systematically underestimates the QCD background.

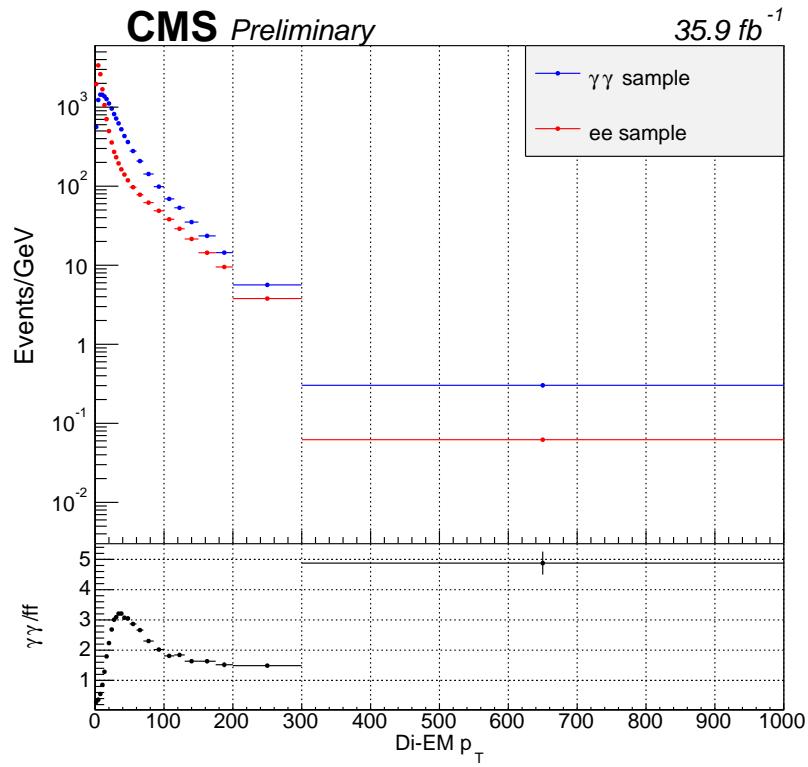


Figure A.2: Comparison of the di-EM  $p_T$  distributions of the  $\gamma\gamma$  and  $ee$  samples.

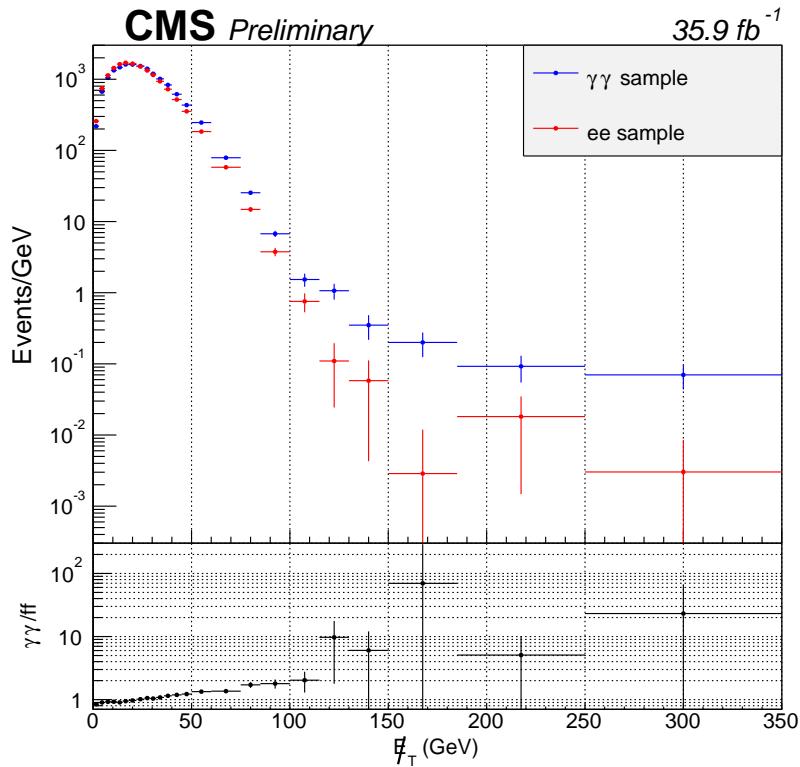


Figure A.3:  $E_T^{\text{miss}}$  distributions of the candidate sample and the reweighted  $ee$  sample. The  $ee$  distribution has been normalized to the  $E_T^{\text{miss}} < 50$  GeV region of the candidate sample. Contributions to  $ee$  from  $t\bar{t}$  and  $ZZ$  events have been subtracted.

## APPENDIX B

### MOVING TOWARD A 2D ANALYSIS

One signature of the GMSB models of interest in this analysis is the presence of two or more high energy jets in the final state. Using the hadronic activity as a handle to discriminate between signal and background would be a fruitful way to improve the sensitivity of the analysis. In this appendix, I will present a few of the 2D analysis methods that were explored in the 2016 data set, highlighting the challenges that we faced.

#### B.1 Bin in number of jets

The most straightforward way to include hadronic information would be to simply bin in jet multiplicity. This was done in the 8 TeV analysis, but it was dropped in the 2015 analysis because of the limited size of the data set ( $2.3 \text{ fb}^{-1}$ ). It was explored in the 2016 analysis, but we decided not to pursue it when we made the switch from  $ee$  to  $ff$  for our primary QCD control sample. We are already limited enough in  $ff$  statistics that binning further in the number of jets is not feasible.

Additionally, there are issues one must address when using a discrete variable such as the number of jets. Defining the jet ID in such a way that real photons are delineated from  $e/\gamma$  jet constituents is essential. We also found that cutting on the number of jets biased the di-EM  $p_T$  distribution in unexpected ways (Figure B.1).

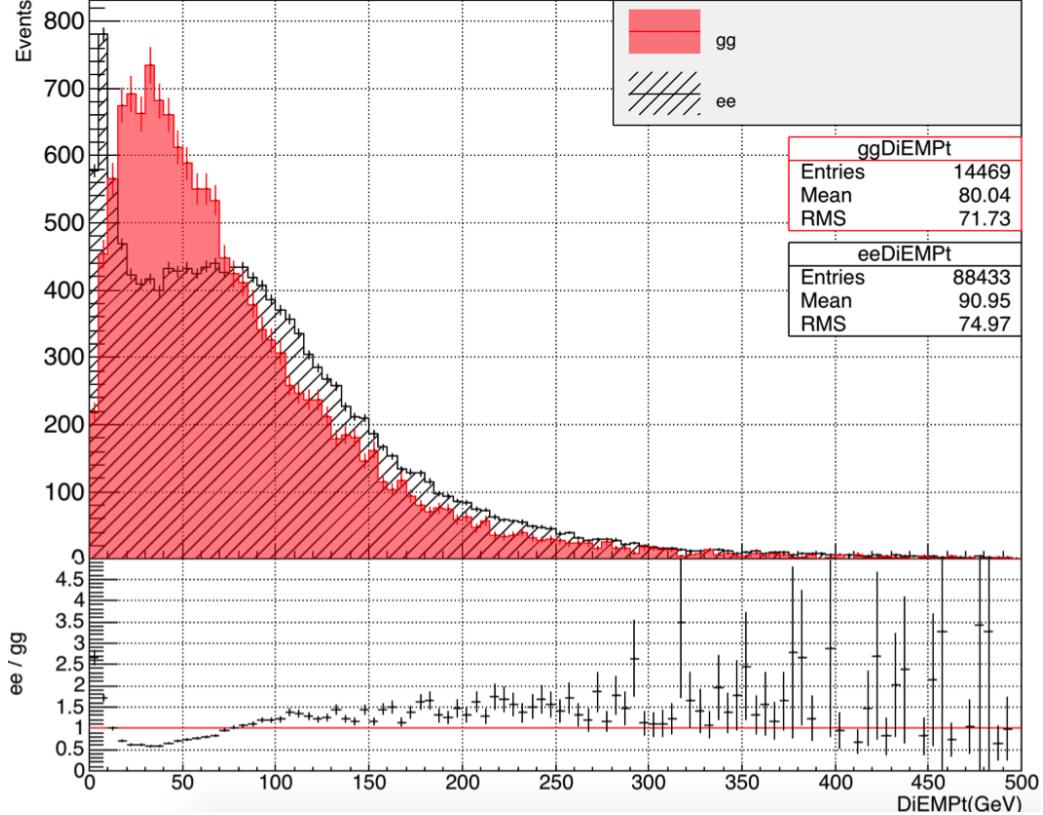


Figure B.1: The di-EM  $p_T$  distributions for  $\gamma\gamma$  (red) and  $ee$  (black) events with  $\geq 2$  jets. The data corresponds to the first  $12.9 \text{ fb}^{-1}$  collected with the CMS detector in 2016. The double peak structure only appears if one applies a cut on the number of jets.

## B.2 ABCD method

The method that we spent the most time exploring is referred to as the “ABCD” method. In this method, two uncorrelated variables are used to discriminate between signal and background. In our case, one variable would represent the amount of hadronic activity in the event (for example  $S_T$ , the scalar sum of the transverse momentum from all visible particles) and the other variable would be either  $E_T^{\text{miss}}$  or a related variable such as  $E_T^{\text{miss}}$  significance (defined as  $E_T^{\text{miss}}/\sqrt{S_T}$ ). In this method, the 2D plane is divided into four regions, as shown in Figure B.2, where Region  $D$  is the signal region and Regions  $A-C$  are control regions with minimal signal

contamination. If the two variables are uncorrelated, then we have the relation

$$\frac{D}{C} = \frac{B}{A} \quad (\text{B.1})$$

which can be rearranged to give a prediction for  $D$  in terms of the yields in the three control regions:

$$D_{exp} = \frac{BC}{A} \quad (\text{B.2})$$

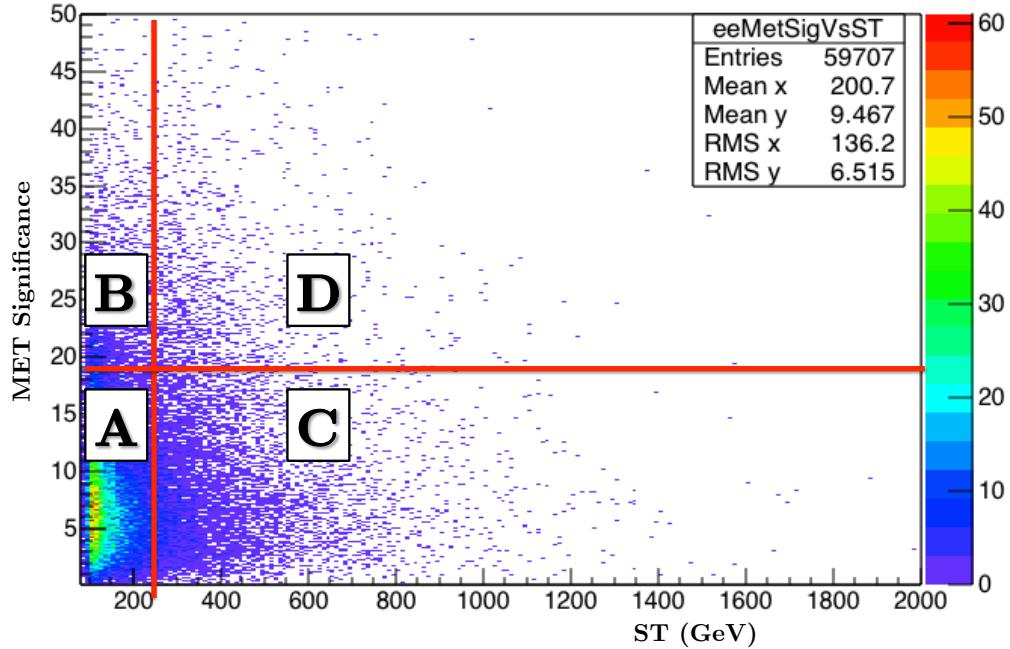


Figure B.2: Example of the ABCD analysis method using  $ee$  events in data. The two discriminating variables in this plot are the scalar sum of all visible energy in the event  $S_T$ , and  $E_T^{\text{miss}}$  significance, defined as  $E_T^{\text{miss}}/\sqrt{S_T}$ . If the two variables are uncorrelated, then  $D_{exp} = \frac{BC}{A}$ .

The problem that we faced was finding two variables that were sufficiently uncor-

related. A wide variety of variables were explored:

- Missing transverse energy  $E_T^{\text{miss}}$
- $S_T$ : scalar sum of the transverse momentum from all photons, leptons, and jets in the event. The sum sometimes includes  $E_T^{\text{miss}}$  as well.
- $H_T$ : scalar sum of the transverse momentum from all jets in the event.
- $E_T^{\text{miss}}$  significance: Defined as  $E_T^{\text{miss}}/\sqrt{S_T}$ . This variable quantifies the likelihood that the  $E_T^{\text{miss}}$  in an event is real missing energy rather than the result of hadronic mismeasurement.
- Missing  $H_T$ : the magnitude of the vector sum of all photons, leptons, and jets passing our ID criteria.
- MT2: sometimes referred to as the “stransverse mass”. It is calculated as a function of the two electromagnetic objects and the  $E_T^{\text{miss}}$ , and it has a cut-off at the mass of the parents of the invisible particles.

To quantify the correlation between two variables  $x$  and  $y$ , I looked at how the ratio  $N_{x>\alpha}/N_{x<\alpha}$  changes as a function of  $y$ , where  $\alpha$  is the boundary between regions  $AC$  and regions  $BD$ . The ABCD method relies on the assumption that this ratio is a constant. Similarly, one can consider  $N_{y>\beta}/N_{y<\beta}$  as a function of  $x$  for the boundary  $\beta$  between  $AB$  and  $CD$ . Figure B.3 shows these ratios for  $S_T$  and  $E_T^{\text{miss}}$  significance. Not only are the ratios not constant, but they exhibit a noticeable change in behavior at  $S_T = 400$  GeV and  $E_T^{\text{miss}}$  significance = 15. This makes it very difficult to extrapolate the behavior from the control region to the signal region. Other pairs of variables followed similar trends.

In the end, we decided that it was more important to publish the results and keep to a schedule that would allow us to be part of the GMSB combination effort. Efforts at a multivariate analysis were tabled until there is time to properly explore it.

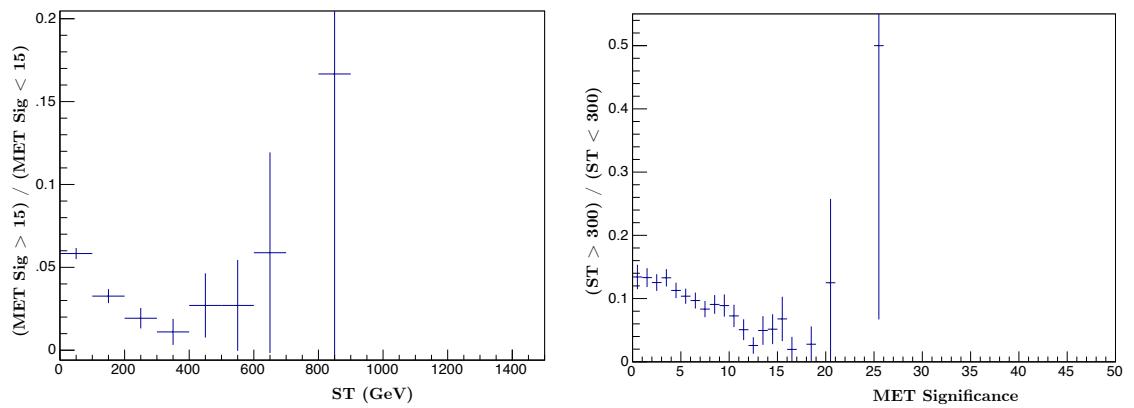


Figure B.3: Exploring the correlation between  $S_T$  and  $E_T^{\text{miss}}$  significance using the ratio method. If the variables are uncorrelated, the ratios should be constant. These ratios are for the  $ee$  control sample in data.

## BIBLIOGRAPHY

1. S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta. The fast simulation of the CMS detector at LHC. *J. Phys. Conf. Ser.*, 331:032049, 2011. doi:10.1088/1742-6596/331/3/032049.
2. S. Abel, M. J. Dolan, J. Jaeckel, and V. V. Khoze. Phenomenology of pure general gauge mediation. *JHEP*, 12:001, 2009. doi:10.1088/1126-6708/2009/12/001.
3. L. Álvarez-Gaumé and J. Ellis. Eyes on a prize particle. *Nature Physics*, 7:2–3, 2010. doi:10.1038/nphys1874.
4. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi:10.1007/JHEP07(2014)079.
5. K. Arun, S. Gudennavar, and C. Sivaram. Dark matter, dark energy, and alternate models: a review. *Advances in Space Research*, 60(1):166–186, 2017. doi:<https://doi.org/10.1016/j.asr.2017.03.043>.
6. ATLAS Collaboration. Search for supersymmetry in a final state containing two photons and missing transverse momentum in  $\sqrt{s} = 13$  TeV  $pp$  collisions at the LHC using the ATLAS detector. *Eur. Phys. J. C*, 76:517, 2016. doi:10.1140/epjc/s10052-016-4344-x.
7. ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phy. Lett. B*, 716(1):1–29, 2012. doi:<https://doi.org/10.1016/j.physletb.2012.08.020>.
8. ATLAS Collaboration. Search for photonic signatures of gauge-mediated supersymmetry in 8 TeV  $pp$  collisions with the ATLAS detector. *Phys. Rev. D*, 92:072001, 2015. doi:10.1103/PhysRevD.92.072001.
9. R. D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi:10.1007/JHEP04(2015)040.
10. W. Beenakker, S. Brensing, M. Krämer, A. Kulesza, E. Laenen, and I. Niessen. Soft-gluon resummation for squark and gluino hadroproduction. *JHEP*, 12:041, 2009. doi:10.1088/1126-6708/2009/12/041.

11. C. Borschensky, M. Krämer, 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. C*, 74:3174, 2014. doi:10.1140/epjc/s10052-014-3174-y.
12. M. Buican, P. Meade, N. Seiberg, and D. Shih. Exploring general gauge mediation. *JHEP*, 03:016, 2009. doi:10.1088/1126-6708/2009/03/016.
13. M. Cacciari, G. P. Salam, and G. Soyez. The anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008. doi:10.1088/1126-6708/2008/04/063.
14. CERN. CERN Document Server (CDS). URL <http://cds.cern.ch>.
15. CMS Collaboration. Search for supersymmetry in events with photons and missing transverse energy in  $pp$  collisions at 13 TeV. *Phys. Lett. B*, 769:391, 2017. doi:10.1016/j.physletb.2017.04.005.
16. CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716(1):30–61, 2012. doi:10.1016/j.physletb.2012.08.021.
17. CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *JINST*, 5(03):T03021, 2010. doi:10.1088/1748-0221/5/03/T03021.
18. CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi:10.1088/1748-0221/3/08/S08004.
19. CMS Collaboration. CMS-DP-2017/003. Twiki, 2017. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResultsCMSDPS2017003>.
20. CMS Collaboration. CMS-DP-2015/057. Twiki, 2015. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResultsCMSDPS2015057>.
21. CMS Collaboration. The CMS electromagnetic calorimeter project : Technical Design Report. CERN-LHCC-97-033, CMS-TDR-4, 1997. URL <http://cds.cern.ch/record/349375>.
22. CMS Collaboration. Changes to CMS ECAL electronics : addendum to the Technical Design Report. CERN-LHCC-2002-027, CMS-TDR-4-add-1, 2002. URL <http://cds.cern.ch/record/581342>.
23. CMS Collaboration. Jet energy scale and resolution in the CMS experiment in  $pp$  collisions at 8 TeV. *JINST*, 12:P02014, 2017. doi:10.1088/1748-0221/12/02/P02014.
24. CMS Collaboration. Search for supersymmetry with photons in  $pp$  collisions at  $\sqrt{s} = 8$  TeV. *Phys. Rev. D*, 92:072006, 2015. doi:10.1103/PhysRevD.92.072006.

25. CMS Collaboration. The CMS Level-1 trigger system for LHC Run II. *JINST*, 12(03):C03021, 2017. doi:10.1088/1748-0221/12/03/C03021.
26. CMS Collaboration. Level-1 trigger Phase I upgrade diagrams. Twiki, 2017. URL <https://twiki.cern.ch/twiki/bin/view/CMS/L1TriggerPhaseIDiagrams>.
27. CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12:P10003, 2017. doi:10.1088/1748-0221/12/10/P10003.
28. CMS Collaboration. Search for supersymmetry in events with at least one photon, missing transverse momentum, and large transverse event activity in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 12:142, 2017. doi:10.1007/JHEP12(2017)142.
29. CMS Collaboration. Search for gauge-mediated supersymmetry in events with at least one photon and missing transverse momentum in pp collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B.*, 780:118, 2018. doi:10.1016/j.physletb.2018.02.045.
30. CMS Collaboration. EGM ID recipes Run II. Twiki, 2018. URL [https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Photon\\_efficiencies\\_and\\_scale\\_fa](https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Photon_efficiencies_and_scale_fa).
31. CMS Collaboration. CMS supersymmetry physics results. Twiki, 2018. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
32. CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *JINST*, 10:P06005, 2015. doi:10.1088/1748-0221/10/06/P06005.
33. CMS Collaboration. CMS luminosity measurements for the 2016 data-taking period. CMS-PAS-LUM-17-001, 2017. URL <http://cds.cern.ch/record/2257069>.
34. CMS Collaboration. The CMS magnet project: Technical Design Report. CERN-LHCC-97-010, CMS-TDR-1, 1997. URL <https://cds.cern.ch/record/331056>.
35. CMS Collaboration. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *JINST*, 10:P08010, 2015. doi:10.1088/1748-0221/10/08/P08010.
36. CMS Collaboration. The CMS tracker system project: Technical Design Report. CERN-LHCC-98-006, CMS-TDR-5, 1997. URL <https://cds.cern.ch/record/368412>.
37. CMS Collaboration. The CMS tracker: addendum to the Technical Design Report. CERN-LHCC-2000-016, CMS-TDR-5-add-1, 2000. URL <https://cds.cern.ch/record/490194>.

38. M. Dine and A. E. Nelson. Dynamical supersymmetry breaking at low-energies. *Phys. Rev. D*, 48:1277, 1993. doi:10.1103/PhysRevD.48.1277.
39. M. Dine, A. E. Nelson, and Y. Shirman. Low-energy dynamical supersymmetry breaking simplified. *Phys. Rev. D*, 51:1362, 1995. doi:10.1103/PhysRevD.51.1362.
40. M. Dine, A. E. Nelson, Y. Nir, and Y. Shirman. New tools for low-energy dynamical supersymmetry breaking. *Phys. Rev. D*, 53:2658, 1996. doi:10.1103/PhysRevD.53.2658.
41. N. Dragon, U. Ellwanger, and M. G. Schmidt. Supersymmetry and supergravity. *Prog. Part. Nucl. Phys.*, 18:1–91, 1987. doi:[https://doi.org/10.1016/0146-6410\(87\)90008-1](https://doi.org/10.1016/0146-6410(87)90008-1).
42. F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. doi:10.1103/PhysRevLett.13.321.
43. L. Evans and P. Bryant. LHC machine. *JINST*, 3(08):S08001, 2008. doi:10.1088/1748-0221/3/08/S08001.
44. C. W. Fabjan and F. Gianotti. Calorimetry for particle physics. *Rev. Mod. Phys.*, 75:1243–1286, 2003. doi:10.1103/RevModPhys.75.1243.
45. P. Fayet. Mixing between gravitational and weak interactions through the massive gravitino. *Phys. Lett. B*, 70:461, 1977. doi:10.1016/0370-2693(77)90414-2.
46. P. Fayet. Scattering cross-sections of the photino and the goldstino (gravitino) on matter. *Phys. Lett. B*, 86:272, 1979. doi:10.1016/0370-2693(79)90836-0.
47. P. Fayet. Lower limit on the mass of a light gravitino from  $e^+ e^-$  annihilation experiments. *Phys. Lett. B*, 175:471, 1986. doi:10.1016/0370-2693(86)90626-X.
48. R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi:10.1016/0168-9002(87)90887-4.
49. S. L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22(4):579–588, 1961. doi:[https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
50. S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D.*, 2:1285–1292, 1970. doi:10.1103/PhysRevD.2.1285.
51. J. Goldstone, A. Salam, and S. Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, 1962. doi:10.1103/PhysRev.127.965.
52. G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, 1964. doi:10.1103/PhysRevLett.13.585.
53. P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, 1964. doi:10.1103/PhysRevLett.13.508.

54. Y. Iiyama. *Search for supersymmetry in pp collisions at  $\sqrt{s} = 8$  TeV with a photon, lepton, and missing transverse energy*. PhD thesis, Carnegie Mellon University, 2015. URL <http://inspirehep.net/record/1429554>.
55. T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999. doi:10.1016/S0168-9002(99)00498-2.
56. K. Kiesel, C. Autermann, and L. Feld. Search for supersymmetry in events with at least one photon, missing transverse momentum and large event activity with data recorded at the CMS detector in 2016. CMS Analysis Note, AN2016-470, 2016. URL <http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2016/470>.
57. A. Kulesza and L. Motyka. Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC. *Phys. Rev. D*, 80:095004, 2009. doi:10.1103/PhysRevD.80.095004.
58. S. P. Martin. A supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 21: 1–153, 2010. doi:10.1142/9789812839657\_0001.
59. P. Meade, N. Seiberg, and D. Shih. General gauge mediation. *Prog. Theor. Phys. Suppl.*, 177:143, 2009. doi:10.1143/PTPS.177.143.
60. C. Patrignani et al. Particle Data Group. Review of particle physics. *Chin. Phys. C.*, 40(10), 2016 and 2017 update. doi:10.1088/1674-1137/40/10/100001.
61. A. L. Read. Presentation of search results: the  $CL_s$  technique. *J. Phys. G*, 28: 2693, 2002. doi:10.1088/0954-3899/28/10/313.
62. T. Sjöstrand, S. Mrenna, and P. Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008. doi:10.1016/j.cpc.2008.01.036.
63. SNO Collaboration. Measurement of the rate of  $\nu_e + d \rightarrow p + p + e^-$  interactions produced by  $^8b$  solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001. doi:10.1103/PhysRevLett.87.071301.
64. SNO Collaboration. Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002. doi:10.1103/PhysRevLett.89.011301.
65. Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998. doi:10.1103/PhysRevLett.81.1562.
66. M. Terwort. ATLAS and CMS Collaborations. Searches for GMSB at the LHC. ATL-PHYS-CONF-2008-017, ATL-COM-PHYS-2008-075, 2008.
67. S. Van der Meer. Calibration of the effective beam height in the ISR. CERN-ISR-PO-68-31, 1968. URL <https://cds.cern.ch/record/296752>.

68. S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi:10.1103/PhysRevLett.19.1264.
69. J. Wittmann, B. Arnold, H. Bergauer, M. Jeitler, T. Matsushita, D. Rabady, B. Rahbaran, and C.-E. Wulz. The upgrade of the CMS Global Trigger. *JINST*, 11(02):C02029, 2016. doi:10.1088/1748-0221/11/02/C02029.

*This document was prepared & typeset with pdfL<sup>A</sup>T<sub>E</sub>X, and formatted with NDDiss2<sub>ε</sub> classfile (v3.2013[2013/04/16]) provided by Sameer Vijay and updated by Megan Patnott.*