

A Search in the Two-Photon Final State for  
Evidence of New Particle Production in pp  
Collisions at  $\sqrt{s} = 7$  TeV

Rachel P. Yohay

University of Virginia

rpy3y@virginia.edu

April 18, 2012

# Contents

<b>1 Overview of the Standard Model of Particle Physics</b>	<b>3</b>
1.1 Particle Content . . . . .	6
1.2 Electroweak Symmetry Breaking and the Higgs Mechanism . . . . .	6
1.3 The Hierarchy Problem, The Origins of Mass, and Fine Tuning . . . . .	6
<b>2 The Supersymmetric Extension to the Standard Model</b>	<b>7</b>
2.1 Supermultiplet Representation . . . . .	7
2.2 The Unbroken SUSY Lagrangian . . . . .	8
2.3 Soft SUSY Breaking . . . . .	14
2.4 Gauge-Mediated SUSY Breaking . . . . .	16
2.5 Phenomenology of General Gauge Mediation . . . . .	19
2.6 Experimental Status of SUSY . . . . .	23
<b>3 Event Selection</b>	<b>29</b>
3.1 Object Reconstruction . . . . .	30
3.1.1 Photons . . . . .	30
3.1.2 Electrons . . . . .	39
3.1.3 Jets and Missing Transverse Energy . . . . .	39
3.2 HLT . . . . .	44
3.3 Event Quality . . . . .	48
3.4 Photon Identification Efficiency . . . . .	50

3.4.1	Tag and Probe Method . . . . .	52
3.4.2	Photon Efficiency Scale Factor $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$ . . . . .	55
<b>4</b>	<b>Data Analysis</b>	<b>61</b>
4.1	Modeling the QCD Background . . . . .	64
4.1.1	Outline of the Procedure . . . . .	64
4.1.2	Reweighting . . . . .	67
4.1.3	Normalization . . . . .	77
4.2	Modeling the Electroweak Background . . . . .	77
4.3	Errors on the Background Prediction . . . . .	80
4.4	Results . . . . .	86
<b>A</b>	<b>Monte Carlo Samples</b>	<b>90</b>
A.0.1	List of Samples . . . . .	90
A.0.2	Explanation of Naming Conventions . . . . .	91

# Chapter 1

## Overview of the Standard Model of Particle Physics

In the 1960s, Sheldon Glashow, Steven Weinberg, and Abdus Salam proposed a mathematical framework that unified the electromagnetic and weak forces at an energy scale in the hundreds of GeV/c, as well as a mechanism for breaking the electroweak symmetry at low energies [1]. At the same time, Murray Gell-Mann introduced the concept of quarks to describe hadron spectroscopy, a concept that would later grow into quantum chromodynamics (QCD), the full theory of the strong force [2]. These two key developments motivated the unified representation of particle physics as a set of fields whose dynamics are invariant under the Standard Model gauge group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_{EM} \tag{1.1}$$

where  $SU(3)_C$  describes the quark QCD interactions,  $SU(2)_L$  describes the weak interactions among quarks and leptons, and  $U(1)_{EM}$  describes the electromagnetic interaction.

The Standard Model, in particular the electroweak theory, has been an extremely successful predictor of particle production and interaction cross-sections and decay

rates, as well as of the exact masses of the electroweak force carriers. The case for the validity of the Standard Model was bolstered by the many precision QCD and electroweak measurements carried out at the Large Electron-Positron (LEP) collider, which ran from 1989-2000 at center-of-mass energies between 65 and 104 GeV/c [3]. Figure 1.1 shows some of the highlights of the LEP program.

However, there are still deep theoretical problems with the Standard Model, stemming from the introduction of the Higgs scalar into the theory to break electroweak symmetry [4]. Since the Higgs self-energy diagram is quadratically sensitive to the ultraviolet cutoff scale(footnote: this is a general property of scalar fields), and assuming that there are no new important energy scales of physics between the weak scale ( $\mathcal{O}(10^2 \text{ GeV}/c)$ ) and the Planck scale ( $\mathcal{O}(10^{19} \text{ GeV}/c)$ ), in order to be consistent with experimental measurements, this diagram must include a remarkable 17-orders-of-magnitude cancellation that is otherwise poorly motivated [5]. The quest to find new physics at an intermediate energy scale between the weak and Planck scales, and thus extend the Standard Model, was the driving force behind the construction of the Large Hadron Collider (LHC) in 2009, the world's highest energy particle accelerator to date.

In this chapter I will briefly describe the Standard Model particle content, the theory and major results of electroweak symmetry breaking (EWSB), and the problems that the Standard Model is as yet ill-prepared to address.

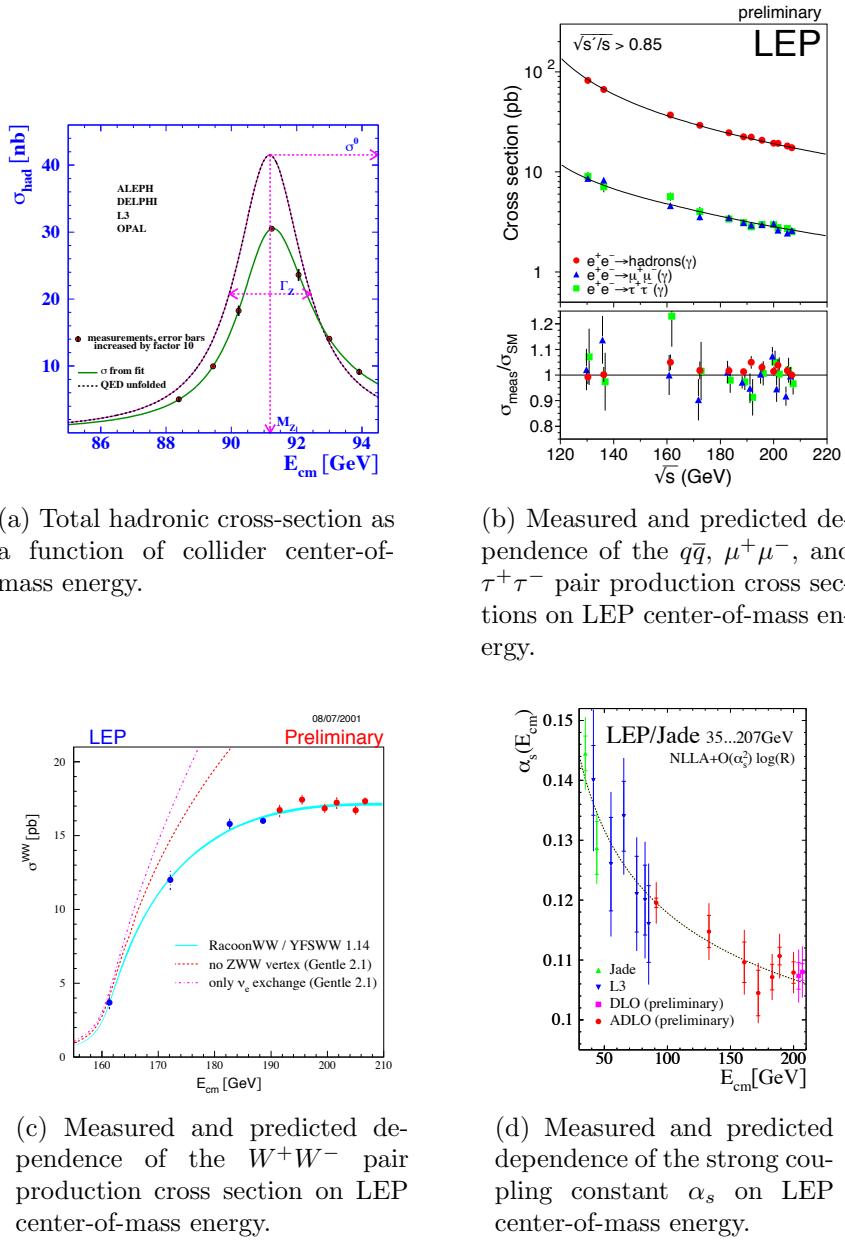


Figure 1.1: Selected LEP measurements demonstrating its contribution to the precise understanding of the Standard Model. Reprinted from [3].

## 1.1 Particle Content

## 1.2 Electroweak Symmetry Breaking and the Higgs Mechanism

## 1.3 The Hierarchy Problem, The Origins of Mass, and Fine Tuning

# Chapter 2

## The Supersymmetric Extension to the Standard Model

The following introduction to SUSY focuses primarily on the aspects of the formalism that are relevant to phenomenology. In particular, most of the details of SUSY breaking (about which there is little theoretical consensus) are omitted, except where they are relevant to experiment. The notation is similar to that used in refs. [5] and [6], and much of the information presented is culled from those references.

### 2.1 Supermultiplet Representation

The Standard Model is extended to include supersymmetry by the introduction of a supersymmetry transformation that takes fermionic states to bosonic states and vice versa. The resulting model is called the *minimal supersymmetric Standard Model* (MSSM). In analogy with the known symmetries of the Standard Model, the SUSY transformation has associated generators that obey defining commutation and anti-commutation relations, and a fundamental representation. All SM particles and their *superpartners* fall into one of two *supermultiplet* representations. Using the property that

$$n_F = n_B, \quad (2.1)$$

where  $n_F$  is the number of fermionic degrees of freedom per supermultiplet and  $n_B$  is the number of bosonic degrees of freedom, the two types of supermultiplets are

1. *Chiral supermultiplets*: one Weyl fermion (two helicity states  $\Rightarrow n_F = 2$ ) and one complex scalar field (with two real components  $\Rightarrow n_B = 2$ )
2. *Gauge supermultiplets*: One spin-1 vector boson (two helicity states  $\Rightarrow n_B = 2$ ) and one Weyl fermion (two helicity states  $\Rightarrow n_F = 2$ )

In the gauge supermultiplet, the vector boson is assumed massless (i.e. before EWSB generates a mass for it). Since the superpartners to the SM particles have not yet been discovered, they must be significantly heavier than their SM counterparts. Unbroken SUSY predicts that the SM particles and their superpartners must have exactly the same mass, so ultimately a mechanism for SUSY breaking must be introduced to generate masses for the superpartners (see Sec. 2.3). Tables 2.1 and 2.2 show the chiral and gauge supermultiplets of the MSSM, respectively. Note that the scalar partners to the SM fermions are denoted by placing an “s” in front of their names, while the chiral fermion partners to the SM gauge bosons are denoted by appending “ino” to their names.

## 2.2 The Unbroken SUSY Lagrangian

The first piece of the full unbroken SUSY Lagrangian density consists of the kinetic and interacting terms related to the chiral supermultiplets. As explained in Sec. 2.1, a chiral supermultiplet consists of a Weyl fermion  $\psi$  (the ordinary fermion) and a complex scalar  $\phi$  (the sfermion). For a collection of such chiral supermultiplets, the Lagrangian is

Table 2.1: Chiral supermultiplets of the supersymmetric Standard Model. Adapted from Table 1.1 of ref. [6].

Type of supermultiplet	Notation	Spin-0 component	Spin-1/2 component	Representation under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$
Left-handed quark/squark doublet ( $\times 3$ families)	$Q$	$(\tilde{u}_L \tilde{d}_L)$	$(u_L d_L)$	$(\mathbf{3}, \mathbf{2}, \frac{1}{6})$
Right-handed up-type quark/squark singlet ( $\times 3$ families)	$\bar{u}$	$\tilde{u}_R^*$	$u_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, -\frac{2}{3})$
Right-handed down-type quark/squark singlet ( $\times 3$ families)	$\bar{d}$	$\tilde{d}_R^*$	$d_R^\dagger$	$(\bar{\mathbf{3}}, \mathbf{1}, \frac{1}{3})$
Left-handed lepton/slepton doublet ( $\times 3$ families)	$L$	$(\tilde{\nu}_{eL} \tilde{e}_L)$	$(\bar{\nu}_{eL} e_L)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$
Right-handed lepton/slepton singlet ( $\times 3$ families)	$\bar{e}$	$\tilde{e}_R^*$	$e_R^\dagger$	$(\bar{\mathbf{1}}, \mathbf{1}, 1)$
Up-type Higgs/Higgsino doublet	$H_u$	$(H_u^+ H_u^0)$	$(\tilde{H}_u^+ \tilde{H}_u^0)$	$(\mathbf{1}, \mathbf{2}, \frac{1}{2})$
Down-type Higgs/Higgsino doublet	$H_d$	$(H_d^0 H_d^-)$	$(\tilde{H}_d^0 \tilde{H}_d^-)$	$(\mathbf{1}, \mathbf{2}, -\frac{1}{2})$

Table 2.2: Gauge supermultiplets of the supersymmetric Standard Model. Adapted from Table 1.2 of ref. [6].

Type of supermultiplet	Spin-1/2 component	Spin-1 component	Representation under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$
Gluon/gluino	$\tilde{g}$	$g$	$(\mathbf{8}, \mathbf{1}, 0)$
W/wino	$\tilde{W}^\pm \tilde{W}^0$	$W^\pm W^0$	$(\mathbf{1}, \mathbf{3}, 0)$
B/bino	$\tilde{B}^0$	$B^0$	$(\mathbf{1}, \mathbf{1}, 0)$

$$\begin{aligned}\mathcal{L}_{\text{chiral}} = & -\partial^\mu \phi^{*i} \partial_\mu \phi_i - V_{\text{chiral}}(\phi, \phi^*) - i\psi^{\dagger i} \bar{\sigma}^\mu \partial_\mu \psi_i - \frac{1}{2} M^{ij} \psi_i \psi_j \\ & - \frac{1}{2} M_{ij}^* \psi^{\dagger i} \psi^{\dagger j} - \frac{1}{2} y^{ijk} \phi_i \psi_j \psi_k - \frac{1}{2} y_{ijk}^* \phi^{*i} \psi^{\dagger j} \psi^{\dagger k}\end{aligned}\quad (2.2)$$

where  $i$  runs over all supermultiplets in Table 2.1,  $\bar{\sigma}^\mu$  are  $-1 \times$  the Pauli matrices (except for  $\sigma^0 = \bar{\sigma}^0$ ),  $M^{ij}$  is a mass matrix for the fermions,  $y^{ijk}$  are the Yukawa couplings between one scalar and two spinor fields, and  $V_{\text{chiral}}(\phi, \phi^*)$  is the scalar potential

$$\begin{aligned}V_{\text{chiral}}(\phi, \phi^*) = & M_{ik}^* M^{kj} \phi^{*i} \phi_j + \frac{1}{2} M^{in} y_{jkn}^* \phi_i \phi^{*j} \phi^{*k} \\ & + \frac{1}{2} M_{in}^* y^{jkn} \phi^{*i} \phi_j \phi_k + \frac{1}{4} y^{ijn} y_{klm}^* \phi_i \phi_j \phi^{*k} \phi^{*l}.\end{aligned}\quad (2.3)$$

The Lagrangian can also be written as the kinetic terms plus derivatives of the *superpotential*  $W$ :

$$\begin{aligned}\mathcal{L}_{\text{chiral}} = & -\partial^\mu \phi^{*i} \partial_\mu \phi_i - i\psi^{\dagger i} \bar{\sigma}^\mu \partial_\mu \psi_i \\ & - \frac{1}{2} \left( \frac{\delta^2 W}{\delta \phi^i \delta \phi^j} \psi_i \psi_j + \frac{\delta^2 W^*}{\delta \phi_i \delta \phi_j} \psi^{\dagger i} \psi^{\dagger j} \right) - \frac{\delta W}{\delta \phi^i} \frac{\delta W^*}{\delta \phi_i}\end{aligned}\quad (2.4)$$

where

$$W = M^{ij} \phi_i \phi_j + \frac{1}{6} y^{ijk} \phi_i \phi_j \phi_k. \quad (2.5)$$

The second part of the Lagrangian involves the gauge supermultiplets. In terms of the spin-1 ordinary gauge boson  $A_\mu^a$  and the spin-1/2 Weyl spinor gaugino  $\lambda^a$  of

the gauge supermultiplet, where  $a$  runs over the number of generators for the SM subgroup (i.e. 1-8 for  $SU(3)_C$ , 1-3 for  $SU(2)_L$ , and 1 for  $U(1)_Y$ ), this part of the Lagrangian is

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}F_{\mu\nu}^a F^{\mu\nu a} - i\lambda^{\dagger a}\bar{\sigma}^\mu D_\mu \lambda^a + \frac{1}{2}D^a D^a \quad (2.6)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c \quad (2.7)$$

( $g$  is the coupling constant and  $f^{abc}$  are the structure constants for the particular SM gauge group),

$$D_\mu \lambda^a = \partial_\mu \lambda^a + g f^{abc} A_\mu^b \lambda^c, \quad (2.8)$$

and  $D^a$  is an auxiliary field that does not propagate (in the literature, it is used as a bookkeeping tool and can be removed via its algebraic equation of motion).

To build a fully supersymmetric and gauge-invariant Lagrangian, the ordinary derivatives in  $\mathcal{L}_{\text{chiral}}$  (Eq. 2.2) must be replaced by covariant derivatives

$$D_\mu \phi_i = \partial_\mu \phi_i - ig A_\mu^a (T^a \phi)_i \quad (2.9)$$

$$D_\mu \phi^{*i} = \partial_\mu \phi^{*i} + ig A_\mu^a (\phi^* T^a)^i \quad (2.10)$$

$$D_\mu \psi_i = \partial_\mu \psi_i - ig A_\mu^a (T^a \psi)_i. \quad (2.11)$$

This leads to the full Lagrangian

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_{\text{chiral}} + \mathcal{L}_{\text{gauge}} \\
&\quad - \sqrt{2}g(\phi^{*i}T^a\psi_i)\lambda^a - \sqrt{2}g\lambda^{\dagger a}(\psi^{\dagger i}T^a\phi_i) + g(\phi^{*i}T^a\phi_i)D^a \\
&= -\partial^\mu\phi^{*i}\partial_\mu\phi_i - i\psi^{\dagger i}\bar{\sigma}^\mu\partial_\mu\psi_i + ig\partial^\mu\phi^{*i}A_\mu^a(T^a\phi)_i - ig\partial_\mu\phi_iA^{\mu a}(\phi^*T^a)^i \\
&\quad - g^2A^{\mu a}(\phi^*T^a)^iA_\mu^a(T^a\phi)_i - g\psi^{\dagger i}\bar{\sigma}^\mu A_\mu^a(T^a\psi)_i - V_{\text{chiral}}(\phi, \phi^*) \\
&\quad - \frac{1}{2}M^{ij}\psi_i\psi_j - \frac{1}{2}M_{ij}^*\psi^{\dagger i}\psi^{\dagger j} - \frac{1}{2}y^{ijk}\phi_i\psi_j\psi_k - \frac{1}{2}y_{ijk}^*\phi^{*i}\psi^{\dagger j}\psi^{\dagger k} \\
&\quad - \frac{1}{4}F_{\mu\nu}^aF^{\mu\nu a} - i\lambda^{\dagger a}\bar{\sigma}^\mu\partial_\mu\lambda^a - ig\lambda^{\dagger a}\bar{\sigma}^\mu f^{abc}A_\mu^b\lambda^c + \frac{1}{2}D^aD^a \\
&\quad - \sqrt{2}g(\phi^{*i}T^a\psi_i)\lambda^a - \sqrt{2}g\lambda^{\dagger a}(\psi^{\dagger i}T^a\phi_i) + g(\phi^{*i}T^a\phi_i)D^a. \tag{2.12}
\end{aligned}$$

Writing out  $F_{\mu\nu}^a$  and  $V_{\text{chiral}}(\phi, \phi^*)$  explicitly combining the  $D^a$  terms using the equation of motion  $D^a = -g\phi^{*i}T^a\phi_i$ , and rearranging some terms, the final unbroken SUSY Lagrangian is

$$\begin{aligned}
\mathcal{L} = & -\partial^\mu \phi^{*i} \partial_\mu \phi_i - i\psi^{\dagger i} \bar{\sigma}^\mu \partial_\mu \psi_i \\
& - \frac{1}{4} (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a) (\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a}) - i\lambda^{\dagger a} \bar{\sigma}^\mu \partial_\mu \lambda^a \\
& - M_{ik}^* M^{kj} \phi^{*i} \phi_j - \frac{1}{2} M^{ij} \psi_i \psi_j - \frac{1}{2} M_{ij}^* \psi^{\dagger i} \psi^{\dagger j} \\
& + ig \partial^\mu \phi^{*i} A_\mu^a (T^a \phi)_i - ig \partial_\mu \phi_i A^{\mu a} (\phi^* T^a)^i - g \psi^{\dagger i} \bar{\sigma}^\mu A_\mu^a (T^a \psi)_i \\
& - ig \lambda^{\dagger a} \bar{\sigma}^\mu f^{abc} A_\mu^b \lambda^c \\
& - \frac{1}{4} g f^{abc} [(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a) A^{\mu b} A^{\nu c} + A_\mu^b A_\nu^c (\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a})] \\
& - \frac{1}{2} M_{in}^* y^{jkn} \phi^{*i} \phi_j \phi_k - \frac{1}{2} M^{in} y_{jkn}^* \phi_i \phi^{*j} \phi^{*k} \\
& - \frac{1}{2} y^{ijk} \phi_i \psi_j \psi_k - \frac{1}{2} y_{ijk}^* \phi^{*i} \psi^{\dagger j} \psi^{\dagger k} \\
& - \sqrt{2} g (\phi^{*i} T^a \psi_i) \lambda^a - \sqrt{2} g \lambda^{\dagger a} (\psi^{\dagger i} T^a \phi_i) \\
& - g^2 A^{\mu a} (\phi^* T^a)^i A_\mu^a (T^a \phi)_i - \frac{1}{4} g^2 f^{abc} A_\mu^b A_\nu^c f^{abc} A^{\mu b} A^{\nu c} \\
& - \frac{1}{4} y^{ijn} y_{kln}^* \phi_i \phi_j \phi^{*k} \phi^{*l} - \frac{1}{2} g^2 (\phi^{*i} T^a \phi_i)^2. \tag{2.13}
\end{aligned}$$

The above Lagrangian applies to chiral supermultiplets interacting with one kind of gauge supermultiplet (i.e. one SM gauge group). In the general case, there are additional terms corresponding to interactions with all three SM gauge groups.

The following list gives a description of the terms in Eq. 2.13:

- First two lines: kinetic terms for the four types of fields  $\phi_i$ ,  $\psi_i$ ,  $A_\mu^a$ , and  $\lambda^a$
- Third line: mass terms for the  $\phi_i$  and  $\psi_i$  (see Figs. 2.1(a) and 2.1(b))
- Fourth and fifth lines: cubic couplings in which  $\phi_i$ ,  $\psi_i$ , or  $\lambda^a$  radiates an  $A_\mu^a$  (see Figs. 2.1(c), 2.1(d), and 2.1(e))
- Sixth line: triple gauge boson couplings (see Fig. 2.1(f))
- Seventh line: triple sfermion couplings (see Fig. 2.1(g))

- Eighth line: cubic couplings in which  $\psi_i$  radiates a  $\phi_i$  (see Fig. 2.1(h))
- Ninth line:  $\phi_i$ - $\psi_i$ - $\lambda^a$  vertices (see Fig. 2.1(i))
- 10<sup>th</sup> line:  $A_\mu^a$ - $A_\mu^a$ - $\phi_i$ - $\phi_i$  and quadruple gauge boson couplings (see Figs. 2.1(j) and 2.1(k))
- 11<sup>th</sup> line:  $\phi_i^4$  vertices (see Figs. 2.1(l) and 2.1(m))

## 2.3 Soft SUSY Breaking

Since quadratic divergences in sfermion masses vanish to all orders in perturbation theory in plain unbroken SUSY [6] due to the presence of gauge and Yukawa interactions with the necessary relationships between coupling constants, it is desirable that the terms that break SUSY not disturb this property. In addition, SUSY should be broken spontaneously, as electroweak symmetry is broken in the Standard Model, so that it is only made manifest at high energies. To satisfy these constraints, SUSY-breaking terms are simply added to the unbroken SUSY Lagrangian in Eq. 2.13 such that  $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{unbroken}} + \mathcal{L}_{\text{breaking}}$ . The coefficients of terms in  $\mathcal{L}_{\text{breaking}}$  must have positive mass dimension in order not to contribute quadratically divergent loop corrections to the scalar masses (like the Higgs mass).<sup>1</sup> This form of SUSY breaking is called *soft*, and all coefficients of soft SUSY breaking terms are expected to be of order  $m_{\text{soft}}$  or  $m_{\text{soft}}^2$ .

---

<sup>1</sup>This point can be argued via dimensional analysis. Radiative corrections take the form  $\Delta m_S^2$ , where  $m_S$  is the mass of the scalar particle in question. The dimensions of  $\Delta m_S^2$  are mass<sup>2</sup>.  $\Delta m_S^2$  is proportional to some coupling constant or mass coefficient  $k$  multiplied by a function of  $\Lambda_{\text{UV}}$ , the high energy cutoff scale. The function of  $\Lambda_{\text{UV}}$  is determined by a loop integral, and thus typically takes the form  $\Lambda_{\text{UV}}^2$  (quadratically divergent) or  $\ln \frac{\Lambda_{\text{UV}}}{m_{\text{low}}}$  (logarithmically divergent, where  $m_{\text{low}}$  is some other lower-mass scale in the problem). Now, if  $k$  already contributes at least one power of mass to  $\Delta m_S^2$ , then the high-energy behavior—the function of  $\Lambda_{\text{UV}}$ —can only contribute at most one power of the dimensionful parameter  $\Lambda_{\text{UV}}$ . However, there are typically no loop integrals that diverge linearly in  $\Lambda_{\text{UV}}$ , so by forcing  $k$  to have positive mass dimension, the form of the radiative corrections contributed by SUSY-breaking terms is limited to  $\Delta m_S^2 \sim m_{\text{low}}^2 \ln \frac{\Lambda_{\text{UV}}}{m_{\text{low}}}$ . In effect, the possibility of dangerous corrections proportional to  $\Lambda_{\text{UV}}^2$  is excluded by dimensional analysis if the requirement that  $k$  contribute at least one power of mass is enforced.

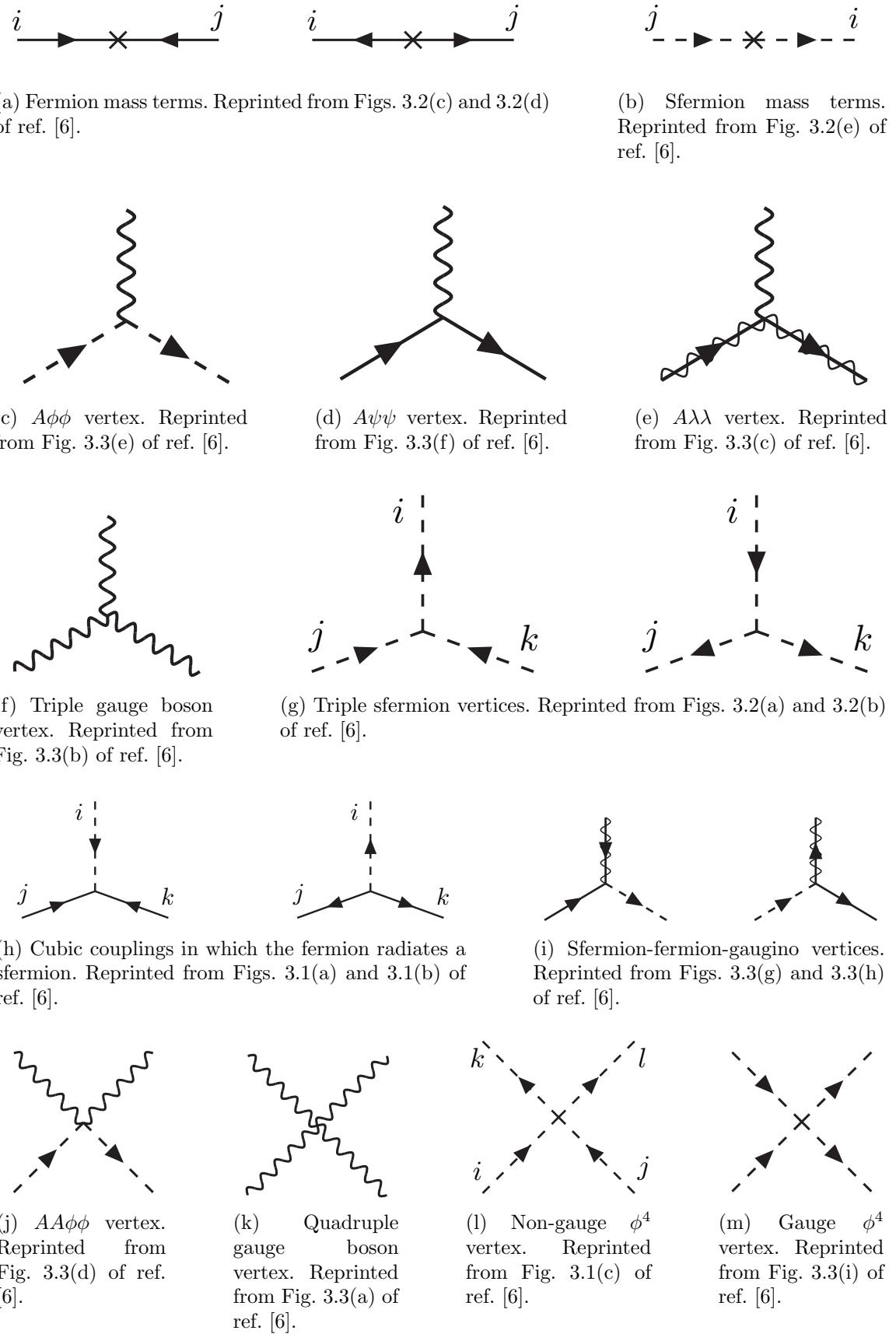


Figure 2.1: Interactions in the unbroken SUSY Lagrangian.

Soft SUSY breaking terms give masses to the sfermions and gauginos and introduce a cubic sfermion vertex. The soft terms are given by

$$\begin{aligned}
\mathcal{L}_{\text{soft}} = & -\frac{1}{2}(M_3 \tilde{g}^a \tilde{g}^a + M_2 \tilde{W}^a \tilde{W}^a + M_1 \tilde{B} \tilde{B} + \text{h.c.}) \\
& - (a_u^{ij} \tilde{u}_{Ri}^* \tilde{Q}_j H_u - a_d^{ij} \tilde{d}_{Ri}^* \tilde{Q}_j H_d - a_e^{ij} \tilde{e}_{Ri}^* \tilde{L}_j H_d + \text{h.c.}) \\
& - m_{\tilde{Q}_{ij}}^2 \tilde{Q}_i^\dagger \tilde{Q}_j - m_{\tilde{L}_{ij}}^2 \tilde{L}_i^\dagger \tilde{L}_j \\
& - m_{\tilde{u}_{ij}}^2 \tilde{u}_{Ri} \tilde{u}_{Rj}^* - m_{\tilde{d}_{ij}}^2 \tilde{d}_{Ri} \tilde{d}_{Rj}^* - m_{\tilde{e}_{ij}}^2 \tilde{e}_{Ri} \tilde{e}_{Rj}^* \\
& - m_{H_u}^2 H_u^* H_u - m_{H_d}^2 H_d^* H_d - (b H_u H_d + \text{h.c.})
\end{aligned} \tag{2.14}$$

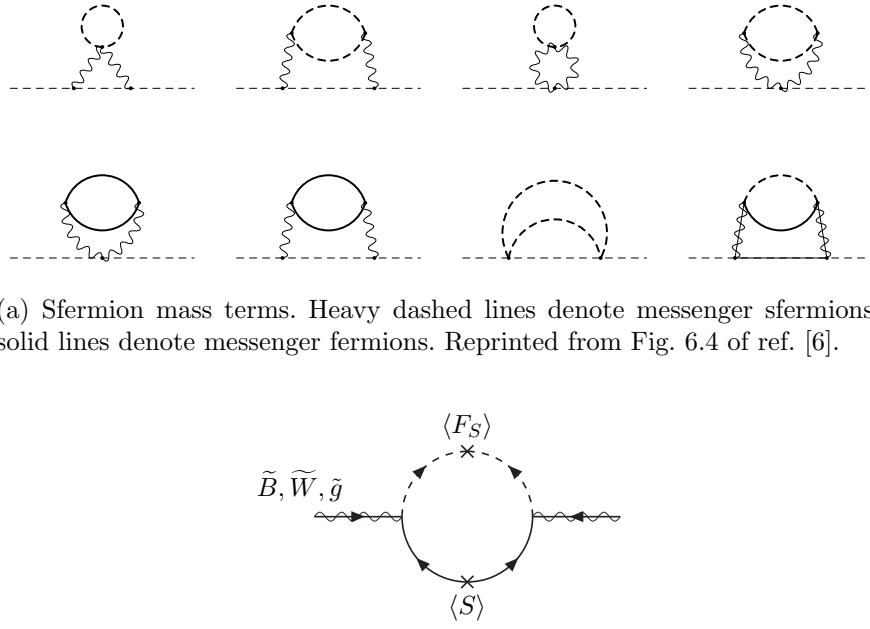
where  $a$  runs from 1-8 for  $\tilde{g}^a$  and from 1-3 for  $\tilde{W}^a$ , and  $i, j$  run over the three families.

The color indices are not shown. The first line of Eq. 2.14 contains the gaugino mass terms. The second line contains cubic scalar couplings that contribute to mixing between the left- and right-handed third generation sfermions (it is assumed in the supersymmetric Standard Model that the  $a_u^{ij}$ ,  $a_d^{ij}$ , and  $a_e^{ij}$  are negligible unless  $i = j = 3$ ). The third and fourth lines of Eq. 2.14 contain squark and slepton mass terms, and finally the last line contains the Higgs mass terms.

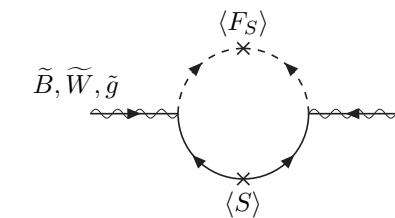
Many viable models of achieving soft SUSY breaking have been studied over the last 30 years. For an overview, see Sec. 6 of ref. [6]. However, this thesis will only cover *gauge-mediated SUSY breaking* (GMSB), because the two-photon search performed is far more sensitive to this model than to other models of SUSY breaking.

## 2.4 Gauge-Mediated SUSY Breaking

In gauge-mediated models [7], “hidden” fields spontaneously break the supersymmetry of very heavy chiral *messenger* supermultiplets. There are a number of competing models (see ref. [7]) that explain the precise mechanism of spontaneous SUSY



(a) Sfermion mass terms. Heavy dashed lines denote messenger sfermions; solid lines denote messenger fermions. Reprinted from Fig. 6.4 of ref. [6].



(b) Gaugino mass term. The  $\langle S \rangle$  part of the loop is a messenger fermion contribution; the  $\langle F_S \rangle$  part is a messenger sfermion contribution. Reprinted from Fig. 6.3 of ref. [6].

Figure 2.2: Contributions to sfermion and gaugino masses from loop interactions with messenger particles in the GMSB framework.

breaking, but fortunately the details of those models mostly decouple from the phenomenology of GMSB. The messengers then communicate the SUSY breaking to the sparticles via loop diagrams of gauge interaction strength (i.e. via vertices like those shown in Figs. 2.1(c), 2.1(d), 2.1(i), 2.1(j), and 2.1(m), which are proportional to the SM gauge coupling constants). Feynman diagrams corresponding to gaugino and sfermion mass terms are shown in Figure 2.2.

Historically, GMSB and gravity-mediated SUSY breaking, or mSUGRA [8], have been the two most thoroughly experimentally studied scenarios of SUSY breaking. One advantage of GMSB over mSUGRA is that it naturally suppresses flavor violation, a generic prediction of supersymmetry. Flavor violation is introduced in the scalar<sup>3</sup> couplings and sfermion mass terms of  $\mathcal{L}_{\text{soft}}$  (second, third, and fourth lines of

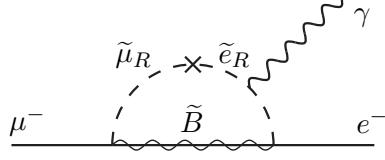


Figure 2.3: Possible contribution to  $\mu \rightarrow e\gamma$  from  $m_{\tilde{e}ij}$  soft term. Reprinted from Fig. 5.6(a) of ref. [6].

Eq. 2.14). Since  $a_u^{ij}$ ,  $a_d^{ij}$ ,  $a_e^{ij}$ ,  $m_{\tilde{Q}ij}$ ,  $m_{\tilde{L}ij}$ ,  $m_{\tilde{u}ij}$ ,  $m_{\tilde{d}ij}$ , and  $m_{\tilde{e}ij}$  are matrices in family space, any nonzero off-diagonal elements will lead to mixing between sfermions of different families. This can lead, for example, to contributions to the diagram  $\mu \rightarrow e\gamma$  (Figure 2.3) exceeding the experimental bounds. To avoid this disaster, *universality* conditions are assumed:

$$\mathbf{m}_{\tilde{\mathbf{Q}}}^2 = m_{\tilde{Q}}^2 \mathbf{1}, \mathbf{m}_{\tilde{\mathbf{L}}}^2 = m_{\tilde{L}}^2 \mathbf{1}, \mathbf{m}_{\tilde{\mathbf{u}}}^2 = m_{\tilde{u}}^2 \mathbf{1}, \mathbf{m}_{\tilde{\mathbf{d}}}^2 = m_{\tilde{d}}^2 \mathbf{1}, \mathbf{m}_{\tilde{\mathbf{e}}}^2 = m_{\tilde{e}}^2 \mathbf{1} \quad (2.15)$$

i.e. all sfermion mass matrices arising from the soft terms are assumed to be proportional to the unit matrix  $\mathbf{1}$ , such that there can be no flavor mixing from these terms and contributions to flavor-changing processes are drastically reduced.<sup>2</sup> In mSUGRA models, universality is assumed from the beginning, while in GMSB it is a natural consequence of the fact that the sparticle-messenger vertices are flavor-blind.

In minimal GMSB (mGMSB), there are four messenger supermultiplets  $q$ ,  $\bar{q}$ ,  $l$ ,  $\bar{l}$  providing the messenger (s)quarks and (s)leptons. There is one breaking scale  $\Lambda$ . The gaugino masses computed from diagrams like Fig. 2.2(b) are given by

$$M_a = \frac{\alpha_a}{4\pi} \Lambda \quad (2.16)$$

---

<sup>2</sup>Universality also includes some assumptions about the form of  $a_{u_{ij}}$ ,  $a_{d_{ij}}$ , and  $a_{e_{ij}}$  and the stipulation that the soft terms not introduce any CP-violating phases.

where  $a$  runs from 1-3 and the  $\alpha_a$  are the SM gauge coupling constants. The sfermion masses computed from diagrams like Fig. 2.2(a) are given by

$$m_{\phi_i}^2 = 2\Lambda^2 \sum_{a=1}^3 \left(\frac{\alpha_a}{4\pi}\right)^2 C_a(i) \quad (2.17)$$

where  $C_a(i)$  are group theory factors that are identical for all particles residing in the same type of supermultiplet (e.g. for all left-handed (s)quarks or left-handed (s)leptons). As explained in the previous paragraph, the gaugino and sfermion masses do not depend on fermion family.

In recent years, much theoretical progress has been made in unifying models of gauge mediation and developing less restrictive models than mGMSB. *General gauge mediation* (GGM) [9] retains the essential features of mGMSB, such as flavor degeneracy and communication of SUSY breaking via messengers, but does not make assumptions about the specific messenger sector or SUSY breaking scale. Many different collider final states can be interpreted in terms of GGM, and conversely, GGM implies a wealth of signatures, only a small fraction of which have been searched for at colliders [10, 11, 12]. The following section discusses the aspects of GGM collider phenomenology relevant to this thesis.

## 2.5 Phenomenology of General Gauge Mediation

The main distinguishing feature of all GMSB phenomenology is that the gravitino  $\tilde{G}$  is very light (eV-keV). In general, the gravitino mass is proportional to  $\langle F \rangle / M_P$ , where  $\langle F \rangle$  is the vacuum expectation value (VEV) of a field  $F$  that spontaneously breaks SUSY in the vacuum state and  $M_P$  is the Planck mass. In GGM models,  $\langle F \rangle \sim 10^8$  GeV, leading to a very light gravitino. In contrast, mSUGRA predicts  $\langle F \rangle \sim 10^{20}$  GeV. The fact that the gravitino is so much lighter than any other particles in the

supersymmetric Standard Model, and that it interacts only gravitationally (and thus extremely feebly), leads to two important phenomenological consequences:

1. All sparticle decay chains end with the production of a gravitino.
2. The gravitino escapes  $4\pi$ , hermetic collider detectors without interacting, leaving a signature of “missing” momentum transverse to the beam direction.

Even if the gravitino were lighter than any other sparticle, but heavier than an ordinary SM particle, it still could not decay to the SM particle due to *R-parity*. R-parity is a conserved quantity of the supersymmetric Standard Model that enforces baryon and lepton number conservation, which would otherwise be generically allowed at levels in conflict with experiment (e.g. the non-observation of baryon- and lepton-number-violating proton decay). All sparticles have R-parity -1, while all ordinary SM particles have R-parity +1, and R-parity conservation dictates that at any vertex, the product of the R-parities of each leg must be +1. This leads to two more important consequences:

1. Since conservation of energy only allows it to decay to ordinary SM particles, but R-parity prevents a sparticle-particle-particle vertex, the *lightest supersymmetric particle* (LSP) must be absolutely stable. All sparticle decays proceed through the *next-to-lightest supersymmetric particle* (NLSP), which in turn decays to the LSP. The fact that it is stable and only gravitationally interacting makes the gravitino a candidate dark matter particle (see Sec. 2.6).
2. In colliders, sparticles are produced in pairs (particle + particle  $\rightarrow$  sparticle + sparticle).

In GMSB, then, the gravitino is the LSP. If the NLSP is a gaugino,<sup>3</sup> then the possible decays depend on mixing among the gauginos. Due to the effects of EWSB,

---

<sup>3</sup>In principle, the NLSP could be anything, but in most popular GGM models, it is either a gaugino or a stau. The stau NLSP search is not the subject of this thesis, so it will not be considered in this section.

the four neutral gauginos  $\tilde{H}_u^0$ ,  $\tilde{H}_d^0$ ,  $\tilde{B}$ ,  $\tilde{W}^0$  mix into four *neutralino* mass eigenstates  $\tilde{\chi}_1^0$ ,  $\tilde{\chi}_2^0$ ,  $\tilde{\chi}_3^0$ ,  $\tilde{\chi}_4^0$ , and the four charged gauginos  $\tilde{H}_u^+$ ,  $\tilde{H}_d^-$ ,  $\tilde{W}^+$ ,  $\tilde{W}^-$  mix into two *chargino* mass eigenstates  $\tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_2^\pm$  (two mass eigenstates each with two possible charges = four particles). In the limit that EWSB effects are small, the neutralino and chargino masses can be written as the gauge eigenstate masses plus a small perturbation:

$$m_{\tilde{\chi}_1^0} = M_1 - \frac{m_Z^2 \sin^2 \theta_W (M_1 + \mu \sin 2\beta)}{\mu^2 - M_1^2} + \dots \quad (2.18)$$

$$m_{\tilde{\chi}_2^0} = M_2 - \frac{m_W^2 (M_2 + \mu \sin 2\beta)}{\mu^2 - M_2^2} + \dots \quad (2.19)$$

$$m_{\tilde{\chi}_3^0} = |\mu| + \frac{m_Z^2 (\text{sgn}(\mu) - \sin 2\beta)(\mu + M_1 \cos^2 \theta_W + M_2 \sin^2 \theta_W)}{2(\mu + M_1)(\mu + M_2)} + \dots \quad (2.20)$$

$$m_{\tilde{\chi}_4^0} = |\mu| + \frac{m_Z^2 (\text{sgn}(\mu) + \sin 2\beta)(\mu - M_1 \cos^2 \theta_W - M_2 \sin^2 \theta_W)}{2(\mu - M_1)(\mu - M_2)} + \dots \quad (2.21)$$

$$m_{\tilde{\chi}_1^\pm} = M_2 - \frac{m_W^2 (M_2 + \mu \sin 2\beta)}{\mu^2 - M_2^2} + \dots \quad (2.22)$$

$$m_{\tilde{\chi}_2^\pm} = |\mu| + \frac{m_W^2 \text{sgn}(\mu)(\mu + M_2 \sin 2\beta)}{\mu^2 - M_2^2} + \dots \quad (2.23)$$

where  $\tan \beta = \langle H_u^0 \rangle / \langle H_d^0 \rangle$ .

The two scenarios studied in ref. [12] are the following:

- **Bino NLSP:**  $M_1 \sim$  few hundred GeV,  $M_2, |\mu| \gg M_1$ . All but the lightest neutralino are effectively inaccessible at the LHC due to their large masses. The NLSP can always decay to  $\gamma + \tilde{G}$ , and if it is heavy enough, to  $Z + \tilde{G}$  or  $H + \tilde{G}$ .
- **Wino NLSP:**  $M_2 \sim$  few hundred GeV,  $M_1, |\mu| \gg M_2$ . The lightest neutralino and the lightest chargino are nearly degenerate in mass, and are the only two particles to play a role at the LHC. The decays described in the previous bullet point can happen, as well as chargino decays to  $W + \tilde{G}$ .

The subject of this thesis is the classic bino NLSP decay  $\gamma + \tilde{G}$ .

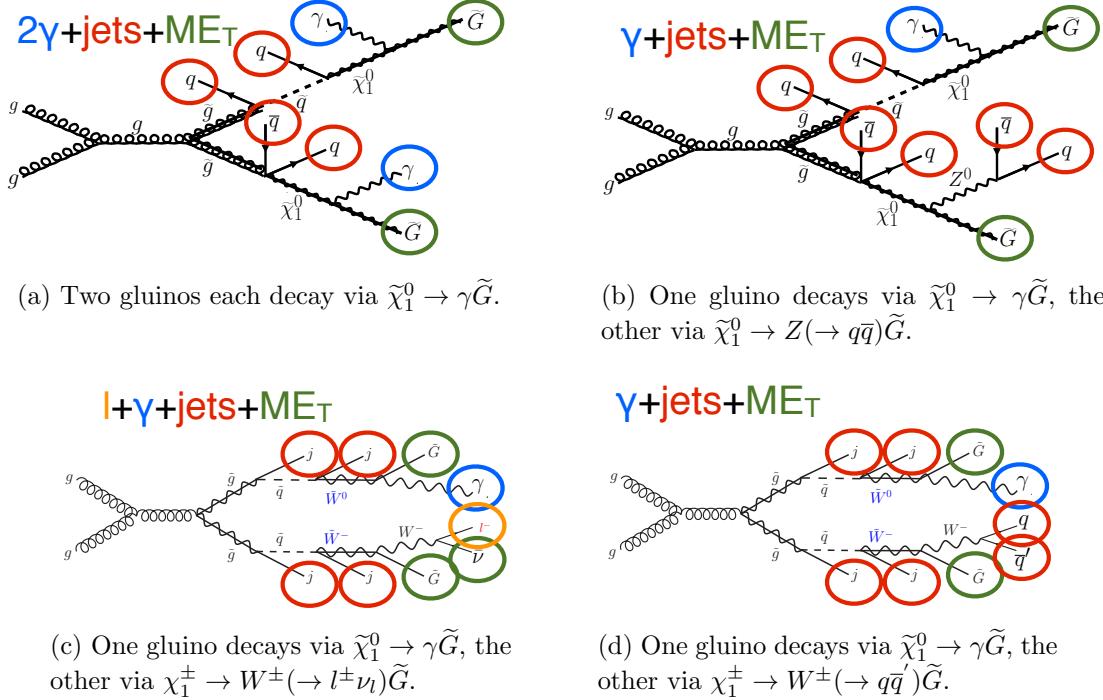


Figure 2.4: Typical LHC signatures of the bino and wino NLSP scenarios.

Since strong production of SUSY particles dominates over electroweak production at the LHC due to the enhanced  $gg$  parton luminosity over the  $q\bar{q}$  parton luminosity, early LHC searches are particularly sensitive to light squarks and gluinos. General gauge mediation makes no a priori restrictions on the mass splitting between the strongly interacting sparticles and the weakly interacting sparticles, so models with light squarks and gluinos are viable. In fact, such models could not be probed as well at the Tevatron<sup>4</sup> as they are at the LHC due to the aforementioned parton luminosities.

Typical LHC signatures of the bino and wino NLSP scenarios are shown in Figure 2.4.

<sup>4</sup>Located on the Fermilab site in Batavia, Illinois, the Tevatron was a proton-antiproton collider operating at 1.96 TeV center-of-mass energy. The Tevatron ran from 1987 to 2011 [13].

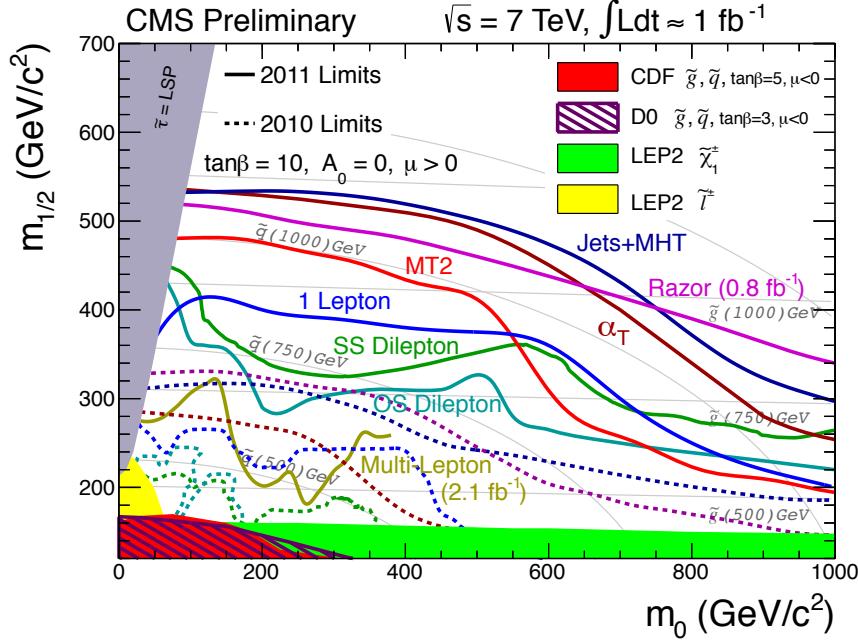


Figure 2.5: CMS limits on mSUGRA with  $\tan \beta = 10$ . The limits set by individual searches are shown as separate colored lines. Solid lines refer to 2011 searches (i.e. using an integrated luminosity of  $\sim 1 \text{ fb}^{-1}$ ), while dashed lines refer to 2010 searches ( $\sim 36 \text{ pb}^{-1}$ ). Reprinted from ref. [16].

## 2.6 Experimental Status of SUSY

Collider searches for evidence of supersymmetry began in earnest in the 1980s [14] and continue to this day. Most recently, the LHC and Tevatron experiments have set the strictest limits on a variety of SUSY breaking scenarios, including GMSB and mSUGRA.

Figure 2.5 shows the current limits set by the CMS experiment on the mSUGRA model (with  $\tan \beta = 10$ ) in the  $m_0$ - $m_{1/2}$  plane. (Note that although the plot is truncated at  $m_0 = 1000 \text{ GeV}/c^2$ , some searches are sensitive out to  $m_0 \sim 2000 \text{ GeV}/c^2$ .) Although the LHC has pushed  $m_0$  above  $\sim 1 \text{ TeV}/c^2$  for  $m_{1/2}$  up to  $\sim 400 \text{ GeV}/c^2$ , casting some doubt onto the theory's prospects for solving the hierarchy problem, there is still a sizable chunk of mSUGRA parameter space that is not ruled out by collider experiments. Furthermore, parts of the CMS unexplored regions overlap with areas allowed by astrophysics experiments [15].

Figure 2.6 shows the most up-to-date limit (using  $1 \text{ fb}^{-1}$  of integrated luminosity collected by the ATLAS experiment [17] at the LHC) on the Snowmass Points and Slopes (SPS) model of mGMSB, dubbed SPS8 [18]. The best limits on a variety of GGM models are shown in Figure 2.7 [19]. In these models, no assumptions are made about the specific parameters common to many gauge mediation models (e.g. the number of messengers or the relationship between the messenger mass and the SUSY breaking scale). Instead, it is only assumed that the lightest neutralino is light enough to be produced on-shell at the LHC (by setting  $M_1$  and  $M_2$  appropriately, see Sec. 2.5) and that it decays to a gravitino, that the gravitino is extremely relativistic (mass of order eV-keV), and that the gravitino is stable. The one-dimensional scan over SUSY breaking scales in the SPS8 model (in which the full sparticle spectrum is specified by the model parameters) is replaced by a two-dimensional scan over gluino and lightest neutralino mass in the GGM models (in which all sparticles except the gluino, first- and second-generation squarks, and neutralinos are forced to be at  $\sim 1.5 \text{ TeV}/c^2$ , effectively decoupling them from the dynamics that can be probed with  $1 \text{ fb}^{-1}$  at a  $7 \text{ TeV}/c$  pp collider).

In general, the lifetime of the lightest neutralino in GMSB models can take on any value between hundreds of nanometers to a few kilometers depending on the mass of the lightest neutralino and the SUSY breaking scale [6]. The search published in ref. [10] (from which Figs. 2.6 and ?? are culled) considers only *prompt* neutralino variants, i.e. with neutralino lifetime short enough that the distance traveled by the neutralino before decay cannot be resolved by the detector. The most recent limits on non-prompt SPS8-style neutralino models were set by the Collider Detector at Fermilab (CDF) collaboration with  $570 \text{ pb}^{-1}$ , and are shown in Figure 2.8 [11].

Finally, if the gravitino is to make up some or all of the dark matter, constraints on the form of gauge mediation must come from cosmological considerations and astronomical observations. The gravitino in gauge mediation models is usually very

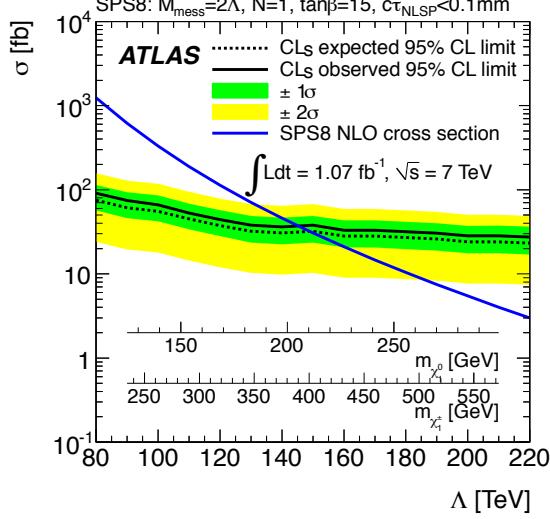


Figure 2.6: ATLAS cross section upper limit on the SPS8 [18] model of mGMSB as a function of SUSY breaking scale  $\Lambda$ , lightest neutralino mass  $m_{\tilde{\chi}_1^0}$ , or lightest chargino mass  $m_{\tilde{\chi}_1^\pm}$ . Values of  $\Lambda$ ,  $m_{\tilde{\chi}_1^0}$ , or  $m_{\tilde{\chi}_1^\pm}$  below the intersection point between the blue (predicted SPS8 cross section) and black (observed cross section upper limit) curves are excluded. The model parameters listed above the plot are defined in Secs. 2.4 and 2.5, except for  $\tau_{\text{NLSP}}$ , which is the neutralino lifetime. Reprinted from ref. [10].

light ( $\mathcal{O}(\text{eV-MeV})$ ) because it is proportional to the SUSY breaking scale divided by the Planck mass, and in GMSB the breaking scale is typically only of order a few hundred TeV ([6] and Sec. 2.5). A light, highly relativistic dark matter particle might have been produced, for instance, in the early, radiation-dominated period of the universe [20]. This *warm dark matter* (WDM) may be responsible for all of the dark matter needed to account for galactic structure, or it may share the duties with *cold dark matter* (CDM, weakly interacting particles with masses in the GeV range). In any viable model, the predicted relic density of the dark matter species must match the observed value of  $\Omega h^2 \sim 0.1$  [21]. For many GMSB models, this measurement constrains the gravitino mass to the keV range [22]. This constraint, however, does not translate into a very strong bound on the lifetime of the lightest neutralino. Using the following equation (taken from [22]):

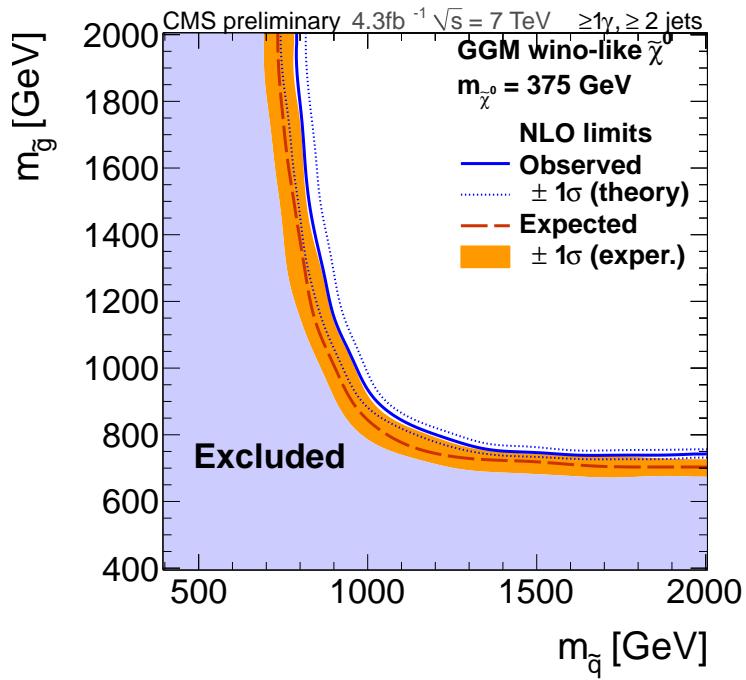
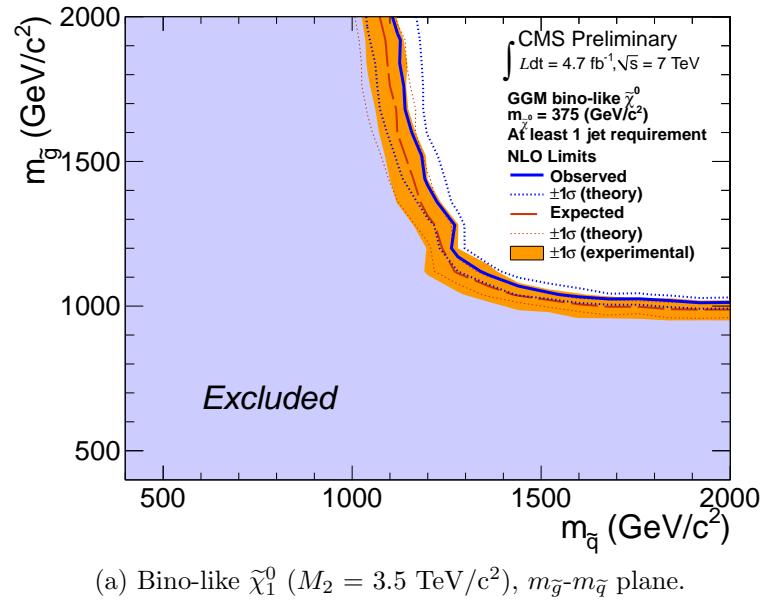


Figure 2.7: CMS exclusion contours for GGM models using  $5 \text{ fb}^{-1}$  of LHC data. Reprinted from ref. [12].

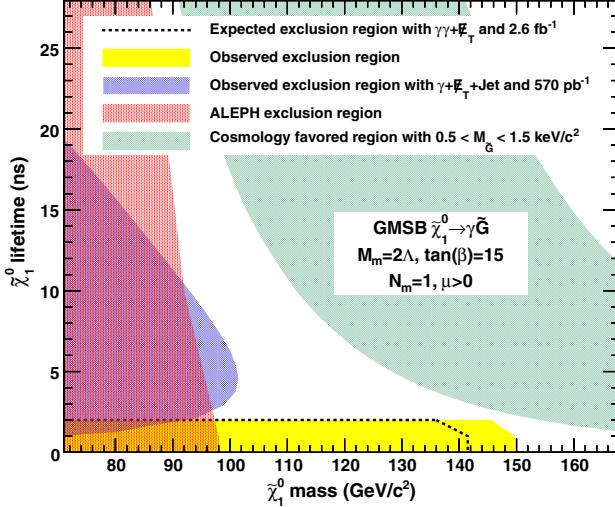


Figure 2.8: CDF exclusion contour in the  $\tau_{\tilde{\chi}_1^0}$ - $m_{\tilde{\chi}_1^0}$  plane, where  $\tau_{\tilde{\chi}_1^0}$  is the lifetime of the neutralino. Reprinted from ref. [11].

$$\tau_{\tilde{\chi}_1^0} \sim 130 \left( \frac{100 \text{ GeV}}{m_{\tilde{\chi}_1^0}} \right)^5 \left( \frac{\sqrt{\langle F \rangle}}{100 \text{ TeV}} \right)^4 \mu\text{m} \quad (2.24)$$

and applying the gravitino mass constraint  $\sqrt{\langle F \rangle} \lesssim 3000 \text{ TeV}$  (cf. the first paragraph of Sec. 2.5 with  $m_{\tilde{G}} \sim \text{keV}$ ) and  $m_{\tilde{\chi}_1^0} = 100 \text{ GeV}$ , the upper bound on the neutralino lifetime is 100 meters. For  $\sqrt{\langle F \rangle} \sim 100 \text{ TeV}$ , the neutralino lifetime is detectable on collider time scales.

Recently, a lower bound on the WDM particle mass in either pure warm or mixed warm and cold dark matter scenarios was set using observations of the Lyman- $\alpha$  forest. For pure WDM,  $m_{\text{WDM}} > 8 \text{ keV}$ , while for some mixed WDM-CDM scenarios,  $m_{\text{WDM}} > 1.1\text{-}1.5 \text{ keV}$  [20, 23]. These bounds and others have motivated the development of more complicated gauge mediation models [23]. However, rather than focus on a specific GMSB model, of which there are many, the search detailed here is interpreted in a minimally model dependent way. With this approach, the results can be applied to many competing models. The remainder of this thesis is devoted to the experimental details of the search, analysis strategy, and presentation of the results. The work described in this thesis forms the basis for the CMS public result “Search

for Supersymmetry in Events with Photons and Missing Energy” [19], published in April 2012.

# Chapter 3

## Event Selection

In keeping with the phenomenology described in Sec. 2.5, the candidate GGM events selected in this search consist of two high- $E_T$  photons and a significant momentum imbalance transverse to the beam, indicating the production of an escaping gravitino. This momentum imbalance is usually referred to as *missing transverse energy* and is denoted by the symbol  $\cancel{E}_T$ .

However, in order to use real CMS data (as opposed to simulation) to derive predictions for the backgrounds to the search, *control samples* distinct from the *candidate* two-photon sample must be collected. These samples consist of different numerical combinations of photons, electrons, and jets, and are explained in more detail in Chapter 4. Since this search is performed in the high- $\cancel{E}_T$  tail of the  $\cancel{E}_T$  distribution, where adequate detector simulation is very difficult, it is advantageous to use *data-driven* background estimates, which capture the true detector response, over numbers derived from simulation.

In the following sections, the reconstruction of photons, electrons, jets, and  $\cancel{E}_T$  is explained. Sec. 3.1 begins with an explanation of the high level reconstruction. It is followed by Sec. 3.2, which describes the triggers used to collect the candidate and control samples. Finally, the chapter concludes with a measurement of the photon

identification efficiency in Sec. 3.4.

## 3.1 Object Reconstruction

This section describes the *offline* object reconstruction, i.e. the reconstruction of particle objects from events that have already been triggered and written to permanent storage, as opposed to the building of trigger objects explained in Secs. ?? and 3.2.

### 3.1.1 Photons

#### Uncalibrated EB/EE Hits

Photon reconstruction begins with the ADC count value for each of the 10 recorded time samples per ECAL crystal per trigger. To construct an *uncalibrated hit*, the gain (1, 6, or 12; see Sec. ??) of each sample is determined and the ADC count value scaled appropriately. The pedestal is estimated from the average of the first three samples, which, for a properly time in hit, should contain no signal. This pedestal value is subtracted from the rest of the samples. Finally, the amplitude of the pulse is reconstructed using a predetermined weight for each sample [24]. The weights correspond to the pulse shape expected from the MGPA and shaping circuit response. The time of the hit is also reconstructed using the ratios between neighboring time samples [25]. A typical ECAL channel pulse shape is shown in Figure 3.1.

#### Calibrated EB/EE Hits

In the next phase of the photon reconstruction, calibrations are applied to the uncalibrated hits to form *calibrated hits* with energy measured in GeV. Channels are excluded from seeding calibrated hits if

- they are excessively noisy,

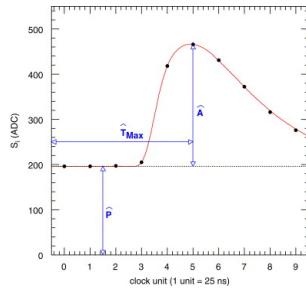


Figure 3.1: Typical ECAL channel pulse shape.  $\hat{P}$  is the pedestal value,  $\hat{A}$  is the pulse amplitude, and  $\hat{T}_{\max}$  is the hit time. The red line is the assumed pulse shape from which the weights are derived. Reprinted from ref. [24].

- they are stuck in fixed gain,
- they are totally dead,
- they have one or more neighboring dead channels, or
- they do not have good trigger primitives (i.e. trigger primitive is missing, saturated, or spike-like).

In addition, no uncalibrated hits that are spike-like are eligible for calibration. The calibrations applied are crystal transparency loss corrections measured continuously by the laser/LED system, energy intercalibrations (relative energy calibration between crystals), absolute scale calibrations between ADC counts and GeV,<sup>1</sup> and time intercalibrations (relative time calibration between crystals).

The ECAL crystals were pre-calibrated before installation in CMS using laboratory light yield and photodetector gain measurements [27]. In addition, some EB and EE crystals were intercalibrated using test beams [28], and all EB crystals were intercalibrated with cosmic ray muons [29]. EE precalibrations were validated with LHC *splash events* in 2009 [29, 30], in which the beam was dumped onto a collimator approximately 150 meters upstream of CMS, causing a spray of muons to enter CMS at

---

<sup>1</sup>The ADC-GeV scale factors (one for EB and one for EE) are defined such that the sum of fully calibrated and scaled hits in a particular  $5 \times$  cluster of crystals (plus the associated energy deposited in ES) is 50 GeV for a 50 GeV incident unconverted photon [26].

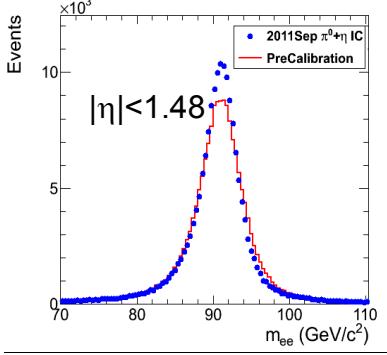


Figure 3.2:  $Z$  peak reconstructed using pre-LHC calibration constants (red) or September 2011  $\pi^0/\eta$ -derived intercalibration constants (blue). Reprinted from ref. [32].

one endcap and exit at the other. Splash events were also used to derive time intercalibration constants. Before colliding beam operations commenced, the intercalibration precision was estimated to be 0.5%-2.2% in EB and 1%-5% in EE [31].

Three calibration methods were employed once colliding beam operations began:

- $\phi$  symmetry relative calibration between crystals, exploiting the azimuthal symmetry of CMS
- $\pi^0$  and  $\eta$  relative calibration between crystals, using the diphoton decays of these particles
- $E/p$  absolute calibration, comparing the momentum measured in the tracker  $p$  to the energy measured in the ECAL  $E$  of a sample of electrons from  $Z$  decay

By September 2011, the intercalibration precision in EB was measured to be between 0.3% and 1.1% using the  $\pi^0/\eta$  method [32]. Figure 3.2 shows the improvement in  $Z$  reconstruction from pre-LHC calibration constants to the latest  $\pi^0/\eta$ -derived constants.

## Calibrated ES Hits

ES calibrated hits are formed from the three samples read out per sensor. Just as in the case of EB/EE crystals, ES uncalibrated hits gain-adjusted, pedestal-subtracted, and reconstructed using weights. To make a calibrated ES hit, intercalibration constants, angle correction constants, and a MIP-GeV absolute scale factor are applied.

## Clustering

After calibrated ECAL hits are formed, they must be clustered into shapes that represent the energy deposit from a single particle. *Basic clusters* are formed around seed hits, defined as a hit that

**Add description of timing cut and topological spike rejection variables, or earlier.**

- has calibrated  $E_T > 1(0.18)$  GeV in EB(EE),
- does not originate from a dead channel or one with faulty hardware,
- is not poorly calibrated,
- was reconstructed with the standard algorithm (i.e. not a special recovery algorithm for channels with subpar data integrity),
- is not saturated,
- is not spike-like (**define spike rejection variables here**), and
- is in time (EB).

EB basic clusters are formed around the seeds via the *hybrid* algorithm, while EE basic clusters are formed with the *multi5x5* algorithm [33]. In addition to non-radiating electrons and unconverted photons, both algorithms are designed to also recover all of the energy associated with electron bremsstrahlung deposits and photon conversions.

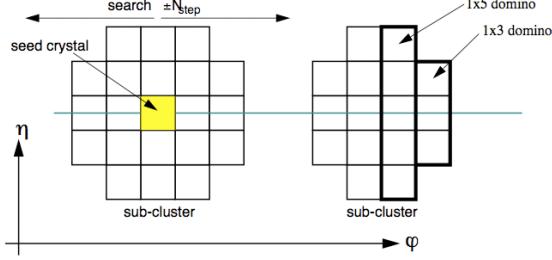


Figure 3.3: Hybrid algorithm in EB. The shower extent is essentially constant in  $\eta$ , but spreads out in  $\phi$  as the two sub-clusters (or basic clusters) are grouped into the same supercluster. Reprinted from ref. [33].

The geometry of the CMS magnetic field means that bremsstrahlung and conversions will tend to spread the shower out in  $\phi$ , not  $\eta$ . Both algorithms work by forming basic clusters around seeds, then combining the basic clusters into *superclusters* (SC) by searching in a window extended in the  $\phi$  direction for all basic clusters consistent with bremsstrahlung radiation from the primary electron, or with a photon conversion. Figure 3.3 illustrates the hybrid algorithm in EB. In EE, the energy deposited in ES must also be added into the total clustered energy sum.

Figure 3.4 shows the effect of superclustering on  $Z \rightarrow ee$  reconstruction.

### Supercluster Corrections

The total clustered ECAL energy is defined as

$$E = F \times \sum_{i=1}^{n_{\text{crystal}}} G \times c_i \times A_i \quad (3.1)$$

where  $G$  is the ADC-GeV or MIP-GeV scale factor,  $c_i$  are the intercalibration constants,  $A_i$  is the uncalibrated hit amplitude in ADC counts, and  $F$  is SC correction factor.  $G$  and  $c_i$  were explained in Sec. 3.1.1.  $F$  is a product of three factors for hybrid SCs (two for multi5x5 SCs) [33]:

1.  $C_{EB}(\eta)$ , which compensates for lateral energy leakage due to the crystal off-

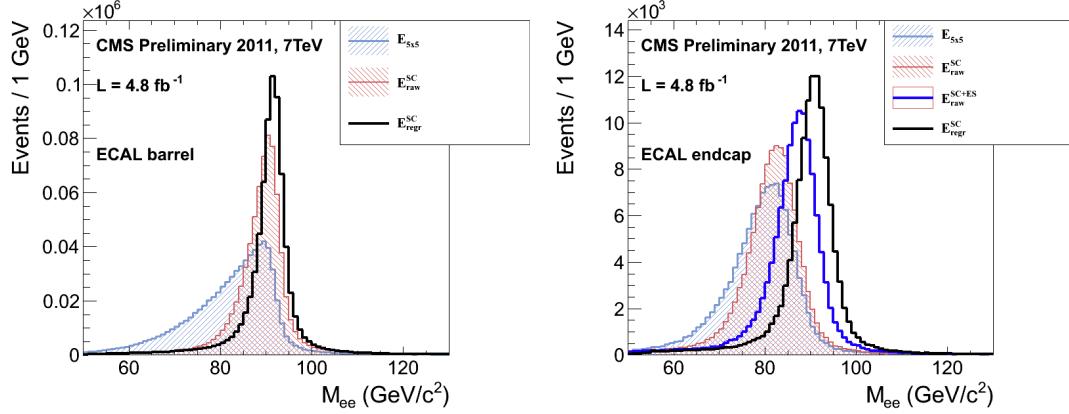


Figure 3.4:  $Z$  peak reconstructed in the dielectron channel for different kinds of clustering. The left plot is for EB and the right plot for EE. The constituent hits were calibrated with the best available intercalibrations and laser calibrations as of December 2011. The light blue histogram shows the reconstruction using a  $5 \times 5$  energy sum, the red histogram shows the reconstruction using the SC energy for crystals only (the dark blue histogram on the right-hand side adds in the energy from ES), and the black histogram shows the reconstruction after the SCs are corrected using a multivariate method [?]. Reprinted from Fig. 30 of ref. [?].

pointing in EB. These corrections are taken from MC simulation [33] and were confirmed in test beams [28].

2.  $f(\text{brem})$ , which corrects for the biases in the clustering algorithms for showers characterized by differing amounts of bremsstrahlung. These corrections are taken from MC simulation [33].
3. Residual correction  $f(E_T, \eta)$ , due to the variation in  $\eta$  of detector material traversed by a primary electron or photon, and to any residual  $E_T$  dependence of the reconstruction. These corrections are determined from  $Z \rightarrow ee$  data samples.

As a benchmark of ECAL calibration performance, the extra energy smearing in MC needed to achieve data/MC agreement in the  $Z$  width was between  $\sim 0.9\%$  (in the central part of EB for electrons with little bremsstrahlung) and  $\sim 3.3\%$  (in the outer part of EE for heavily radiating electrons) [34].

## From Supercluster to Photon

The CMS photon object is any SC with  $E_T > 10$  GeV and  $H/E < 0.5$ , unless the SC  $E_T > 100$  GeV, in which case the  $H/E$  requirement is dropped.  $H/E$  is defined as the ratio of energy in the HCAL in a 0.15 cone around the SC centroid, directly behind the SC, to the SC energy. SCs with  $R9 > 0.94(0.95)$  in EB(EE), where  $R9$  is defined as the ratio of the energy in the central  $3 \times 3$  cluster of crystals divided by the SC energy  $E_{3 \times 3}/E_{SC}$ , are the best calibrated and most accurate type of electromagnetic shower. Therefore, for these objects, the photon energy is defined as the energy sum of the fully calibrated hits in the central  $5 \times 5$  cluster around the seed (with  $C_{EB}(\eta)$  applied for EB photons). For all other SCs, the photon energy is equal to the fully corrected SC energy (cf. Sec. 3.1.1).

In this search, candidate photons and *fake photons* ( $f$ , “fakes”) are further selected according to the criteria listed in Table 3.1. Fakes are used in the determination of the QCD background, as explained in Chapter 4.  $I_{\text{comb}}$  is defined as

$$I_{\text{comb}} = I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}} \quad (3.2)$$

where  $I_{\text{ECAL}}$ ,  $I_{\text{HCAL}}$ , and  $I_{\text{track}}$  are  $E_T$  sums in the annular regions defined in Figure 3.5 and  $\rho$  is the average pileup energy density in the calorimeters (per unit  $\eta \cdot \phi$ ) as measured with the Fastjet algorithm [35, 36]. Note that the ECAL and track isolation veto strips at constant  $\eta$  ensure that the isolation cuts are similarly efficient for converted photons, radiating electrons, and unconverted photons.

Figure 3.6 shows the  $\rho$  distribution for a sample of events with at least two 25 GeV EM objects passing the  $|\eta|$ ,  $H/E$ , and  $R9$  requirements in Table 3.1, and passing the trigger requirements in Table 3.3, representing the full 2011 dataset. Since average  $\rho$  is  $\sim 5$  GeV, and there is a long tail above this average value, it is necessary to

Table 3.1: Selection criteria for photons and fakes.

Variable	Cut ( $\gamma$ )	Cut ( $f$ )
SC $ \eta $	$< 1.4442$	$< 1.4442$
$H/E$	$< 0.05$	$< 0.05$
$R9$	$< 1$	$< 1$
Has pixel seed	No	No
$I_{\text{comb}}, \sigma_{i\eta i\eta}$	$< 6 \text{ GeV} \&\& < 0.011$	$(\geq 6 \text{ and } < 20 \text{ GeV}) \parallel \geq 0.011$

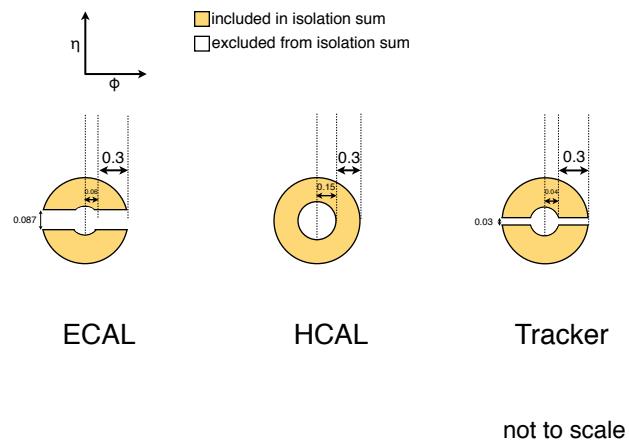


Figure 3.5: ECAL, HCAL, and track Isolation cones.

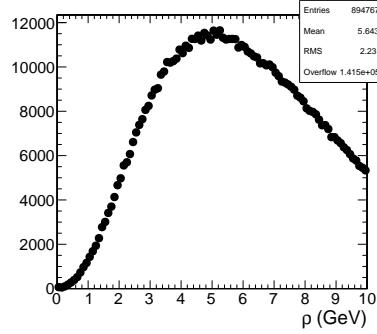


Figure 3.6:  $\rho$  distribution for a sample of events with at least two 25 GeV EM objects passing the  $|\eta|$ ,  $H/E$ , and  $R9$  requirements in Table 3.1, and passing the trigger requirements in Table 3.3. This sample covers the full 2011 dataset.

subtract pileup energy from the ECAL and HCAL isolation cones to recover otherwise clean photons in events with large pileup. The ECAL and HCAL *effective areas* of 0.0792 and 0.0252 (**Change these to the final values used in the analysis.**), respectively, are calculated by fitting the average ECAL or HCAL isolation energy vs.  $\rho$  in a sample of  $Z \rightarrow ee$  events to a straight line. The slope of the line—which has the units of  $\eta \cdot \phi$ , or area—is the effective area.

#### **Summarize the combined isolation lower bound optimization here.**

The upper bound on fake photon combined isolation guarantees that poorly isolated dijet events, with  $\cancel{E}_T$  resolution dissimilar to the candidate diphoton events, do not enter the  $ff$  sample. The exact value of 20 GeV (cf. Table 3.1) arises from a low- $\cancel{E}_T$   $ff/\gamma\gamma \chi^2$  optimization procedure [37]. Figure 3.7 shows the value of the Neyman’s  $\chi^2$  between the  $ff$  and  $\gamma\gamma$   $\cancel{E}_T$  distributions, truncated at either 25 or 50 GeV, vs. upper bound on fake combined isolation. As shown in the figure, 20 GeV very nearly minimizes the  $\chi^2$ , while also being large enough that a sufficient number of  $ff$  events may be collected.

#### **Summarize the effective area calculation here?**

A “pixel seed” is defined as a hit in the pixel detector consistent with a track extrapolated from the position of the ECAL SC back to the primary vertex. Real photons, having no charge and therefore no bending in the magnetic field, should not

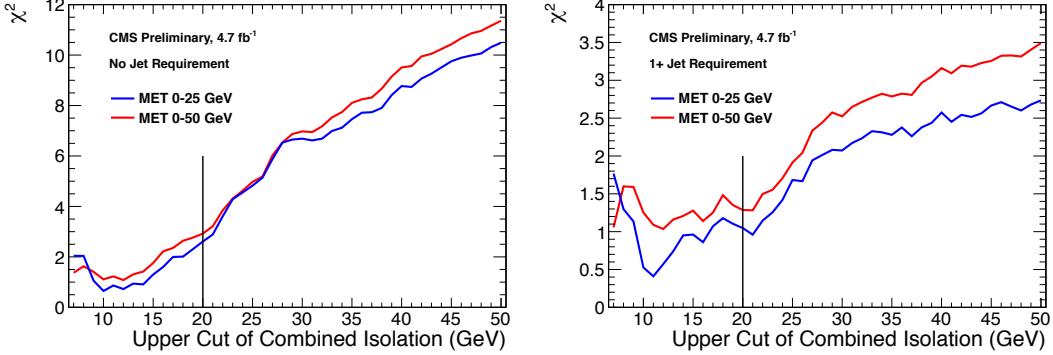


Figure 3.7: Neyman’s  $\chi^2$  between the  $ff$  and  $\gamma\gamma \not{E}_T$  distributions, truncated at either 25 (red) or 50 (blue) GeV, vs. upper bound on fake combined isolation. The left plot includes all events; the right plot is for events with  $\geq 1$  jet defined as in Table 3.2, but with the  $\Delta R$  cleaning criteria applied to the two primary EM objects and all additional electrons, photons, and fake photons. The full reweighting and normalization procedure is employed in the  $\not{E}_T$  calculation (see Sec. 4.1). Error bars include statistical, reweighting, and normalization error (see Sec. ??). Reprinted from Fig. 9 of ref. [37].

have a pixel seed.

### 3.1.2 Electrons

Electrons are reconstructed identically to photons, except that in the electron case the presence of a pixel seed is enforced, rather than vetoed.<sup>2</sup> Photons and electrons are defined by very similar criteria so that  $Z \rightarrow ee$  events can be used to model the QCD background in the two-photon sample without introducing any bias in the electron energy measurement (cf. Sec. 4.1).

### 3.1.3 Jets and Missing Transverse Energy

#### Particle Flow

In this analysis, jets and  $\not{E}_T$  are formed from *particle flow* (PF) candidates. The particle flow algorithm [39, 40] uses information from all CMS subdetectors to reconstruct

<sup>2</sup>In many CMS analyses, electrons are reconstructed very differently from photons. In particular, a special tracking algorithm [38] is used to best follow a radiating electron. However, in this analysis, the electron tracking is not used.

as accurately as possible the positions and momenta of all visible jet constituents, exploiting the fine granularity of the tracker and ECAL to achieve a greatly improved momentum resolution over calorimeter-only jets [41]. The PF algorithm is summarized below [42].

1. Reconstruct the fundamental detector objects via iterative procedures
  - Tracks in the inner silicon layers
    - High efficiency and low fake rate for charged hadrons in jets
    - Relaxed primary vertex constraint allows photon conversions, particles originating from nuclear interactions in the silicon, and long-lived particles to be reconstructed
  - Calorimeter clusters
  - Muon tracks in the outer muon layers
2. Create a “block” of linked fundamental objects
  - Link silicon tracks to calorimeter clusters via  $\Delta R_{\text{track-cluster}}$  (account for electron bremsstrahlung)
  - Link clusters in one calorimeter layer to clusters in a separate layer via  $\Delta R_{\text{cluster-cluster}}$
  - Link silicon tracks to muon tracks via global track  $\chi^2$
3. ID the particles in the block
  - If global (silicon + muon layers) muon  $p_T$  is compatible with silicon track  $p_T$ , ID as a muon and remove corresponding tracks from block
  - ID electron tracks via special algorithm and removed all corresponding tracks and cluster from block

- Remove fake tracks from the block
- Remove excess track-cluster links via  $\Delta R$  track – cluster minimization (but allow multiple tracks to be associated to one cluster)
- If the cluster energy is significantly larger than the energy of the linked track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- If the cluster is not linked to a track, ID as a PF photon or PF neutral hadron and remove corresponding clusters from the block
- Remaining track-cluster links are PF charged hadrons

## Jets

PF candidates are clustered into jets by means of the anti- $k_T$  algorithm with  $R = 0.5$  [43]. In this algorithm, all possible pairs of PF candidates  $i, j$  are looped over, and the momenta of the pair that minimize the distance variable

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2 \max(k_{Ti}^2, k_{Tj}^2)} \quad (3.3)$$

are combined, where  $k_{Ti}$  is the transverse momentum of “combined” PF candidate  $i$ . The constituent PF candidates are clustered together. The process is repeated until  $d_{ij} > 1/k_{Ti}^2$  for all pairs of clustered PF momenta [44]. An illustration is given in Figure 3.8. The anti- $k_T$  algorithm is infrared and collinear safe, leading to well-behaved theoretical predictions and ease of comparison between data and MC simulation. It also tends to form circular jets, making it easy for experimental effects such as expected out-of-cone energy and fiducial acceptance to be measured or simulated. For these reasons, the anti- $k_T$  jet clustering algorithm was chosen for this analysis.

Once jets are clustered, they must be corrected for biases in the energy mea-

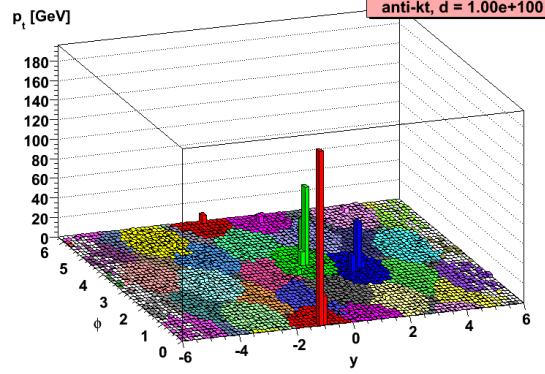


Figure 3.8: Example event display showing jets clustered via the anti- $k_T$  algorithm.  $y$  is pseudorapidity.

surement due to non-compensation [45], invisible energy (lost to overcoming nuclear binding energy, in neutrinos, or in unclustered muons, for example) [45], detector geometry and cracks [46], zero suppression and trigger inefficiencies [47], pileup, and effects of the clustering algorithm [46]. Four multiplicative correction factors are applied to the raw jet four-momentum  $p_\mu^{\text{raw}}$  [41]:

- $C_{\text{offset}}(p_T^{\text{raw}})$ , which accounts for extra energy due to noise, pileup, and the underlying event;
- $C_{\text{MC}}(C_{\text{offset}}p_T^{\text{raw}}, \eta)$ , which is derived from MC and accounts for most of the  $p_T$  and  $\eta$  dependence;
- $C_{\text{rel}}(\eta)$ , which accounts for the remaining differences in uniformity over the entire calorimeter between data and MC; and
- $C_{\text{abs}}(C_{\text{rel}}C_{\text{MC}}C_{\text{offset}}p_T^{\text{raw}})$ , which accounts for the remaining differences in linearity over the full  $p_T$  range between data and MC.

Figure 3.9 shows the total jet energy correction factor  $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$  vs.  $\eta$  for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . The PF jet corrections are more uniform across  $\eta$  than those of CALO jets (composed of simple calorimeter towers) or JPT jets (Jet Plus Tracks; composed of calorimeter energies replaced,

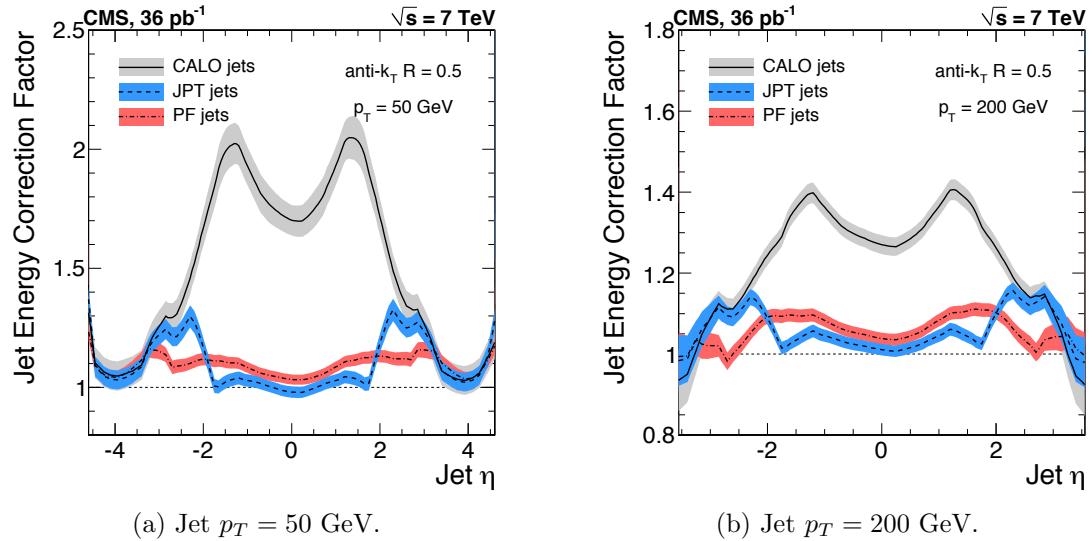


Figure 3.9: Total jet energy correction factor  $C_{\text{offset}}C_{\text{MC}}C_{\text{rel}}C_{\text{abs}}$  vs.  $\eta$ , including uncertainty band, for jets reconstructed with the anti- $k_T$  algorithm,  $R = 0.5$ . Reprinted from Fig. 26 of ref. [41].

where possible, with matching track  $p_T$ ) [48]. In addition, for  $p_T$  in the range 30–200 GeV and  $|\eta|$  up to 2.0, the PF jet energy correction uncertainty is lower than that of the other two types of jets, and never exceeds  $\sim 3\%$  [41]. The superior performance of PF jets motivates their use in this search.

In this analysis, candidate and QCD control events are binned by number of jets satisfying the criteria in Table 3.2.

## Missing Transverse Energy

To be consistent with the jet reconstruction,  $\cancel{E}_T$  in this analysis is also reconstructed from PF candidates. Raw  $\cancel{E}_T$  is defined as

$$\mathcal{E}_{T\text{raw}} = \left| - \sum_{i=1}^{n_{\text{PF}}} \vec{p}_T i \right| \quad (3.4)$$

where  $n_{\text{PF}}$  is the number of PF candidates in the event.  $\cancel{E}_{\text{Raw}}$  may be corrected for the same effects that necessitate jet corrections, since  $\cancel{E}_{\text{Raw}}$  is usually the result of jet mis-measurement (except, of course, in electroweak physics processes that include an energetic neutrino, or SUSY production). CMS *Type-I*  $\cancel{E}_T$  corrections simply involve replacing the PF jets with their corrected energies (cf. Sec 3.1.3) and recalculating  $\cancel{E}_T$ . Only jets with *electromagnetic fraction* (EMF) below 90% and  $p_T > 20$  GeV are replaced. This ensures that very electromagnetic jets (as well as isolated leptons, which also receive no correction), which consist chiefly of neutral pions and are measured accurately by the ECAL, do not receive a correction derived for jets with a large fraction of their energy in charged hadrons. In addition, the  $p_T$  cut guarantees that jet corrections are only applied where they are known to within a few percent. For this search, the level of agreement between the SM background estimate and the two-photon search sample in a low- $\cancel{E}_T$  control region is the same regardless of whether the  $\cancel{E}_T$  is corrected or not, so for simplicity the Type-I  $\cancel{E}_T$  corrections are not used (see Sec. ??).

Figure 3.10 shows the  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\cancel{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25$  GeV. Again, PF  $\cancel{E}_T$  outperforms  $\cancel{E}_T$  constructed of calorimeter towers or track-corrected calorimeter deposits.

## 3.2 HLT

From the objects described in Sec. 3.1, four samples of events are formed:

- $\gamma\gamma$  candidate sample, in which the two highest  $E_T$  objects are photons,
- $e\gamma$  control sample, in which the two highest  $E_T$  objects are one electron and one photon,
- $ee$  control sample, in which the two highest  $E_T$  objects are electrons, and

Table 3.2: Definition of HB/HE hadronic jets. **Add a footnote describing the PF electron and PF muon definitions, with references.**

Variable	Cut
Algorithm	L1FastL2L3Residual corrected PF
$p_T$	$> 30 \text{ GeV}$
$ \eta $	$< 2.6$
Neutral hadronic energy fraction	$< 0.99$
Neutral electromagnetic energy fraction	$< 0.99$
Number of constituents	$> 1$
Charged hadronic energy	$> 0.0 \text{ GeV}$ if $ \eta  < 2.4$
Number of charged hadrons	$> 0$ if $ \eta  < 2.4$
Charged electromagnetic energy fraction	$< 0.99$ if $ \eta  < 2.4$
$\Delta R$ to nearest electron, muon, or one of the two primary EM objects	$> 0.5$

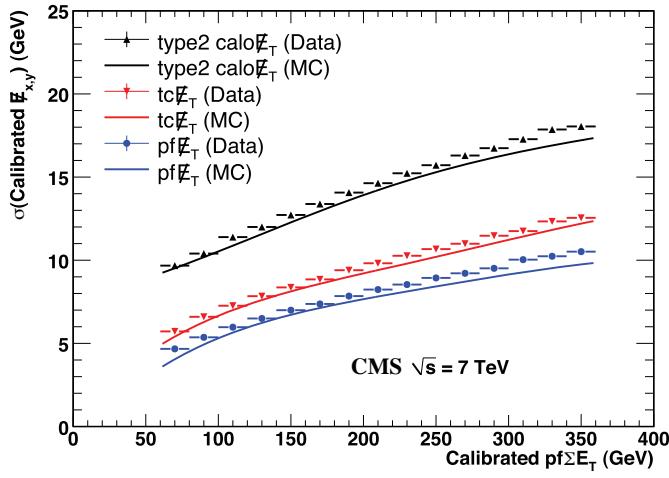


Figure 3.10:  $\sigma$  of a Gaussian fit to the x- and y-components of calibrated  $\not{E}_T$  vs. the calibrated PF  $E_T$  sum in a sample of events containing at least two jets with  $p_T > 25 \text{ GeV}$ .  $\sigma$  is calibrated such that the  $\not{E}_T$  scale is equal for all three algorithms. PF  $\sum E_T$  is corrected, on average, to the particle level using a Pythia v8 simulation [49]. The blue markers (data) and line (MC) refer to PF jets. Reprinted from Fig. 13 of ref. [47].

- $ff$  control sample, in which the two highest  $E_T$  objects are fakes.

In all samples, the leading EM object is required to have  $E_T > 40$  GeV, while the trailing EM object is required to have  $E_T > 25$  GeV. The high level triggers used to select the four samples, by run range, are listed in Table 3.3. No trigger is prescaled.

Table 3.3: HLT paths triggered by the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  samples, by run range. No triggers are prescaled.

Run range	$\gamma\gamma$	$e\gamma$	$ee$	$ff$
160404-163261	Photon26_-	Photon26_-	Photon26_-	Photon26_-
	IsoVL_-	IsoVL_-	IsoVL_-	IsoVL_-
	Photon18	Photon18	Photon18	Photon18
161216-166967	Photon36_-	Photon36_-	Photon36_-	Photon36_-
	CaloIdL_-	CaloIdL_-	CaloIdL_-	CaloIdL_-
	Photon22_-	Photon22_-	Photon22_-	Photon22_-
166347-180252	CaloIdL	CaloIdL	CaloIdL	CaloIdL
				Photon36_-
				CaloIdL_-
				IsoVL_-
				Photon22_-
				CaloIdL_-
	Photon36_-	Photon36_-	Photon36_-	IsoVL
	CaloIdL_-	CaloIdL_-	CaloIdL_-	
	IsoVL_-	IsoVL_-	IsoVL_-	
	Photon22_-	Photon22_-	Photon22_-	Photon36_-
	CaloIdL_-	CaloIdL_-	CaloIdL_-	CaloIdL_-
	IsoVL	IsoVL	IsoVL	R9Id
				Photon36_-
				R9Id_-
				Photon22_-
				CaloIdL_-
				IsoVL
				Photon36_-
				R9Id_-
				Photon22_-
				R9Id

Each piece of the HLT path name is defined as follows.

- “Photon”: Energy deposit in the ECAL that fired an L1 trigger (cf. Sec. ??).

For Photon26\_IsoVL\_Photon18, the L1 seed  $E_T$  threshold is 12 GeV, while for all other triggers in Table 3.3 it is 20 GeV.

- Integer following the word “Photon”:  $E_T$  threshold in GeV for offline reconstructed photon, using the full photon reconstruction of Sec. 3.1.1 minus the laser calibrations and assuming the primary vertex at (0, 0, 0).
- “CaloIdL”: For EB photons,  $H/E < 0.15$  and  $\sigma_{i\eta i\eta} < 0.014$ .
- “IsoVL”:  $I_{\text{ECAL}} < 0.012E_T + 6$  GeV,  $I_{\text{HCAL}} < 0.005E_T + 4$  GeV, and  $I_{\text{track}} < 0.002E_T + 4$  GeV.
- “R9Id”:  $R9 > 0.8$ .

In addition, the versions of HLT\_Photon26\_IsoVL\_Photon18 and Photon36\_CaloIdL\_Photon22\_CaloIdL that were active during runs 160404-163268 included a cut  $E_{\max}/E_{5\times 5} < 0.98$  for spike rejection.  $E_{\max}$  is the energy in the highest energy crystal of the EM cluster and  $E_{5\times 5}$  is the energy in the  $5\times 5$  crystal matrix around the seed crystal. For runs after 163268, Swiss cross spike rejection of individual crystals from HLT quantities was performed (cf. Sec. 3.1.1). All information about the evolution of the CMS HLT settings can be found in the HLT configuration browser at <http://j2eeps.cern.ch/cms-project-confdb-hltdev/browser/>.

As an example of the naming convention just described, the HLT path Photon36\_CaloIdL\_IsoVL\_Photon22\_R9Id is fired if one photon is found with  $E_T > 36$  GeV passing the CaloIdL and IsoVL requirements, and another is found with  $E_T > 22$  GeV passing the R9Id requirement.

**Add a discussion of trigger efficiency here, with plots.**

### 3.3 Event Quality

To suppress instrumental backgrounds, a set of event quality cuts are applied to the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  samples. First, all events are required to pass a good run selection, as determined by the CMS Physics Validation Team (<https://twiki.cern.ch/twiki/bin/view/CMS/PVTMain>, CERN computing ID needed). The good run selection excludes luminosity sections during which a sufficient part of the CMS detector was unpowered or malfunctioning. Such conditions could occur if, for example, a high voltage supply trips off in the middle of a run, or a DAQ error corrupts data quality but is not spotted until after the data have been collected. The severity of a detector problem is judged by its effect on a wide range of analyses and reconstruction algorithms. Of the **N**  $\text{fb}^{-1}$  of integrated luminosity delivered by the LHC in 2011, 4.68  $\text{fb}^{-1}$  passed the good run selection. This analysis is performed on the entire 2011 certified dataset.

Second, all events must contain at least one good interaction vertex. The criteria for a good vertex are:

- $\chi^2 \neq 0 \parallel \text{ndof} \neq 0 \parallel N_{\text{tracks}} \neq 0$ , where  $\chi^2$  and  $\text{ndof}$  are calculated for the track fit to the vertex, and  $N_{\text{tracks}}$  is the number of tracks in the vertex fit
- $\text{ndof} > 4$
- $|z| < 24 \text{ cm}$ , where  $z$  is the  $z$ -coordinate of the vertex position
- $|\rho| < 2 \text{ cm}$ , where  $\rho$  is the transverse displacement of the vertex position from the beam line

The good vertex requirement eliminates non-collision backgrounds such as beam scraping, beam halo, cosmic muon interactions, and instrumental effects.

Third, the two electromagnetic objects in the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events must be separated in  $\phi$  by at least 0.05. This requirement protects against beam halo

bremsstrahlung, in which a halo muon traveling parallel to the beam line radiates an energetic photon while itself depositing a large amount of energy in the ECAL. In this case, the two ECAL hits would likely be at the same  $\phi$  (and  $\rho$ ). An example cartoon is shown in Figure ???. **Insert figure.**

Fourth, the two EM objects must be separated in  $R$  by at least 0.6. Since the isolation cone size used is 0.3, this ensures that the isolation energy of one EM object cannot be in the veto strip (Fig. 3.5) of the other.

Finally, the  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events must pass an HCAL noise filter and ECAL dead channel filter. The HCAL noise filter guarantees that all HCAL reconstructed hits are inconsistent with any noise source. Noise sources [50] include:

- Ion feedback in the HPDs absent any true incident photons, in which a thermal electron ionizes a molecule in the HPD acceleration gap, faking a real signal
- HPD discharge affecting nearly all channels in the same HPD [51], partially explained by the effect of the 4 T CMS magnetic field on the flashover voltage of the dielectric [53]
- Concurrent signals in nearly all 72 channels of a single RBX, as yet unexplained
- HF PMT window hits (as opposed to the usual quartz fiber hits)
- ADC saturation

Since HCAL noise may induce fake jets or  $E_T$ , events are rejected if any of the following criteria are true:

- Any HPD has  $> 17$  hits
- A single HPD has  $> 10$  hits, but every other HPD has zero hits
- An RBX has  $> 10$  zero-ADC-count hits

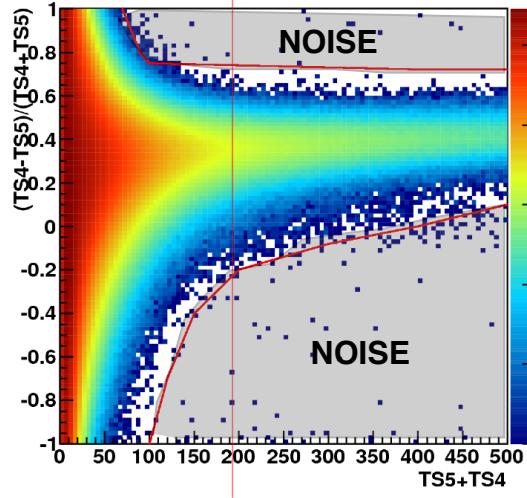


Figure 3.11:  $(TS4 - TS5)/(TS4 + TS5)$  vs.  $TS4 + TS5$  for a minimum bias sample. HB/HE hits are considered noisy if they lie in the sparsely populated gray region labeled "NOISE" defined by the curved red lines. Adapted from ref. [52].

- Any HB/HE reconstructed hit corresponding to an RBX with  $> 50$  GeV of energy fails a two-dimensional cut defined by the variables  $(TS4 - TS5)/(TS4 + TS5)$  vs.  $TS4 + TS5$ , where  $TS4(TS5)$  is the hit amplitude in the fourth(fifth) time sample read out for that hit. The cut is defined in Fig. 3.11.

The ECAL dead channel filter is designed to flag events in which significant EM energy was deposited in a masked region of the ECAL by using the trigger primitive information for the corresponding trigger tower. Energy deposited in a masked region of ECAL can cause fake  $\cancel{E}_T$ . Events are rejected if the trigger primitive  $E_T$  exceeds the maximum value of 63.75 GeV in any trigger tower that is masked in the readout.

### 3.4 Photon Identification Efficiency

In order to determine the cross section (or cross section upper limit) for a GGM signal, the photon identification efficiency is needed. Since no suitably large sample of  $Z \rightarrow \mu\mu\gamma$  events in CMS exists yet, the efficiency calculation relies on the similarity between detector response to electrons and photons. A scale factor to correct the MC

photon ID efficiency to the real photon efficiency for the data is obtained from the ratio of the electron efficiency from the data to the electron efficiency from MC.

The different types of photon ID variables—calorimeter and track isolation, ratio of hadronic to electromagnetic energy of the shower, and transverse shower shape—are chosen so that their distributions for isolated electrons and photons are similar.<sup>3</sup> Figure ?? shows distributions of photon ID variables for MC electrons and photons, where a MC electron(photon) is a `reco::Photon` object matched to a generated electron(photon) originating from a  $Z$  decay(the hard interaction) within  $\Delta R = 0.3$ . MC electrons come from the DYJets<sub>LL</sub> sample, while MC photons come from the GJet sample (see Appendix A). **Make these plots.** The shapes of the distributions agree well. **Also include data/MC comparisons?**

The photon selection efficiency is

$$\epsilon_\gamma = \epsilon_\gamma^{\text{MC}} \times \frac{\epsilon_e^{\text{data}}}{\epsilon_e^{\text{MC}}} \quad (3.5)$$

where

- $\epsilon_\gamma$  is the photon ID efficiency in data,
- $\epsilon_\gamma^{\text{MC}}$  is the photon ID efficiency in MC,
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in the data that satisfy the photon ID cuts, and
- $\epsilon_e^{\text{data}}$  is the electron ID efficiency obtained using  $Z \rightarrow ee$  electrons in MC that satisfy the photon ID cuts.

The ratio  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  is defined as the scale factor by which the GGM signal MC photon ID efficiency must be multiplied to give an estimate of the photon ID efficiency

---

<sup>3</sup> $R9$  differs between photons and radiating electrons, but the requirement  $R9 < 1$  is loose enough not to introduce problems with the use of electrons to measure the photon ID efficiency.

in data. The photon ID requirements of Table 3.1 plus the IsoVL HLT requirement described in Sec. 3.2 and Table 3.3 are repeated in Table 3.4.

Table 3.4: Candidate photon ID requirements.

Variable	Cut
$I_{\text{ECAL}}$	$< 0.012E_T + 6 \text{ GeV}$
$I_{\text{HCAL}}$	$< 0.005E_T + 4 \text{ GeV}$
$I_{\text{track}}$	$< 0.002E_T + 4 \text{ GeV}$
$H/E$	$< 0.05$
$\sigma_{i\eta i\eta}$	$< 0.011$
$I_{\text{ECAL}} - 0.0792\rho + I_{\text{HCAL}} - 0.0252\rho + I_{\text{track}}$	$< 6 \text{ GeV}$
$R9$	$< 1$

### 3.4.1 Tag and Probe Method

A *Z tag and probe* method is utilized to measure the efficiency of the photon ID cuts in Table 3.1. The tag is a well-identified electron. The probe, by contrast, is as loosely identified as possible, and all tags must pass the probe criteria in addition to the stringent tag criteria. The tag and probe criteria used in this study are shown in Table 3.5.

The invariant mass of the tag and probe are required to be within a narrow window around  $Z$  mass. Assuming that the probabilities of the tag and probe legs of the  $Z$  decay to pass the photon ID cuts are uncorrelated, the efficiency can be estimated as

$$\epsilon = \frac{N_{\text{tag-pass}}}{N_{\text{tag-pass}} + N_{\text{tag-fail}}} \quad (3.6)$$

where  $N_{\text{tag-pass}}$  is the number of tag-probe pairs in which the probe leg passes the photon ID cuts under study and  $N_{\text{tag-fail}}$  is the number of tag-probe pairs in which the probe leg fails the cuts. Implicit in these definitions is a double counting of pairs in which both electrons pass the tag and probe criteria [54]. In addition, in the rare

Table 3.5: Tag and probe criteria. The superscript 0.4 indicates that the isolation variable was calculated in a cone of  $\Delta R = 0.4$  around the photon candidate. The isolations without superscripts use the standard  $\Delta R = 0.3$  cones.

Variable	Cut	
	Tag	Probe
RECO object	photon	photon
HLT	HLT_Ele17_CaloIdVT_CaloIsoVT_TrkIdT_-TrkIsoVT_SC8_Mass30_v* (must have fired the 17 GeV leg)	—
$H/E$	$< 0.05$	$< 0.15$
$I_{\text{ECAL}}^{0.4}$	$< 0.006E_T + 4.2 \text{ GeV}$	—
$I_{\text{HCAL}}^{0.4}$	$< 0.0025E_T + 2.2 \text{ GeV}$	—
$I_{\text{track}}^{0.4}$	$< 0.001E_T + 2.0 \text{ GeV}$	—
$E_T$	$> 25 \text{ GeV}$	—
SC $E_T$	—	$> 15 \text{ GeV}$
SC $ \eta $	$< 1.4442$	$< 1.4442$
$\sigma_{in\eta}$	$< 0.009$	—
Has pixel seed?	Yes	—
Track match type	General track	—
Track match $\Delta R$	$< 0.04$	—
Track match $p_T$	$> 15 \text{ GeV}$	—
Track match $ \eta $	$< 1.479$	—

circumstance (less than 1% in MC [54]) that two or more probes may be matched to one tag, the pair with invariant mass closest to the  $Z$  mass is chosen.

In practice,  $N_{\text{tag-pass}}$  and  $N_{\text{tag-fail}}$  are returned by a simultaneous unbinned maximum likelihood fit to the invariant mass distributions of tag-pass and tag-fail events, with appropriate signal and background PDF assumptions. The fit form used is

$$\begin{aligned} f_{\text{tag-pass}}(m_{\text{tag-pass}}) &= \epsilon N_S f_S^{\text{pass}}(m_{\text{tag-pass}}) + N_B^{\text{pass}} f_B^{\text{pass}}(m_{\text{tag-pass}}) \\ f_{\text{tag-fail}}(m_{\text{tag-fail}}) &= (1 - \epsilon) N_S f_S^{\text{fail}}(m_{\text{tag-fail}}) + N_B^{\text{fail}} f_B^{\text{fail}}(m_{\text{tag-fail}}) \end{aligned} \quad (3.7)$$

where  $f_{\text{tag-pass}}(m_{\text{tag-pass}})$  and  $f_{\text{tag-fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail PDFs, respectively;  $\epsilon$  is the efficiency;  $N_S$  is the total number of  $Z$  signal events summed over both samples;  $f_S^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_S^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail signal PDFs, respectively;  $N_B^{\text{pass}}$  and  $N_B^{\text{fail}}$  are the numbers of background events in the tag-pass and tag-fail samples, respectively; and  $f_B^{\text{pass}}(m_{\text{tag-pass}})$  and  $f_B^{\text{fail}}(m_{\text{tag-fail}})$  are the tag-pass and tag-fail background PDFs, respectively. This particular implementation of the tag and probe methodology is based on tag `CMSSW_4_2_5` of the CMSSW package `PhysicsTools/TagAndProbe`, and uses the MINUIT2 [55] library, as coded in RooFit [56], for the likelihood maximization. For this study, CMSSWv4.2.8 was used.

For both samples, the signal shape is assumed to be a Crystal Ball function [57] convoluted with the  $Z$  generated lineshape, while the background shape is a PDF that describes the falling background as well as the kinematic turn-on at low invariant mass. The background PDF, called “RooCMSShape” [54], is given by

$$\text{RooCMSShape}(x) = \begin{cases} 1e20 & \text{for } (x - \mu)\gamma < -70 \\ 0 & \text{for } (x - \mu)\gamma > 70 \\ \text{erfc}((\alpha - x)\beta) \exp(-(x - \mu)\gamma) & \text{otherwise} \end{cases} \quad (3.8)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\mu$  are parameters of the fit, most of which are held fixed. In the three lowest  $E_T$  bins, all parameters of the tag-pass and tag-fail background PDFs are left floating, because the effect of the relaxed  $E_T$  cuts has a significant effect on the background shape. More details of the signal and background PDFs are given in Table 3.6. The fixed signal and background parameter values were determined by fitting a small sample ( $0.0 \leq \eta < 0.25$ ) of Fall11 MC signal (DYJetsToLL) and background (QCD\_Pt-20to30\_BCtoE, QCD\_Pt-30to80\_BCtoE, QCD\_Pt-80to170\_BCtoE, GJet\_Pt-20\_doubleEMEnriched, WJetsToLNu, TTJets) with parameters left floating.

Some fit examples are shown in Figures ?? and ???. In Fig. ??, which shows fits to data and MC for  $15 \text{ GeV} \leq \text{probe } E_T \leq 20 \text{ GeV}$ , the kinematic turn-on is below the invariant mass range covered by the plot. The exponentially falling background is easily seen underneath the signal, and is especially pronounced in the background-dominated tag-fail sample. **Include these plots.**

### 3.4.2 Photon Efficiency Scale Factor $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$

Figure 3.12 shows the dependence of the photon ID efficiency scale factor  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}}$  on  $E_T$ ,  $\eta$ ,  $\Delta R_{\gamma\text{jet}}$ , and  $N_{\text{jet}}$ , where jets are defined as in Sec. 3.5.2 (**insert reference to “Jet Selection” section here**).  **$\Delta R_{\gamma\text{jet}}$  plot will be added during second round.** Errors are statistical only. For the plot of the scale factor vs.  $\Delta R_{\gamma\text{jet}}$ , only events with  $\geq 1$  jets are included. There no significant dependence of the scale factor on these variables, so only one scale factor is computed from the entire dataset.

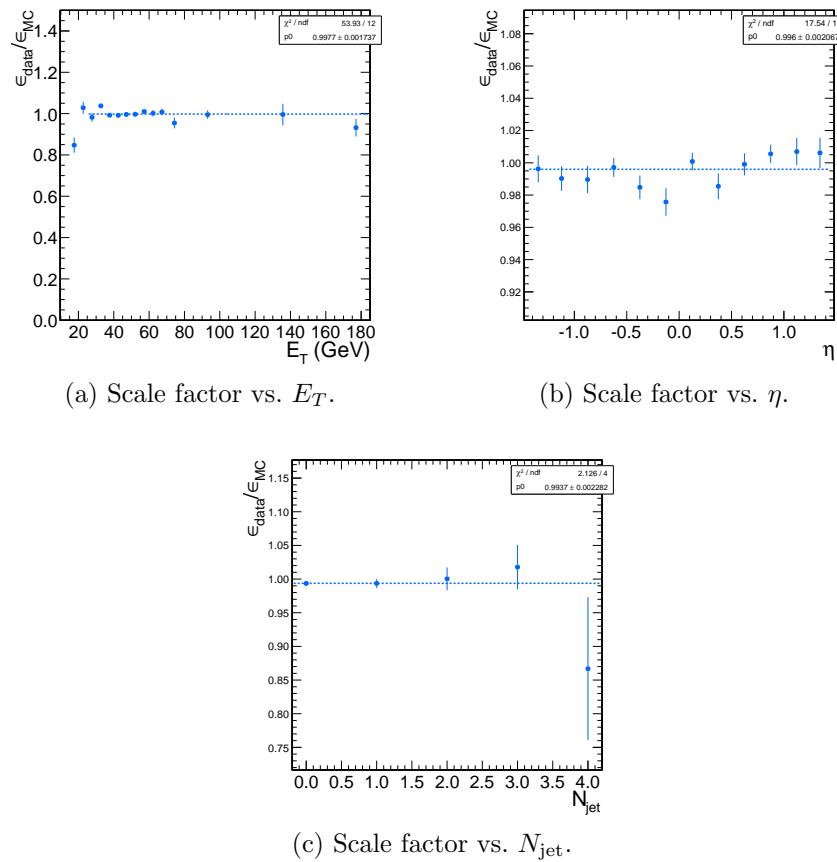


Figure 3.12: Dependence of the photon ID efficiency scale factor on some kinematic variables. Errors are statistical only.

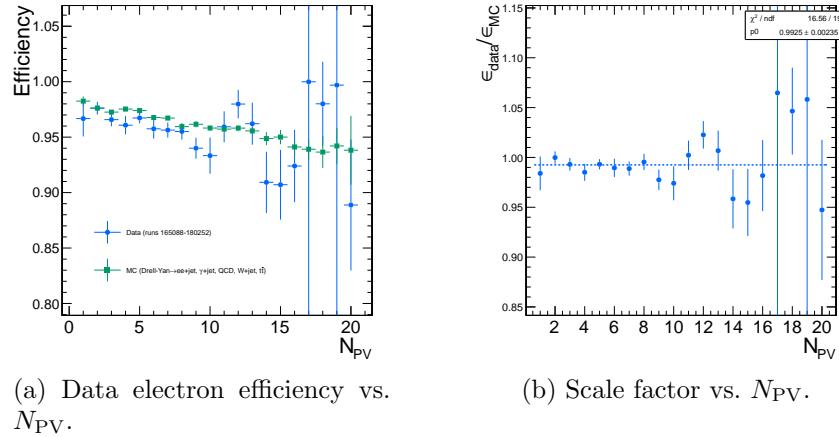


Figure 3.13: Dependence of the photon ID efficiency scale factor on the number of primary vertices per event. Errors are statistical only.

The effect of pileup is studied by comparing the efficiencies  $\epsilon_e^{\text{data}}$  and  $\epsilon_e^{\text{MC}}$  vs. the number of primary vertices ( $N_{PV}$ ) in the event. The efficiency drops for events with large  $N_{PV}$ , even after using pileup-corrected isolation cuts, as can be seen in Figure 3.13a. However, the MC tracks the data, and the scale factor is flat vs.  $N_{PV}$ , as shown in Fig. 3.13b.

The scale factor is measured to be  $\epsilon_e^{\text{data}}/\epsilon_e^{\text{MC}} = 0.994 \pm 0.002(\text{stat.}) \pm ?(\text{syst.})$ . Four main sources of systematic error, in addition to the statistical error of 0.2%, were studied.

**Different behavior of electrons and photons in MC** Even though the photon ID cuts are designed to be similarly efficient for both electrons and photons, there might be a small difference in the performance between the two kinds of particles, e.g. because of electron bremsstrahlung. To check this effect, the MC electron ID efficiency was calculated using a  $Z \rightarrow ee$  sample and the MC photon ID efficiency was calculated using a  $\gamma + \text{jets}$  sample. Both samples were reconstructed in CMSSWv3.6. Half the difference between these two results, 0.5%, was taken as an error on the scale factor. **This result is now out of date because it used an older selection. Should this check be redone?**

**Pileup** To account for the possibility that the MC simulation may not adequately reproduce the data in a high pileup environment, the data/MC scale factor for events with 1-4 good reconstructed primary vertices was calculated, along with the same for events with  $\geq 7$  good reconstructed primary vertices. This particular division of the data was chosen because the pileup distribution in data peaks at  $\sim 7$  primary vertices per event. The difference between the scale factors from both samples, ?, was taken as an error on the scale factor from pileup.

**Signal fit over/underestimation** It was found that the signal fit slightly underestimates the data in the tag-pass sample, and slightly overestimates it in the tag-fail sample. To cover this effect with a systematic error, the efficiencies in data and MC, and then the scale factor, were recalculated using the background (from fit) subtracted integrals of the tag-pass and tag-fail distributions, rather than the fitted signal yields in those distributions. The difference between the scale factor found in this way and the nominal scale factor, ?, was taken as an error on the scale factor. **Not sure this is a problem anymore.**

**Signal and background shape assumption** The largest source of systematic error comes from the signal and background shape assumptions. To assess the magnitude of this error, the tag-pass and tag-fail tail parameters (Crystal Ball ? and n) were varied by  $\pm 1\sigma$ , and the background shape was varied between “RooCMSShape, exponential, power law, and quadratic. All possible combinations of varied parameters were generated, and the data and MC were refit and new scale factors generated according to those combinations. The error was taken as the largest deviation of the scale factor from nominal, ?%. **Also, in many bins, there is zero background fitted under the tag-pass peak. Background is a small effect here anyway, but we should get a sys-**

tematic for this somehow.

Add in a MC estimate of the pixel veto efficiency with varied tracker geometries.

Table 3.6: Parameter values for the signal and background PDFs for the different samples. The background PDF applies to all efficiency bins except the four lowest  $E_T$  bins, which use a floating RooCMSShape background. When a bracketed range is given, the parameter is allowed to float within that range. When a constant is given, the parameter is fixed to that constant.

PDF	Crystal Ball fit parameters				RooCMSShape fit parameters			
	$\mu$	$\sigma$	$\alpha$	n	$\mu$	$\alpha$	$\beta$	$\gamma$
Tag-pass signal	[-1.0, 1.0]	[1.0, 3.0]	0.87	97.0	N/A	N/A	N/A	N/A
Tag-fail signal	[-1.0, 1.0]	[1.0, 3.0]	0.73	134.9	N/A	N/A	N/A	N/A
Tag-pass background	N/A	N/A	N/A	N/A	65.0	61.949	0.04750	0.01908
Tag-fail background	N/A	N/A	N/A	N/A	$\alpha$	[50.0, 100.0]	0.065	0.048

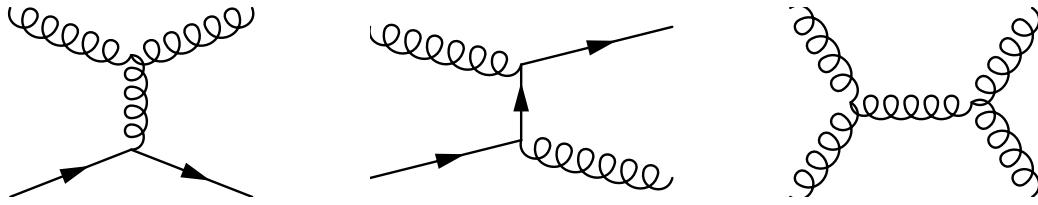
# Chapter 4

## Data Analysis

The signature of GGM SUSY particle production in this search is an excess of two-photon events with high  $\cancel{E}_T$ .  $\cancel{E}_T$  is reconstructed using the particle flow algorithm as described in Sec. 3.1.3. Candidate two-photon events, as well as control events, are selected according to the offline object criteria presented in Secs. 3.1.1, 3.1.2, and 3.1.3; the event quality criteria in Sec. 3.3; and the trigger requirements in Sec. 3.2. These are summarized in Table 4.1.

Table 4.1: Selection criteria for  $\gamma\gamma$ ,  $e\gamma$ ,  $ee$ , and  $ff$  events.

Variable	Cut			
	$\gamma\gamma$	$e\gamma$	$ee$	$ff$
HLT match	IsoVL	IsoVL	IsoVL	IsoVL    R9Id
$E_T$	$> 40/ > 25$ GeV			
$SC  \eta $	$< 1.4442$	$< 1.4442$	$< 1.4442$	$< 1.4442$
$H/E$	$< 0.05$	$< 0.05$	$< 0.05$	$< 0.05$
$R9$	$< 1$	$< 1$	$< 1$	$< 1$
Pixel seed	No/No	Yes/No	Yes/Yes	No/No
$I_{\text{comb}}, \sigma_{inj\eta}$	$< 6$ GeV $\&\&$ $< 0.011$	$< 6$ GeV $\&\&$ $< 0.011$	$< 6$ GeV $\&\&$ $< 0.011$	$< 20$ GeV $\&\&$ $(\geq 6$ GeV $\parallel$ $\geq 0.011)$
JSON	Yes	Yes	Yes	Yes
No. good PVs	$\geq 1$	$\geq 1$	$\geq 1$	$\geq 1$
$\Delta R_{\text{EM}}$	$> 0.6$	$> 0.6$	$> 0.6$	$> 0.6$
$\Delta\phi_{\text{EM}}$	$\geq 0.05$	$\geq 0.05$	$\geq 0.05$	$\geq 0.05$



This search utilizes  $4.7 \text{ fb}^{-1}$  of CMS data collected during the period April–December 2011, corresponding to the following datasets [58]:

- `/Photon/Run2011A-05Jul2011ReReco-ECAL-v1/AOD`
- `/Photon/Run2011A-05Aug2011-v1/AOD`
- `/Photon/Run2011A-030ct2011-v1/AOD`
- `/Photon/Run2011B-PromptReco-v1/AOD`

The search strategy is to model the backgrounds to the GGM SUSY signal using  $\cancel{E}_T$  shape templates derived from the control samples, and then to look for a high- $\cancel{E}_T$  excess above the estimated background in the  $\gamma\gamma$  sample. There are two categories of backgrounds: QCD processes with no real  $\cancel{E}_T$  and electroweak processes with real  $\cancel{E}_T$  from neutrinos. The relevant QCD background processes are multijet production with at least two jets faking photons, photon + jet production with at least one jet faking a photon, diphoton production, and  $Z$  production with a radiated photon where at least one of the  $Z$  decay products (typically a jet) fakes a photon. The relevant electroweak background processes, which are small compared to the QCD background, involve  $W \rightarrow e\nu$  decay with a recoiling jet that fakes a photon or a real radiated photon (the  $W$  may come from the decay of a top quark in  $t\bar{t}$  events). In both cases, the electron is misidentified as a photon due to a small inefficiency in reconstructing the electron pixel seed. The main diagrams contributing to the QCD(electroweak) backgrounds are shown in Figure ??(??). **Generate these Feynman diagrams.**

Figure ?? shows the  $\cancel{E}_T$  spectrum of the  $\gamma\gamma$  search data sample overlaid on the  $\cancel{E}_T$  spectra of MC simulated background components. The MC spectra are normalized to the integrated luminosity of the  $\gamma\gamma$  data sample. **Make this plot.** The dominant background components are QCD inclusive photon processes. The MC is not used in the actual background estimation. It is just shown here to illustrate the breakdown of backgrounds.

Data control samples are used to model all of the backgrounds. The primary control sample used to model the QCD background is the  $ff$  sample, which is similar to the candidate  $\gamma\gamma$  sample but with combined isolation or  $\sigma_{inj\eta}$  cuts inverted. The cuts on these variables are used to distinguish between photons and jets, so by inverting those cuts, the resulting  $ff$  sample becomes enriched with QCD dijets. Because the fake photons are still required to pass a tight cut on  $H/E$ , they are guaranteed to be very electromagnetic jets, with an EM energy scale and resolution similar to that of the candidate photons. This insures that the resulting estimate of the  $\cancel{E}_T$  shape does not have too long of a tail from severe HCAL mis-measurements that are actually rare in the  $\gamma\gamma$  sample, as shown in Figure ??.

**Plot the  $\gamma\gamma/ff$   $\cancel{E}_T$  agreement for different values of the  $ff$   $H/E$  cut in MC. Make the same plot in data for a restricted  $\cancel{E}_T$  range?**

As a cross-check, the  $ee$  sample is also used to model the QCD background. This sample of  $Z$  decays should have no true  $\cancel{E}_T$ , just like the  $ff$  sample, and the electron definition (differing from the photon definition only in the presence of a pixel seed) insures that the electron energy scale and resolution is similar to that of the photon.

Finally, the  $e\gamma$  sample is used to model the electroweak background from  $W \rightarrow e\nu$  decays. The  $e\gamma$   $\cancel{E}_T$  distribution is scaled by the electron $\rightarrow$ photon misidentification rate to predict the number of  $W\gamma$ ,  $W +$  jet, and  $t\bar{t}$  events in the  $\gamma\gamma$  sample.

The remainder of this chapter describes the data analysis procedures and the final results of the search. Sec. 4.1 addresses the QCD background estimation. Sec. ??

addresses the electroweak background estimation. The chapter concludes with a discussion of systematic errors in Sec. ?? and a presentation of the final results in Sec. 4.4.

## 4.1 Modeling the QCD Background

### 4.1.1 Outline of the Procedure

Due to the fact that the CMS ECAL energy resolution is much better than the HCAL energy resolution, the energies of the two candidate photons in the events of the  $\gamma\gamma$  sample are typically measured to far greater accuracy and precision than the energy of the hadronic recoil in those events. Therefore, fake  $E_T$  in the  $\gamma\gamma$  sample is almost entirely the result of hadronic mis-measurement in QCD dijet, photon + jet, and diphoton events. The strategy employed to model this background is to find a control sample in data consisting of two well-measured EM objects, just like the candidate  $\gamma\gamma$  sample, and assign each event a weight to account for the underlying kinematic differences between the control and candidate samples. Once the reweighted  $E_T$  spectrum of the control sample is created, it is then normalized in the low- $E_T$  region, the assumption being that GGM SUSY does not predict a significant amount of events at low  $E_T$ . There are three aspects to this QCD background estimation procedure that bear highlighting:

**Choice of control sample** Since the underlying cause of  $E_T$  in the candidate sample is mis-measured hadronic activity, a control sample with similar hadronic activity to the candidate sample should be chosen. Hadronic activity refers to number of jets, jet  $E_T$ , pileup, etc.

**Reweighting** The control sample is reweighted so that its  $E_T$  spectrum appears as it would if the control sample had the same kinematic properties as the candidate

sample (i.e. particle  $p_T$  and  $\eta$  distributions, etc.). By choosing an appropriate control sample and reweighting it, the control  $\cancel{E}_T$  distribution should now match both the hadronic activity and the kinematics of the candidate sample.

**Normalization** Finally, the control  $E_T$  distribution is normalized in a region of low  $\cancel{E}_T$ , where contamination from the expected GGM SUSY signal is small. This implies an extrapolation of the low- $\cancel{E}_T$  QCD background prediction to the high- $\cancel{E}_T$  signal region.

As explained in the beginning of this chapter, the  $ff$  sample is used as the primary QCD control sample, while the  $ee$  sample is used as a cross-check. Both samples have two well-measured EM objects per event, no real  $\cancel{E}_T$ , and similar hadronic activity to the  $\gamma\gamma$  sample. Figure 4.1 shows a comparison of the shapes of some distributions relevant to hadronic activity between the  $\gamma\gamma$ ,  $ee$ , and  $ff$  samples. In general, the  $ee$  sample has less hadronic activity than the  $\gamma\gamma$  and  $ff$  samples, as shown by the more steeply falling  $ee$  distributions in Figs. 4.1a, 4.1b, 4.1c, and 4.1d. In addition to the kinematic reweighting, there is also a reweighting by number of jets per event, which attempts to correct for this difference (see Sec. 4.1.2).

### 4.1.2 Reweighting

To reweight the control sample events to match the kinematics of the candidate sample events, a weight based on the  $p_T$  of the di-EM-object system and the number of jets in the event is used. As explained in Sec. 4.1.1,  $E_T$  in the  $\gamma\gamma$ ,  $ff$ , and  $ee$  samples is due to the poorly measured hadronic recoil off the well-measured di-EM system. Therefore, the  $p_T$  of the di-EM system is a good handle on the true magnitude of the hadronic recoil, which affects the measured  $\cancel{E}_T$ . The di-EM system is depicted in Figure 4.2. As shown in Figure 4.3,  $\cancel{E}_T$  is largely uncorrelated with di-EM  $p_T$ , so there is little danger of reweighting away a true signal excess.

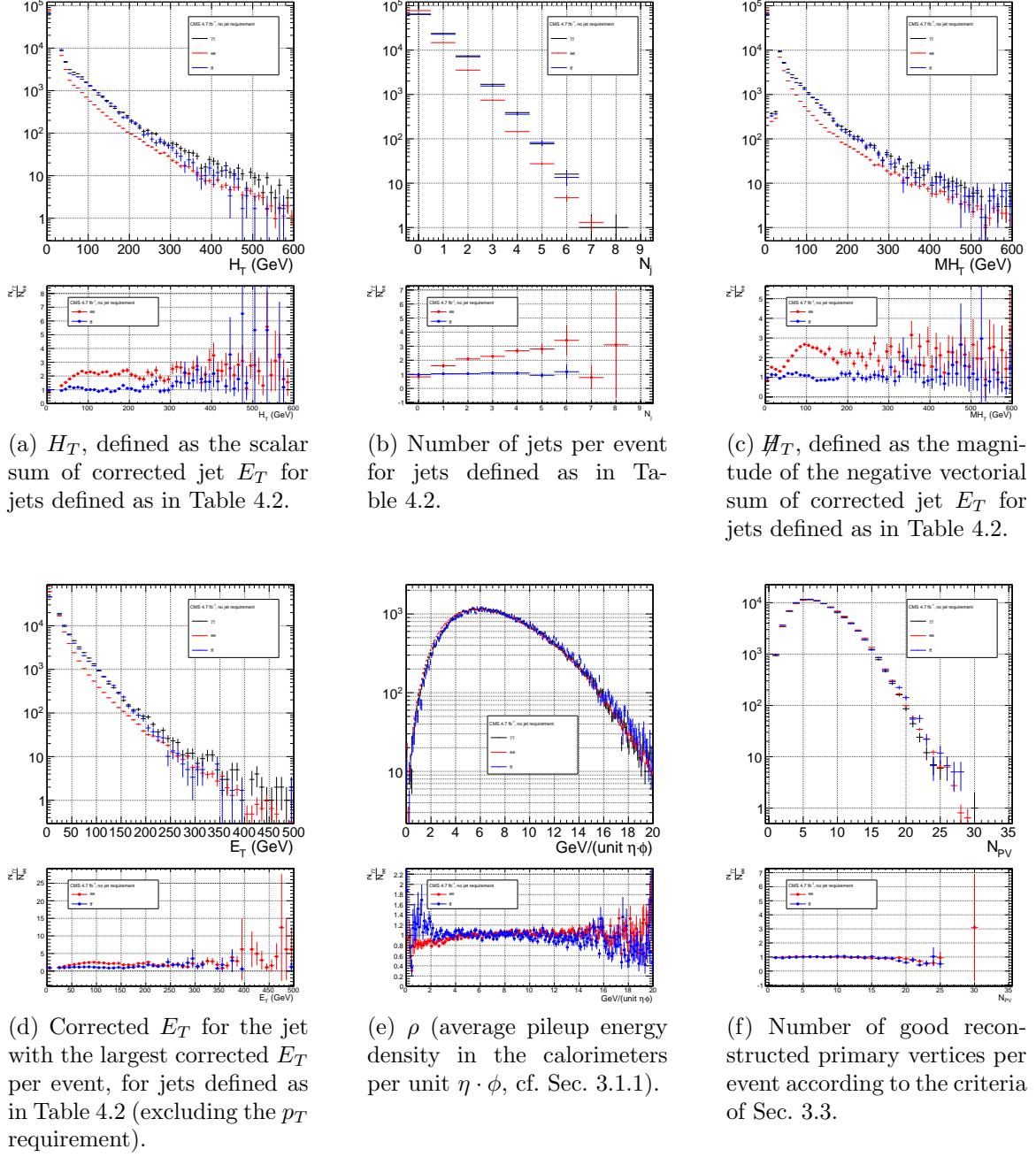


Figure 4.1: Comparison of the shapes of some distributions relevant to hadronic activity between the  $\gamma\gamma$ ,  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ), and  $ff$  samples. The  $ee$  and  $ff$  distributions are normalized to the number of events in the  $\gamma\gamma$  distribution. Errors are statistical only.

Table 4.2: Definition of HB/HE/HF hadronic jets. **Add a footnote describing the PF electron and PF muon definitions, with references.**

Variable	Cut
Algorithm	L1FastL2L3Residual corrected PF (cf. Sec. 3.1.3)
$p_T$	$> 30 \text{ GeV}$
$ \eta $	$< 5.0$
Neutral hadronic energy fraction	$< 0.99$
Neutral electromagnetic energy fraction	$< 0.99$
Number of constituents	$> 1$
Charged hadronic energy	$> 0.0 \text{ GeV}$ if $ \eta  < 2.4$
Number of charged hadrons	$> 0$ if $ \eta  < 2.4$
Charged electromagnetic energy fraction	$< 0.99$ if $ \eta  < 2.4$
$\Delta R$ to nearest electron, muon, or one of the two primary EM objects	$> 0.5$

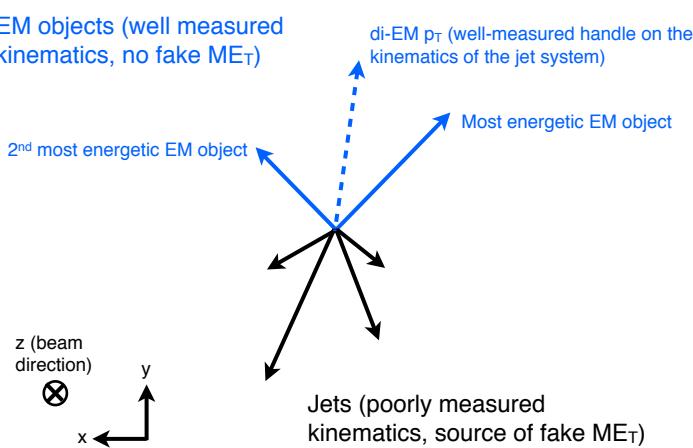


Figure 4.2: Cartoon showing the di-EM system in blue and the hadronic recoil in black. The di-EM  $p_T$  (dashed blue line) is used to reweight the control sample kinematic properties to match those of the candidate  $\gamma\gamma$  sample.

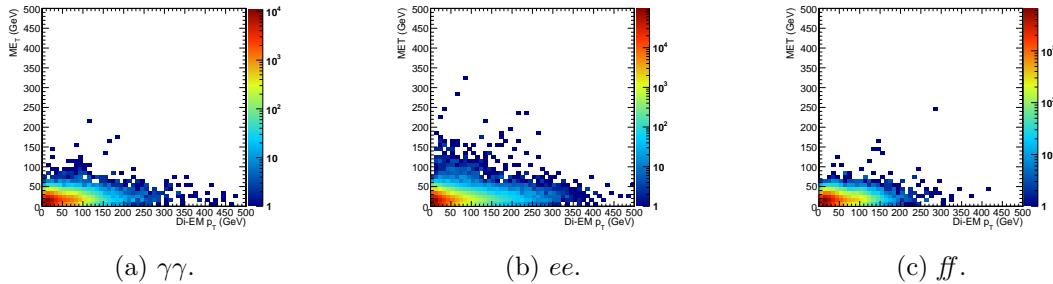


Figure 4.3:  $E_T$  vs. di-EM  $p_T$ .

Whereas the di-EM  $p_T$  reweighting accounts for differences in production kinematics between the control and  $\gamma\gamma$  samples, a simultaneous reweighting based on the number of jets in the event accounts for differences in hadronic activity between the samples, especially between  $ee$  and  $\gamma\gamma$  (cf. Fig. 4.1). Jets are defined as in Table 3.2. Figure 4.4 shows the effect of reweighting by number of jets per event, which is to increase(decrease) the tail of the  $ee(ff)$   $E_T$  spectrum.

Although the electron and photon energies are well measured by the ECAL, the ECAL-only measurement of the fake photon energy (cf. Sec 3.1.1) is biased slightly low due to the fact that fakes (which are really jets) tend to deposit some energy in the HCAL. This can be seen in Figs. 4.5 and 4.6, which show the relative difference between the ECAL-only  $E_T$  measurement and the PF  $E_T$  measurement vs. EMF for electrons, photons, and fakes. PF  $E_T$  is defined as the L1Fast-corrected  $E_T$  of the nearest PF jet with  $p_T \geq 20$  GeV (i.e., the  $E_T$  of the PF jet object reconstructed from the same ECAL shower as the fake photon). On average, the fakes tend to deposit a few percent more energy in the HCAL than the electrons or photons, which is recovered by the PF algorithm. For this reason, the PF  $p_T$  is used in the calculation of di-EM  $p_T$  rather than the ECAL-only  $p_T$ .<sup>1</sup> This leads to a modest improvement in the agreement between the  $ee$  and  $ff$   $E_T$  spectra, as shown in Figure 4.7.

The control sample event weights are defined as

<sup>1</sup>In the few events ( $\mathcal{O}(10^{-3})$ ) in which two PF jet objects, corresponding to the two electrons or fakes, are not found, the ECAL-only  $p_T$  is used.

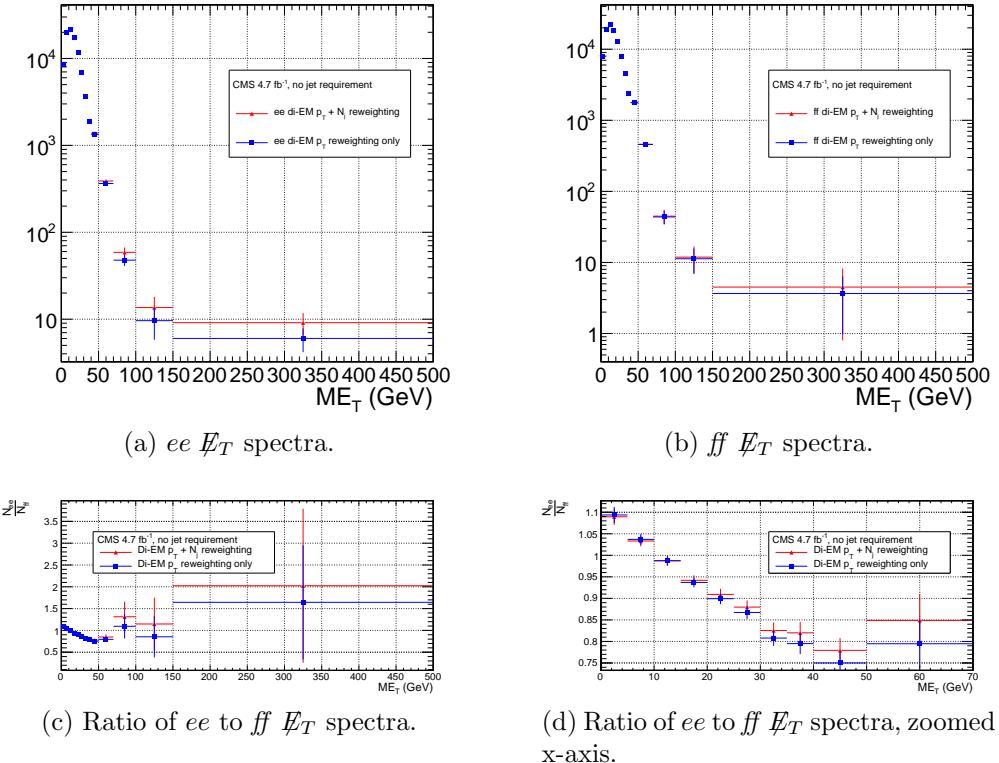


Figure 4.4:  $E_T$  spectra of the reweighted ee ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) and ff control samples. Blue squares indicate di-EM  $p_T$  reweighting only; red triangles indicate di-EM  $p_T + \text{number of jets}$  reweighting. PF  $p_T$  (cf. p. 70) is used to calculate the di-EM  $p_T$ . The full normalization procedure is employed, along with ee sideband subtraction (discussed at the end of this section). Error bars include statistical, reweighting, and normalization error (see Sec. ??).

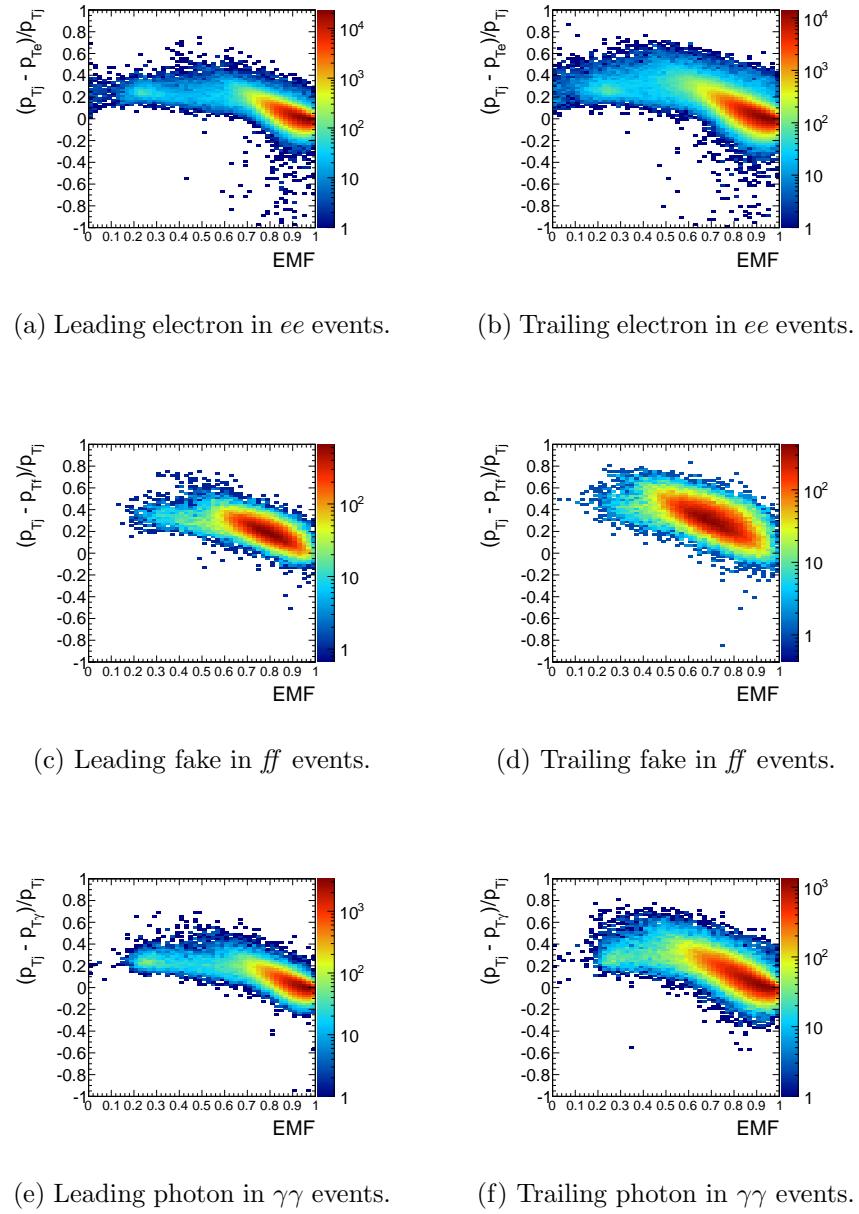


Figure 4.5: Relative difference between the ECAL-only  $E_T$  measurement and the PF  $E_T$  measurement vs. EMF. PF  $E_T$  is defined in the text.

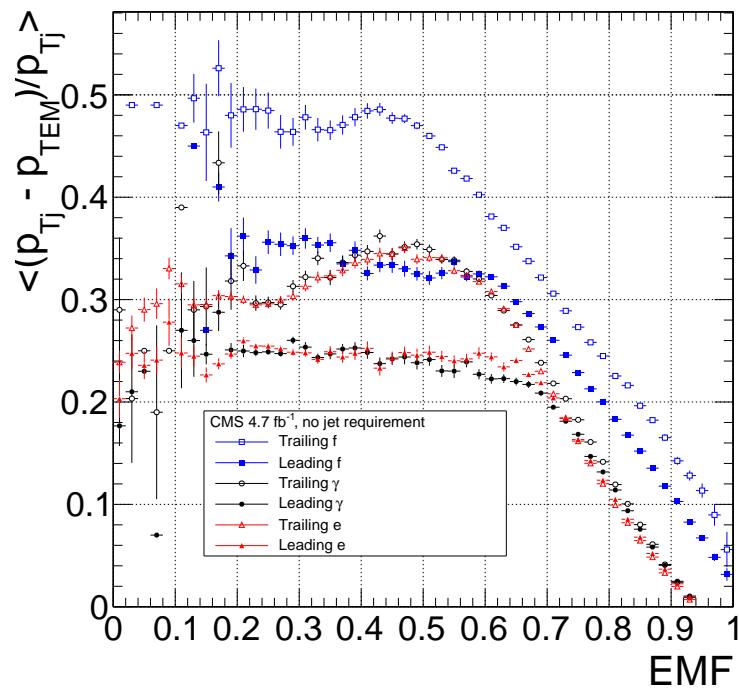


Figure 4.6: Average relative difference between the ECAL-only  $E_T$  measurement and the PF  $E_T$  measurement vs. EMF for the leading (filled marker) and trailing (open marker) electrons in  $ee$  events (red triangles), fakes in  $ff$  events (blue squares), and photons in  $\gamma\gamma$  events (black circles). These are nothing more than profile histograms of Fig. 4.5. PF  $E_T$  is defined in the text. Error bars are statistical only.

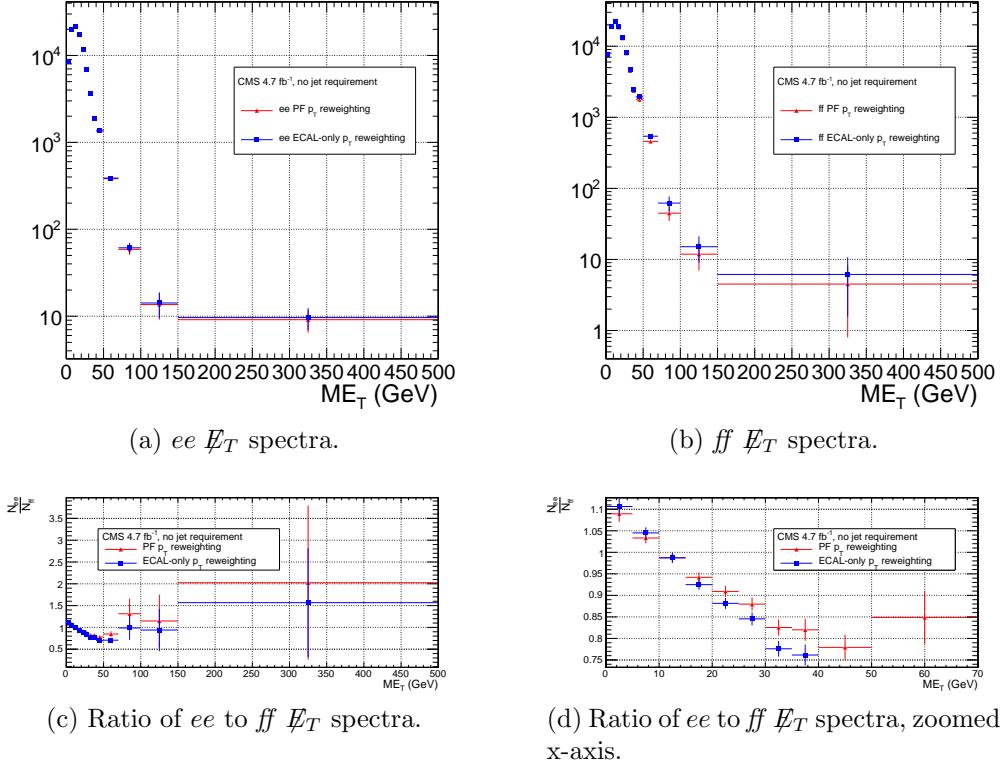


Figure 4.7:  $E_T$  spectra of the reweighted  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) and  $ff$  control samples. Blue squares indicate reweighting using the ECAL-only  $p_T$  estimate; red triangles indicate reweighting using the PF  $p_T$  estimate. The full reweighting and normalization procedure is employed, along with  $ee$  sideband subtraction (discussed at the end of this section). Error bars include statistical, reweighting, and normalization error (see Sec. ??).

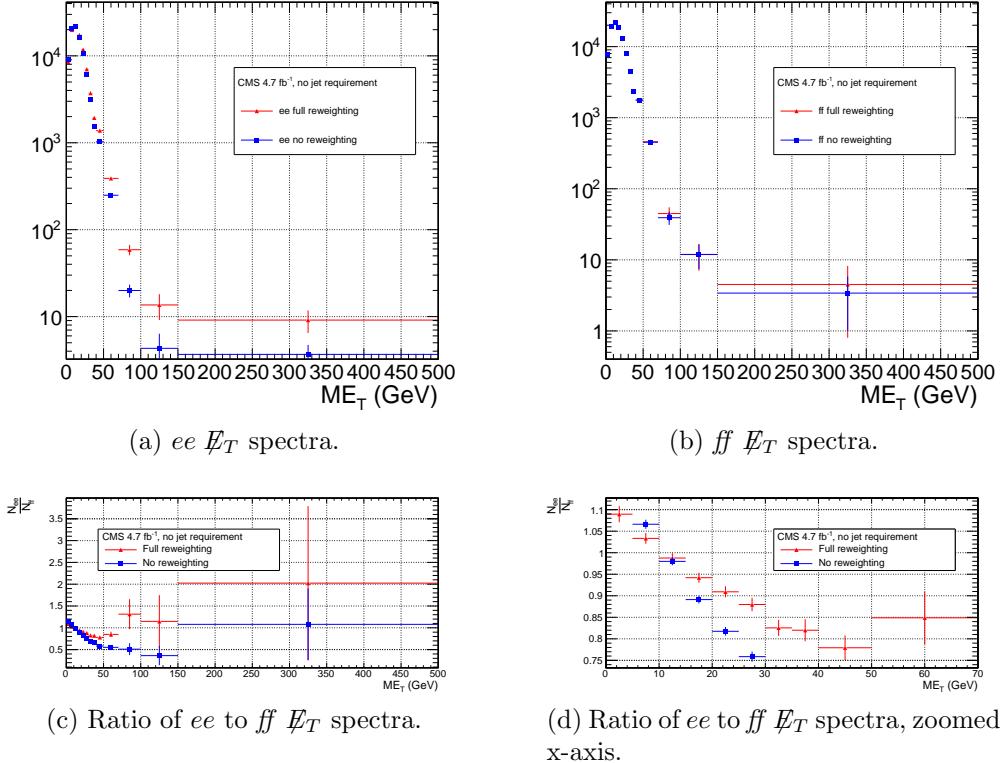


Figure 4.8:  $E_T$  spectra of the ee ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) and ff control samples. Red triangles indicate full di-EM  $p_T +$  number of jets reweighting; blue squares indicate no reweighting. PF  $p_T$  (cf. p. 70) is used to calculate the di-EM  $p_T$ . The full normalization procedure is employed, along with ee sideband subtraction (discussed at the end of this section). Error bars include statistical, reweighting (where appropriate), and normalization error (see Sec. ??).

$$w_{ij} = \frac{N_{\text{control}}}{N_{\gamma\gamma}} \frac{N_{\gamma\gamma}^{ij}}{N_{\text{control}}^{ij}} \quad (4.1)$$

where  $i$  runs over the number of di-EM  $p_T$  bins,  $j$  runs over the number of jet bins,  $N_{\text{control}}$  is the total number of events in the control sample,  $N_{\gamma\gamma}$  is the total number of events in the  $\gamma\gamma$  sample,  $N_{\gamma\gamma}^{ij}$  is the number of  $\gamma\gamma$  events in the  $i^{\text{th}}$  di-EM  $p_T$  bin and  $j^{\text{th}}$  jet bin, and  $N_{\text{control}}^{ij}$  is the number of control sample events in the  $i^{\text{th}}$  di-EM  $p_T$  bin and  $j^{\text{th}}$  jet bin. The effect of the reweighting is more significant for the ee sample than for the ff sample, as shown in Figure 4.8.

The  $ee$  sample contains a non-negligible background of  $t\bar{t}$  events in which both  $W$  bosons decay to electrons. These events have significant real  $\cancel{E}_T$  from the two neutrinos (unlike the  $\gamma\gamma$  events), and therefore inflate the background estimate at high  $\cancel{E}_T$ . In order to remove the  $t\bar{t}$  contribution from the  $ee$  sample, a sideband subtraction method is employed.

Only events in the  $ee$  sample with  $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ , where  $m_{ee}$  is the di-electron invariant mass, are used in the QCD background estimate. This choice maximizes the ratio of  $Z$  signal to background. The sidebands used to estimate the background contribution within the  $Z$  window are defined such that  $71 \text{ GeV} \leq m_{ee} < 81 \text{ GeV}$  and  $101 \text{ GeV} \leq m_{ee} < 111 \text{ GeV}$ .

The full reweighting procedure is applied to the  $Z$  signal region and the two sideband regions independently. Only  $Z$  signal events are used in the calculation of the di-EM  $p_T$  weights for the  $Z$  signal region, and likewise only the events within a given sideband region are used in the calculation of the weights for that region. Assuming a constant  $t\bar{t}$  background shape, the resulting reweighted sideband  $\cancel{E}_T$  distributions are added together and subtracted from the reweighted  $Z$  signal  $\cancel{E}_T$  distribution. The sideband subtracted  $Z$  signal  $\cancel{E}_T$  distribution is then normalized as discussed in Secs. 4.1.1 and 4.1.3. The statistical and reweighting error from the sideband regions is propagated to the error on the final  $ee$  QCD  $\cancel{E}_T$  estimate.

The di-EM  $p_T$  weights for the two  $ee$  sideband regions are shown in figure ???. The overall scale of the weights, as well as the trend with di-EM  $p_T$ , is similar for the two regions (except at high di-EM  $p_T$ , where the statistics are poor anyway). Figure 4.10 shows the  $\cancel{E}_T$  spectra for the two sideband regions and the  $Z$  signal region after subtraction. The shapes of the spectra indicate that the high- $\cancel{E}_T$   $t\bar{t}$  tail, present in the sideband distributions, was successfully subtracted from the  $Z$  signal distribution.

The  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ),  $ff$ , and  $\gamma\gamma$  di-EM  $p_T$  spectra for events with 0, 1, or  $\geq 2$  jets (as in Table 3.2) are shown in Figure 4.11. Broad humps in the  $ff$

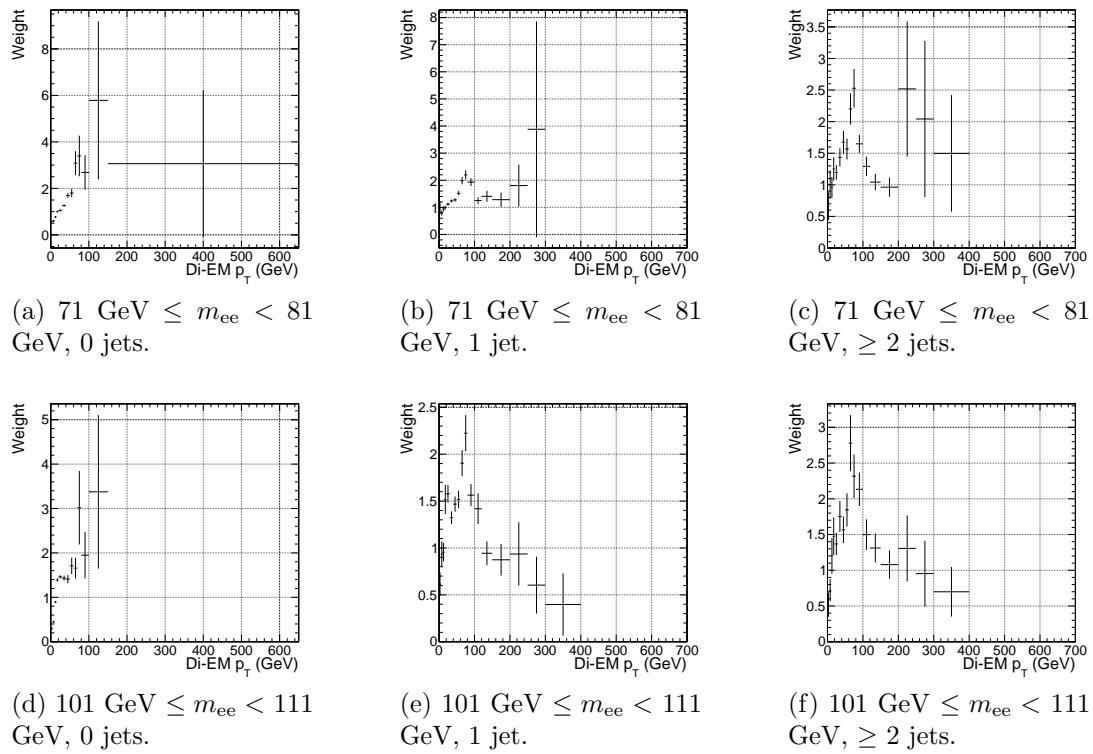


Figure 4.9:  $ee$  sideband di-EM  $p_T$  weights for events with 0, 1, or  $\geq 2$  jets (as in Table 3.2). Errors are statistical only.

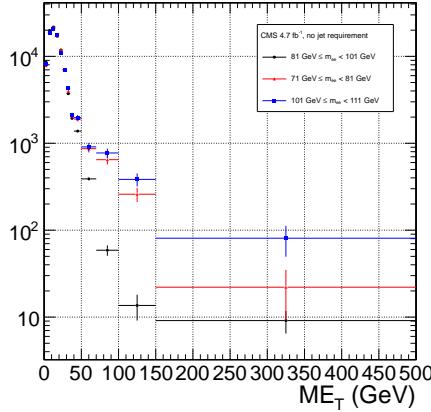


Figure 4.10:  $E_T$  spectra of the  $ee$  sample for  $71 \text{ GeV} \leq m_{ee} < 81 \text{ GeV}$  (red triangles),  $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$  (black circles), and  $101 \text{ GeV} \leq m_{ee} < 111 \text{ GeV}$  (blue squares). The two sideband distributions (red and blue) and the  $Z$  signal distribution (black) are normalized to the total number of  $\gamma\gamma$  events. Errors on the sideband distributions are statistical only, while the error on the  $Z$  signal distribution includes statistical, reweighting, and normalization error (see Sec. ??).

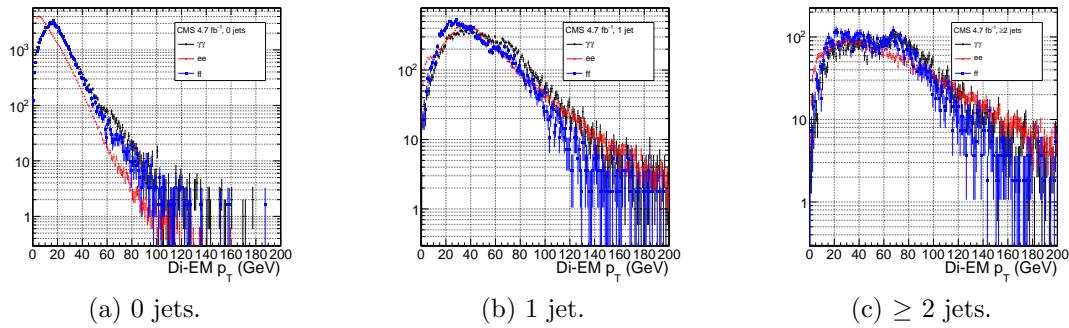


Figure 4.11:  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) (red triangles),  $ff$  (blue squares), and  $\gamma\gamma$  (black circles) di-EM  $p_T$  spectra for events with 0, 1, or  $\geq 2$  jets (as in Table 3.2). Errors are statistical only. **Zoom out the x-axis to show the full tail out to 500 GeV?**

and  $\gamma\gamma$  spectra are due to kinematic  $\Delta R$  and  $p_T$  turn-ons that are suppressed in the  $ee$  sample due to the invariant mass cut. Figure 4.12 shows the weights applied to the  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) and  $ff$   $\cancel{E}_T$  spectra as a function of di-EM  $p_T$  and number of jets per event.

### 4.1.3 Normalization

After reweighting, the  $\cancel{E}_T$  distributions of the QCD control samples are normalized to the  $\cancel{E}_T < 20$  GeV region of the candidate  $\gamma\gamma \cancel{E}_T$  spectrum, where signal contamination is low. The normalization factor is  $(N_{\gamma\gamma}^{\cancel{E}_T < 20\text{GeV}} - N_{e\gamma}^{\cancel{E}_T < 20\text{GeV}})/N_{\text{control}}^{\cancel{E}_T < 20\text{GeV}}$ , where  $N_{e\gamma}^{\cancel{E}_T < 20\text{GeV}}$  is the expected number of electroweak background events with  $\cancel{E}_T < 20$  GeV (discussed in Section 4.2).

## 4.2 Modeling the Electroweak Background

$W\gamma$ ,  $W + \text{jet}$ , and  $t\bar{t}$  processes in which the  $W$  decay electron is misidentified as a photon (due to a failure to properly associate a pixel seed to the electron candidate) can contribute significantly to the high- $\cancel{E}_T$  tail of the  $\gamma\gamma \cancel{E}_T$  spectrum. To estimate this background, the  $e\gamma$  sample, which is enriched in  $W \rightarrow e\nu$  decays, is scaled by

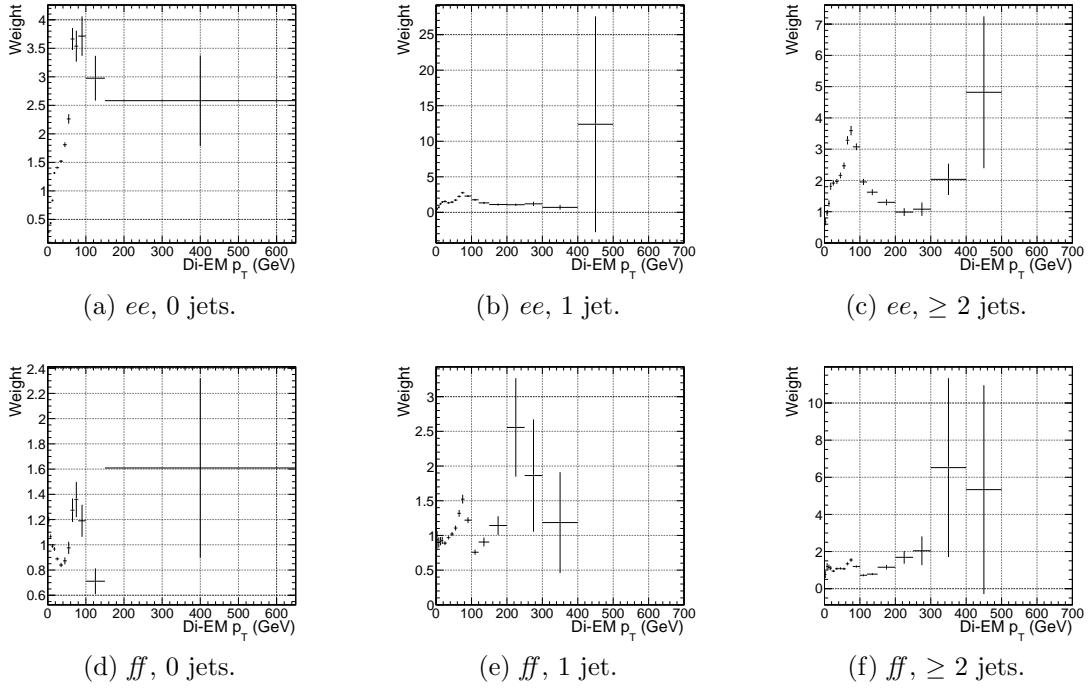


Figure 4.12:  $ee$  ( $81 \text{ GeV} \leq m_{ee} < 101 \text{ GeV}$ ) and  $ff$  di-EM  $p_T$  weights for events with 0, 1, or  $\geq 2$  jets (as in Table 3.2). Errors are statistical only. **Zoom in the x-axis to hide large weights with large statistical errors?**

$f_{e \rightarrow \gamma}/(1 - f_{e \rightarrow \gamma})$ , where  $f_{e \rightarrow \gamma}$  is the rate at which electrons are misidentified as photons.

The derivation of this scaling factor comes from the two equations

$$N_{e\gamma}^W = f_{e \rightarrow e} N_W \quad (4.2)$$

$$N_{\gamma\gamma}^W = (1 - f_{e \rightarrow e}) N_W \quad (4.3)$$

where  $N_{e\gamma}^W$  is the number of  $W$  events in the  $e\gamma$  sample in which the electron was correctly identified,  $f_{e \rightarrow e}$  is the probability to correctly identify an electron,  $N_W$  is the true number of triggered  $W \rightarrow e\nu$  events, and  $N_{\gamma\gamma}^W$  is the number of  $W$  events in the  $\gamma\gamma$  sample in which the electron was misidentified as a photon. The contribution from  $Z \rightarrow ee$  can be neglected (i.e.  $f_{e \rightarrow \gamma}$  is small and the  $Z$  contribution involves  $f_{e \rightarrow \gamma}^2$ , since both electrons have to be misidentified). Since  $f_{e \rightarrow e} = 1 - f_{e \rightarrow \gamma}$ , solving

for  $N_{\gamma\gamma}^W$  gives

$$N_{\gamma\gamma}^W = \frac{f_{e\rightarrow\gamma}}{1 - f_{e\rightarrow\gamma}} N_{e\gamma}^W \quad (4.4)$$

$f_{e\rightarrow\gamma}$  is measured by fitting the  $Z$  peaks in the  $ee$  and  $e\gamma$  samples. The number of  $Z$  events fitted in the  $ee$  and  $e\gamma$  samples, respectively, is given by

$$N_{ee}^Z = (1 - f_{e\rightarrow\gamma})^2 N_Z \quad (4.5)$$

$$N_{e\gamma}^Z = 2f_{e\rightarrow\gamma}(1 - f_{e\rightarrow\gamma}) N_Z \quad (4.6)$$

where  $N_Z$  is the true number of triggered  $Z \rightarrow ee$  events. Solving for  $f_{e\rightarrow\gamma}$  gives

$$f_{e\rightarrow\gamma} = \frac{N_{e\gamma}^Z}{2N_{ee}^Z + N_{e\gamma}^Z} \quad (4.7)$$

A Crystal Ball function is used to model the  $Z$  signal shape in both the  $ee$  and  $e\gamma$  samples, while an exponential convoluted with an error function (“RooCMSShape”, see Sec. 3.4.1) is used to model the background shape. The fixed fit parameters are identical for the two samples, but the other parameters are allowed to float independently. Table 4.3 shows the values and ranges of the fixed and floating fit parameters, respectively. **Edit this to reflect the actual study once done.**

Fits to the  $ee$  and  $e\gamma$  invariant mass spectra are shown in Figure ???. **Make these plots.** Figure ?? indicates that the dependence of  $f_{e\rightarrow\gamma}$  on the electron  $p_T$  and  $\eta$  is small. Applying a  $p_T$ - and  $\eta$ -dependent misidentification rate (with  $p_T$  and  $\eta$  binned as in Fig. ??) makes only a **XXX%** difference in the final electroweak background estimate with respect to a constant rate derived from all  $ee$  and  $e\gamma$  events. Therefore,

the constant rate is used in the final electroweak background estimate, with the largest difference between the constant rate and the  $p_T$ -dependent rate taken as a systematic error.

The signal and background shape assumptions are the main sources of systematic error on  $f_{e \rightarrow \gamma}$ . **Check whether the  $p_T$  dependence or the shapes yield the larger error.** To assess the magnitude of this error,  $f_{e \rightarrow \gamma}$  is recalculated using both linear and quadratic background shapes, and with a Crystal Ball + generated  $Z$  signal shape (as used in Sec. 3.4.1). The largest difference from the nominal shape is taken as the error. **Do it yourself. Also check the misidentification rate in MC with varied tracker radiation lengths to see if there is a dependence on the tracker density.**

Using the integrals of the  $Z$  fits shown in Fig. ??, Eq. 4.7, and the shape and  $p_T$  systematics discussed above,  $f_{e \rightarrow \gamma}$  is calculated to be  $0.015 \pm 0.002(\text{stat.}) \pm 0.005(\text{syst.})$ .

**Replace with your calculated number.** The scaled  $e\gamma$  MET distribution is shown in Figure 4.13.

In the  $36 \text{ pb}^{-1}$  version of this analysis [59], it was shown that the  $ee$  sample could accurately predict the QCD and real  $Z$  contribution to the  $e\gamma$  sample at low  $\cancel{E}_T$ , and that the expectation from  $W \rightarrow e\nu$  MC accounted for the remaining  $W$  contribution at high  $\cancel{E}_T$ . A plot of the  $\cancel{E}_T$  distributions of the 2010  $e\gamma$  sample and the predicted components is shown in Figure 4.14. **Repeat for current selection?** This exercise helps to validate both the QCD and electroweak background prediction methods.

### 4.3 Errors on the Background Prediction

The statistical error on the final background estimate in a particular  $\cancel{E}_T$  bin comes from three sources: the number of control sample events collected in that bin, the statistical error on the weights applied to events in that bin, and the statistics of the

Table 4.3: Parameter values for the signal and background PDFs for the  $ee$  and  $e\gamma$  samples. When a bracketed range is given, the parameter is allowed to float within that range. When a constant is given, the parameter is fixed to that constant. **Edit this to reflect the actual study once done.**

PDF	Crystal Ball fit parameters				RooCMSShape fit parameters			
	$\mu$	$\sigma$	$\alpha$	n	$\mu$	$\alpha$	$\beta$	$\gamma$
$ee$ signal	[-1.0, 1.0]	[1.0, 3.0]	0.87	97.0	N/A	N/A	N/A	N/A
$e\gamma$ signal	[-1.0, 1.0]	[1.0, 3.0]	0.73	134.9	N/A	N/A	N/A	N/A
$ee$ background	N/A	N/A	N/A	N/A	65.0	61.949	0.04750	0.01908
$e\gamma$ background	N/A	N/A	N/A	N/A	$\alpha$	[50.0, 100.0]	0.065	0.048

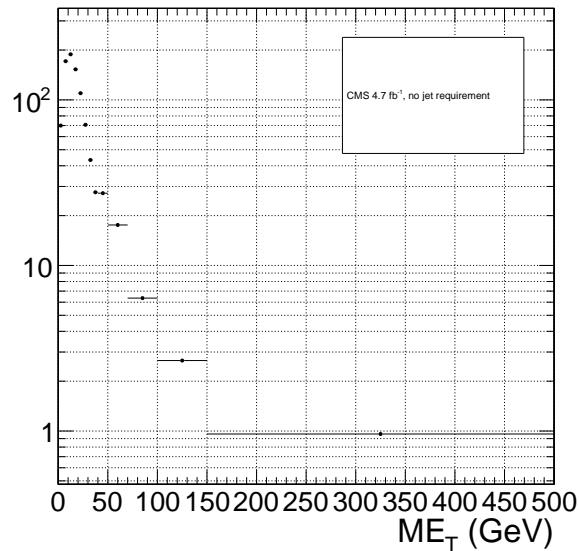


Figure 4.13:  $E_T$  spectrum of the  $e\gamma$  sample after scaling by  $f_{e\rightarrow\gamma}$ . The total error on  $f_{e\rightarrow\gamma}$  is propagated to the total error on the electroweak background estimate. **How to properly treat the error when the same events are used in the  $f_{e\rightarrow\gamma}$  calculation and in the  $e\gamma$  sample? Replace with figure using latest  $f_{e\rightarrow\gamma}$ , and include error bars.**

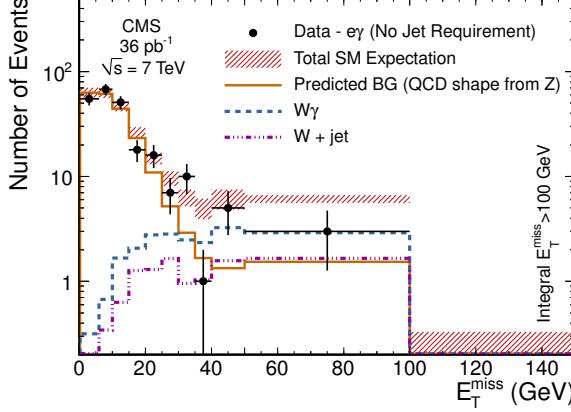


Figure 4.14:  $\cancel{E}_T$  spectrum of the  $e\gamma$  sample in  $35 \text{ pb}^{-1}$  of 2010 LHC data scaled by the 2010 measured  $f_{e \rightarrow \gamma}$  (black dots), QCD and real  $Z$  predicted background from the 2010  $ee$  sample (solid orange line), MC  $W + \text{jet}$  estimate (dash-dotted purple line), and MC  $W\gamma$  estimate (dashed blue line). The total  $e\gamma$  prediction (red band) is the sum of the  $ee$ ,  $W + \text{jet}$ , and  $W\gamma$  predictions. Reprinted from Fig. 2 of ref. [59].

normalization region. In the case of the  $ee$  control sample, there are contributions from the statistics of the  $m_{ee}$  sidebands as well.

In order to estimate the statistical error due to the reweighting procedure, 1000 toy sets of weights are generated. Each set includes a weight for each (di-EM  $p_T$ ,  $N_j$ ) bin, with the values picked from a Gaussian distribution with mean and standard deviation equal to the observed weight for that bin and its statistical error. All bins are uncorrelated. For each of the 1000 experiments, the control sample data are reweighted according to the generated weights, and the background estimates are calculated for each  $\cancel{E}_T$  bin. Since the distribution of the toy background estimates follows a Gaussian distribution in each  $\cancel{E}_T$  bin, the RMS spread of the estimates is taken as the statistical error due to reweighting. This procedure is carried out for the  $ff$ ,  $ee$ , low sideband  $ee$ , and high sideband  $ee$  samples.

The total statistical error on the background estimate per  $\cancel{E}_T$  bin is given by

(4.8)

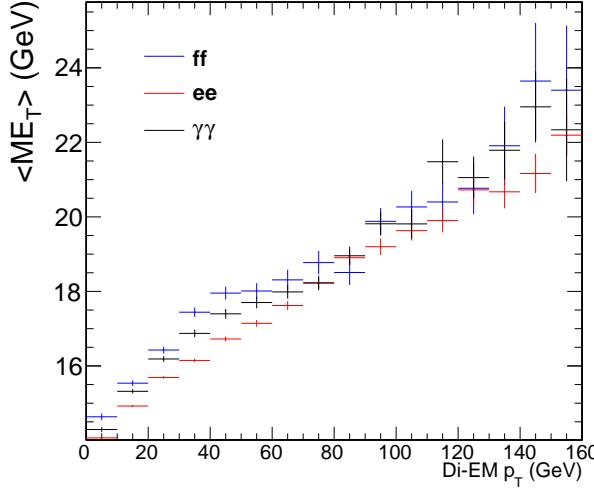


Figure 4.15: Average  $\cancel{E}_T$  vs. di-EM  $p_T$  for the  $ff$  (blue),  $ee$  (red), and  $\gamma\gamma$  (black) samples.

where **Fill in at work.** For  $\cancel{E}_T > 100$  GeV, the errors due to the number of events collected, the normalization, and the reweighting amount to **XXX%**, **XXX%**, and **XXX%** of the total background estimate.

The dominant source of systematic error on the background estimate is the slight difference in hadronic activity between the  $ee$ ,  $ff$ , and  $\gamma\gamma$  samples. This results in a small bias ( $\sim 1$  GeV) of the  $ee$   $\cancel{E}_T$  distribution towards lower values with respect to the  $ff$  and  $\gamma\gamma$  samples, as shown in Figure 4.15. Therefore, the  $ff$  sample is used as the primary QCD background estimator, and the difference between the  $ee$  and  $ff$  predictions is assigned as an error on the knowledge of the hadronic activity. For  $\cancel{E}_T > 100$  GeV, this error amounts to 43% of the total QCD + electroweak background estimate.

The second largest source of systematic error comes from the  $p_T$  dependence of the  $e \rightarrow \gamma$  misidentification rate (see 4.2). For  $\cancel{E}_T > 100$  GeV, the expected electroweak background is  $3.6 \pm \text{XXX}$  events, so this error amounts to **XXX%** of the total QCD + electroweak background estimate.

Finally, the assumption of a constant  $t\bar{t}$  and  $W + \text{jets}$  background shape under

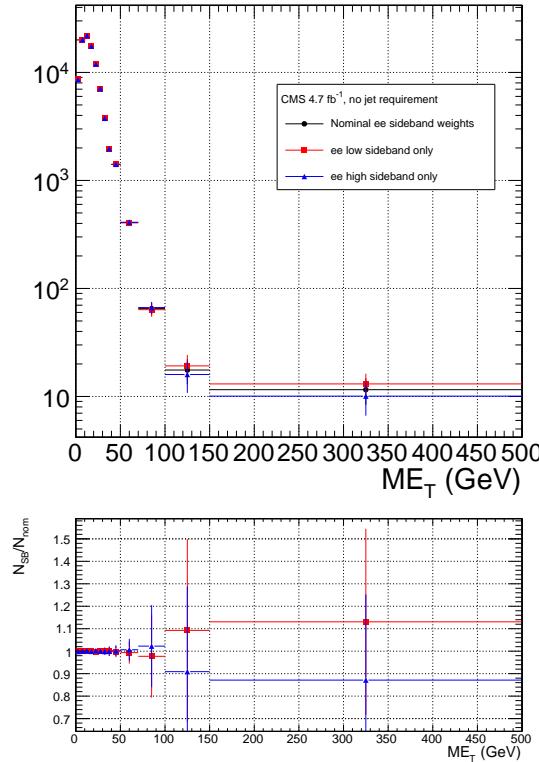


Figure 4.16:  $ee$   $\cancel{E}_T$  distributions using the nominal sideband subtraction (black circles), low sideband only (red squares), and high sideband only (blue triangles). The bottom plot shows the ratio of the low sideband distribution to the nominal (red squares) and the ratio of the high sideband distribution to the nominal (blue triangles).

the  $Z$  peak in the  $ee$  sample induces a systematic error on the  $ee$  sideband-subtracted background prediction. To assess the magnitude of this error, the sideband subtraction (see Sec. 4.1.2) is performed once using only the prediction from the high sideband, and once using only the prediction from the low sideband. In each of these cases, the prediction is weighted by a factor of two, to account for the fact that the sideband regions are only half as wide (10 GeV) as the signal region (20 GeV). The maximum variation from the nominal  $ee$  estimate is taken as the error, which amounts to **XXX%** for  $\cancel{E}_T > 100$  GeV. **Is this the right way to do this?**  $\cancel{E}_T$  distributions using the nominal  $ee$  sideband subtraction, the low-sideband-only subtraction, and the high-sideband-only subtraction are shown in Figure 4.16.

The uncertainty in how to define the (di-EM  $p_T$ ,  $N_j$ ) bins, especially at high di-

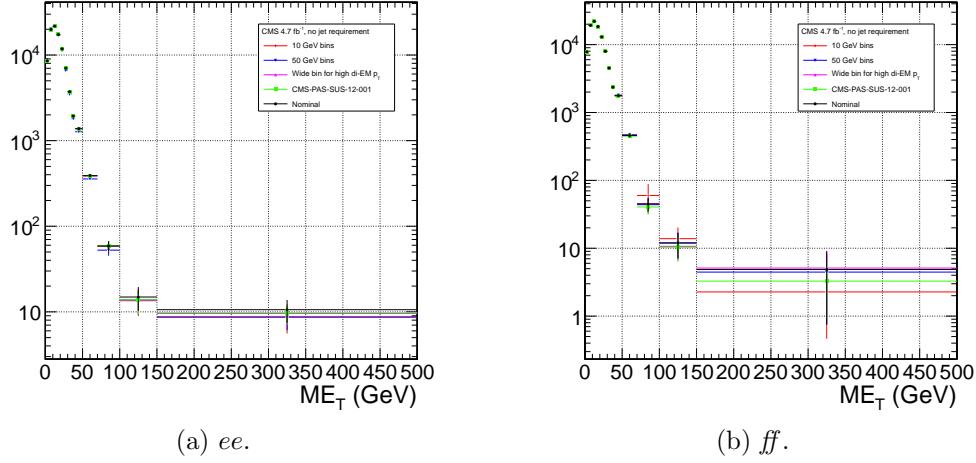


Figure 4.17: Comparison of  $\cancel{E}_T$  distributions for five different di-EM  $p_T$  bin definitions: uniform bins of width 10 GeV (red diamonds); uniform bins of width 50 GeV (blue downward-pointing triangles); bins with lower edges  $\{0.0, 5.0, 10.0, 15.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 750.0\}$  GeV for 0-jet events and  $\{0.0, 5.0, 10.0, 15.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 150.0\}$  GeV for  $\geq 1$ -jet events (magenta upward-pointing triangles), i.e. a single wide bin at high di-EM  $p_T$ ; bins with lower edges  $\{0.0, 5.0, 10.0, 15.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 150.0\}$  GeV for 0-jet events and  $\{0.0, 5.0, 10.0, 15.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 120.0, 150.0, 200.0, 700.0\}$  GeV for  $\geq 1$ -jet events (green squares), i.e. the bins used in ref. [19]; and the nominal bin definitions shown in Fig. 4.12 (black circles).

EM  $p_T$  where the statistics are low, is covered by the 1000-toys procedure as long as the bins are not too coarse. This is shown in Figure 4.17. If the bins are too coarse, the details of the shape of the di-EM  $p_T$  spectra are lost, and the reweighting has a smaller effect.

The use of uncorrected instead of corrected PF  $\cancel{E}_T$  (see Sec. 3.1.3) makes no difference in the agreement of the background predictions and the search sample in a control region at low  $\cancel{E}_T$ , as shown in Figure 4.18. Since the control samples are derived from the same data as the search sample, any biases in the  $\cancel{E}_T$  reconstruction due to jet energy scale are present equally in both samples.

Tables 4.4 and 4.5 list all the errors on the  $ee$  and  $ff$  QCD background predictions, respectively, for the  $\cancel{E}_T$  bins used in the search. Table ?? lists the errors on the

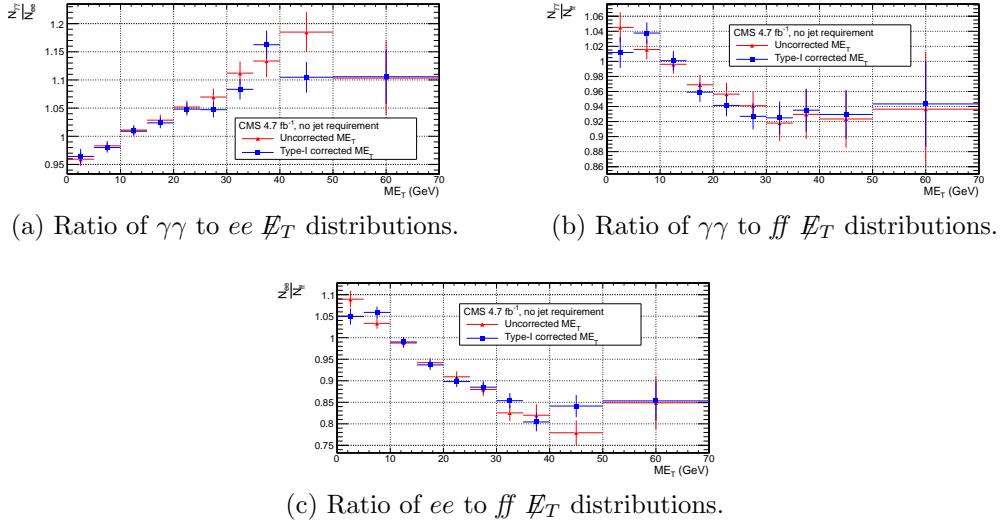


Figure 4.18: Agreement between  $\gamma\gamma$ ,  $ee$ , and  $ff$  samples for uncorrected (red triangles) and corrected (blue squares)  $\not{E}_T$ .

electroweak background prediction. Finally, table ?? shows the errors on the total QCD + electroweak background prediction, broken down by origin (statistical or systematic) and QCD background estimation sample ( $ee$  or  $ff$ ). In the final result, only the  $ff$  QCD estimate is used.

Table 4.4: Errors on the  $ee$  QCD background prediction as a fraction of the  $ee$  prediction.

Source of error	Fractional uncertainty (%)				
	[50, 60)	[60, 70)	[70, 80)	[80, 100)	$\geq 100$
Total	3.9	8.1	16	25	25
Statistics	3.6	7.8	16	24	22
No. events	3.6	7.7	15	24	20
In norm. region	0.43	0.44	0.46	0.55	0.51
In this $\not{E}_T$ bin	3.5	7.7	15	24	20
Reweighting	0.71	1.2	3.6	4.2	7.8
In norm. region	0.2	0.2	0.21	0.25	0.23
In this $\not{E}_T$ bin	0.68	1.2	3.6	4.2	7.8
Systematics	1.4	2	0.72	5.5	12
$f_{e \rightarrow \gamma}$ (in norm. region)	0.0015	0.0015	0.0016	0.0019	0.0018
$m_{ee}$ background shape	1.4	2	0.72	5.5	12

Table 4.5: Errors on the  $ff$  QCD background prediction as a fraction of the  $ff$  prediction.

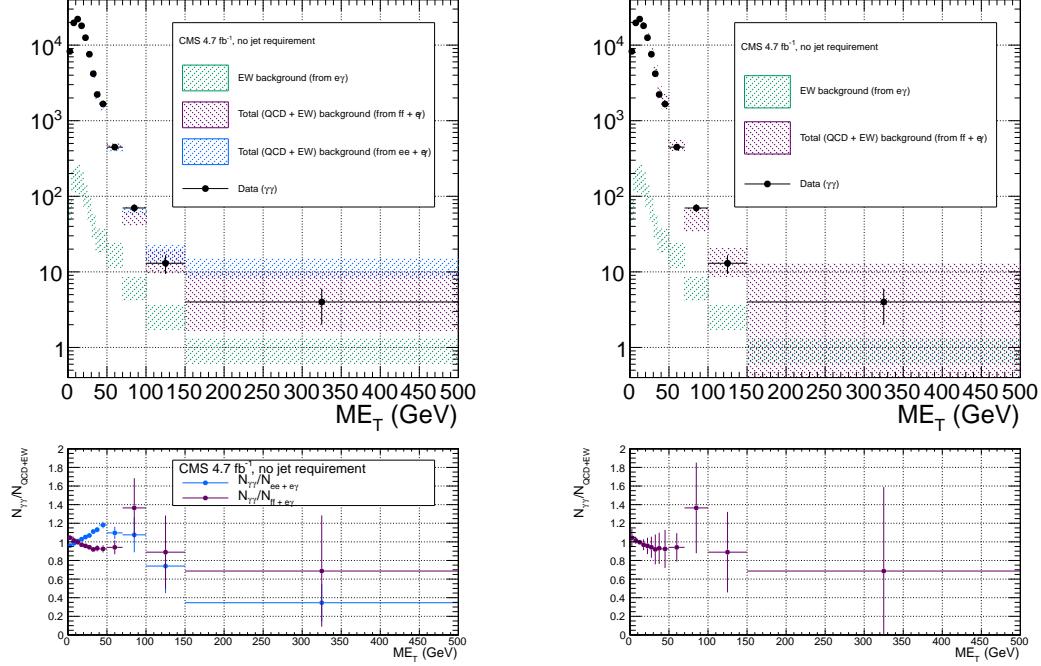
Source of error	Fractional uncertainty (%)				
	[50, 60)	[60, 70)	[70, 80)	[80, 100)	$\geq 100$
Total	15	25	61	34	64
Statistics	7.2	14	30	34	38
No. events	7.1	14	29	33	36
In norm. region	0.64	0.64	0.64	0.64	0.64
In this $E_T$ bin	7.1	14	29	33	36
Reweighting	0.89	2.7	5.2	7.2	13
In norm. region	0.28	0.28	0.28	0.28	0.28
In this $E_T$ bin	0.84	2.7	5.2	7.2	13
Systematics	13	21	53	5.5	52
$ee/ff$ difference	13	21	53	5.5	52
$f_{e \rightarrow \gamma}$ (in norm. region)	0.0015	0.0015	0.0015	0.0015	0.0015

Table 4.6: Errors on the  $e\gamma$  electroweak background prediction as a fraction of the  $e\gamma$  prediction.

Source of error	Fractional uncertainty (%)				
	[50, 60)	[60, 70)	[70, 80)	[80, 100)	$\geq 100$
Total	34	34	35	35	34
Statistics	3.6	5.2	6.7	7.2	6.5
Systematics ( $f_{e \rightarrow \gamma}$ )	34	34	34	34	34

Table 4.7: Errors on the total QCD + electroweak background prediction as a fraction of the total prediction.

Source of error	Fractional uncertainty (%)				
	[50, 60)	[60, 70)	[70, 80)	[80, 100)	$\geq 100$
Total ( $ee + e\gamma$ )	4	7.9	15	22	22
Statistics	3.4	7.3	14	21	18
QCD	3.4	7.2	14	21	18
Electroweak	0.13	0.32	0.57	0.84	0.81
Systematics	1.9	2.8	2.9	6.3	12
QCD	1.4	1.9	0.66	4.8	11
Electroweak	1.3	2.1	2.9	4	4.2
Total ( $ff + e\gamma$ )	14	24	53	30	53
Statistics	6.9	13	26	29	30
QCD	6.9	13	26	29	30
Electroweak	0.12	0.26	0.84	0.88	1.1
Systematics	12	20	47	6.4	43
QCD	12	20	46	4.9	42
Electroweak	1.1	1.7	4.2	4.2	6



(a)  $ee + e\gamma$  and  $ff + e\gamma$ . The widths of the bands correspond to the errors given in Table 4.7, excluding the error associated with the difference between the  $ee$  and  $ff$  QCD estimates for the  $ff + e\gamma$   $E_T$  distribution.

(b)  $ff + e\gamma$ . The widths of the bands correspond to the errors given in Table 4.7, including the error associated with the difference between the  $ee$  and  $ff$  QCD estimates.

Figure 4.19:  $E_T$  distribution of the  $\gamma\gamma$  search sample (black circles) along with the predicted  $E_T$  distributions of the QCD and electroweak backgrounds (blue band for  $ee$  QCD prediction + electroweak prediction, purple band for  $ff$  QCD prediction + electroweak prediction). The electroweak background prediction is shown in green. The bottom plots show the ratio of the  $\gamma\gamma E_T$  distribution to the  $ee + e\gamma$  background distribution (blue) and  $ff + e\gamma$  background distribution (purple).

## 4.4 Results

Figure 4.19 shows the  $E_T$  distribution of the  $\gamma\gamma$  search sample along with the predicted  $E_T$  distributions of the QCD and electroweak backgrounds. The observed number of two-photon events, background estimates and their errors, and expected number of two-photon events from two representative GGM SUSY models are listed in Table 4.8. (Details of the SUSY MC production are given in Chapter ?? and App. A.) No deviation from the Standard Model prediction is observed in the  $\gamma\gamma$  search sample.

Table 4.8: Observed numbers of two-photon events, background estimates and their errors, and expected numbers of two-photon events from two representative GGM SUSY models (details of MC simulation given in Chapter ?? and App. A) for the  $E_T$  bins used in the search. Errors on the background estimates are detailed in Tables 4.4, 4.5, 4.6, and 4.7. Errors on the expected numbers of GGM events are purely statistical.

Source	No. events				
	[50, 60)	[60, 70)	[70, 80)	[80, 100)	$\geq 100$
Data observed ( $\gamma\gamma$ )	354	93	37	33	17
Background predicted $(ff + e\gamma)$	$362 \pm 51.6$	$113 \pm 27.1$	$27.2 \pm 14.5$	$24.1 \pm 7.27$	$20.4 \pm 10.9$
GGM1 expected (fill in)					
GGM2 expected (fill in)					

# Appendix A

## Monte Carlo Samples

A number of MC samples are utilized in this analysis and referred to throughout the text. Below is a list of the MC samples used and an explanation of what the sample names mean.

### A.0.1 List of Samples

1. Drell-Yan:

```
/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/  
Fall11-PU_S6_START42_V14B-v1/AODSIM
```

2. QCD enriched with  $B$  and  $D$  meson decays to electrons:

```
/QCD_Pt-20to30_BCToE_TuneZ2_7TeV-pythia6/  
Fall11-PU_S6_START42_V14B-v1/AODSIM,  
/QCD_Pt-30to80_BCToE_TuneZ2_7TeV-pythia6/  
Fall11-PU_S6_START42_V14B-v1/AODSIM,  
/QCD_Pt-80to170_BCToE_TuneZ2_7TeV-pythia6/  
Fall11-PU_S6_START42_V14B-v1/AODSIM
```

3. Photon + jet doubly enriched with jets passing an EM filter:

```
/GJet_Pt-20_doubleEMEnriched_TuneZ2_7TeV-pythia6/
Fall11-PU_S6_START42_V14B-v1/AODSIM
```

4.  $W$  leptonic decays:

```
/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/
Fall11-PU_S6_START42_V14B-v1/AODSIM
```

5.  $t\bar{t}$ :

```
/TTJets_TuneZ2_7TeV-madgraph-tauola/
Fall11-PU_S6_START42_V14B-v2/AODSIM
```

### A.0.2 Explanation of Naming Conventions

- L: charged lepton
- B:  $B$  hadron
- C:  $D$ , or charmed, hadron
- E: electron or positron
- G: photon
- W:  $W$  boson
- Nu: neutrino
- T: top quark
- TuneZ2: Pythia tune incorporating 2010 LHC data with CTEQ6L1 [60] PDFs [61]
- M-50: Generated  $l^+l^-$  invariant mass threshold of 50 GeV
- 7TeV: Generated center-of-mass energy 7 TeV

- **pythia6**: Parton showering and hadronization simulated with Pythia v6.424 [62]
- **madgraph**: Hard interaction generated with MadGraph [63]
- **tauola**:  $\tau$  decays generated with Tauola [64]
- **PU\_S6**: Generated with S6 pileup scenario
- **START42\_V14B**: Reconstructed with best alignment and calibration constants and magnetic field conditions as of August 3, 2011
- **Pt\_XtoY**:  $X \leq$  generated  $\hat{p}_T < Y$
- **BCToE**: Enriched in  $B$  and  $D$  meson decays to electrons
- **doubleEMEnriched**: Enriched in electromagnetic jets

# Bibliography

- [1] S.L. Glashow, J. Iliopoulos, and L. Maiani, *Phys. Rev. D* **2** (1970) 1285; S.L. Glashow, *Nucl. Phys.* **22(4)** (1961) 579; J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127** (1962) 965; S. Weinberg, *Phys. Rev. Lett.* **19** (1967) 1264; A. Salam and J.C. Ward, *Phys. Lett.* **13(2)** (1964) 168.
- [2] M. Gell-Mann, *Phys. Lett.* **8** (1964) 214; G. Zweig, *CERN 8419/TH. 412* (1964) (unpublished).
- [3] J. Drees, *Int. J. Mod. Phys.* **A17** (2002) 3259.
- [4] P.W. Higgs, *Phys. Lett.* **12(2)** (1964) 132; P.W. Higgs, *Phys. Rev. Lett.* **13** (1964) 508; P.W. Higgs, *Phys. Rev.* **145** (1966) 1156.
- [5] I. Aitchison, *Supersymmetry in Particle Physics: An Elementary Introduction* (Cambridge University Press, Cambridge 2007), p. 4.
- [6] S. P. Martin, *A Supersymmetry Primer* **v4** (2006) 86. arXiv:hep-ph/9709356.
- [7] M. Dine and W. Fischler, *Phys. Lett.* **B110** (1982) 227; C.R. Nappi and B.A. Ovrut, *Phys. Lett.* **B113** (1982) 175; L. Alvarez-Gaumé, M. Claudson, and M.B. Wise, *Nucl. Phys.* **B207** (1982) 96; M. Dine and A.E. Nelson, *Phys. Rev.* **D48** (1993) 1277; M. Dine, A.E. Nelson, and Y. Shirman, *Phys. Rev.* **D51** (1995) 1362; M. Dine, A.E. Nelson, Y. Nir, and Y. Shirman, *Phys. Rev.* **D53** (1996) 2658.

- [8] A.H. Chamseddine, R. Arnowitt, and P. Nath, *Phys. Rev. Lett.* **49** (1982) 970; R. Barbieri, S. Ferrara, and C.A. Savoy, *Phys. Lett.* **B119** (1982) 343; L.E. Ibáñez, *Phys. Lett.* **B118** (1982) 73; L.J. Hall, J.D. Lykken, and S. Weinberg, *Phys. Rev.* **D27** (1983) 2359; N. Ohta, *Prog. Theor. Phys.* **70** (1983) 542; J. Ellis, D.V. Nanopoulos, and K. Tamvakis, *Phys. Lett.* **B121** (1983) 123; L. Alvarez-Gaumé, J. Polchinski, and M. Wise, *Nucl. Phys.* **B221** (1983) 495.
- [9] P. Meade, N. Seiberg, and D. Shih, *Progr. Theor. Phys. Suppl.* **177** (2009) 143.
- [10] G. Aad et al., *CERN-PH-EP-2011-160* (2011).
- [11] T. Aaltonen et al., *Phys. Rev. Lett.* **104** (2010) 011801.
- [12] CMS Collaboration, *CMS-PAS-SUS-11-009* (2011).
- [13] <http://en.wikipedia.org/wiki/Tevatron>.
- [14] E. Fernandez et al., *Phys. Rev. Lett.* **54** (1985) 1118; E. Fernandez et al., *Phys. Rev.* **D35** (1987) 374; D. Decamp et al., *Phys. Lett.* **B237(2)** (1990) 291; F. Abe et al., *Phys. Rev. Lett.* **75** (1995) 613; S. Abachi et al., *Phys. Rev. Lett.* **75** (1995) 618; G. Alexander et al., *Phys. Lett.* **B377(4)** (1996) 273; S. Aid et al., *Z. Phys.* **C71(2)** (1996) 211; S. Aid et al., *Phys. Lett.* **B380(3-4)** (1996) 461; B. Aubert et al., *Phys. Rev. Lett.* **95** (2005) 041802.
- [15] O. Buchmueller et al., *CERN-PH-TH/2011-220* (2011).
- [16] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
- [17] G. Aad et al., *JINST* **3** (2008) S08003.
- [18] B.C. Allanach et al., *Eur. Phys. J.* **C25** (2002) 113.
- [19] CMS Collaboration, *CMS-PAS-SUS-12-001* (2012).

- [20] A. Boyarsky, J. Lesgourges, O. Ruchayskiy, and M. Viel, *CERN-PH-TH/2008-234* (2009).
- [21] E. Komatsu et al., *Astrophys. J. Suppl. Ser.* **180** (2009) 330.
- [22] C.-H. Chen and J.F. Gunion, *Physical Review D***58** (1998) 075005.
- [23] F. Staub, W. Porod, J. Niemeyer, *JHEP* **1001** (2010) 058.
- [24] R. Brunelière, *Nucl. Instr. Meth. Res.* **A572** (2007) 33.
- [25] S. Chatrchyan et al. (CMS Collaboration), *JINST* **5** (2010) T03011.
- [26] P. Meridiani and C. Seez, *CMS IN-2011/002* (2011).
- [27] P. Adzic et al. (CMS Electromagnetic Calorimeter Group), *Eur. Phys. J.* **C44S2** (2006) 1.
- [28] P. Adzic et al. (CMS Electromagnetic Calorimeter Group), *JINST* **3** (2008) P10007.
- [29] M. Malberti, *Nuc. Sci. Symposium Conference Record NSS/MIC IEEE* (2009) 2264.
- [30] S. Chatrchyan et al. (CMS Collaboration), *JINST* **5** (2010) T03010.
- [31] R. Paramatti, *J. Phys. Conf. Ser.* **293** (2011) 012045.
- [32] Y. Yang, [http://www.hep.caltech.edu/cms/posters/Pi0Poster\\_CMSWeekDec2011.pdf](http://www.hep.caltech.edu/cms/posters/Pi0Poster_CMSWeekDec2011.pdf) (2011).
- [33] M. Anderson, A. Askew, A.F. Barfuss, D. Evans, F. Ferri, K. Kaadze, Y. Maravin, P. Meridiani, and C. Seez, *CMS IN-2010/008* (2010).
- [34] The H $\rightarrow \gamma\gamma$  working group, *CMS AN-2011/426* (2011).

- [35] M. Cacciari, *LPTHE-P06-04* (2006).
- [36] M. Cacciari, G.P. Salam, and G. Soyez, *CERN-PH-TH-2011-297* (2011).
- [37] A. Askew, B. Cox, D. Elvira, Y. Gershtein, M. Hildreth, D. Jang, Y.-F. Liu, D. Mason, D. Morse, U. Nauenberg, M. Paulini, R. Stringer, R. Yohay, and S.L. Zang, *CMS AN-2011/515* (2011).
- [38] W. Adam, R. Fr  wirth, A. Strandlie, and T. Todorov, *J. Phys.* **G31** No. 9 (2005).
- [39] CMS Collaboration, *CMS PAS PFT-09-001* (2009).
- [40] CMS Collaboration, *CMS PAS PFT-10-002* (2010).
- [41] S. Chatrchyan et al., *JINST* **6** (2011) P11002.
- [42] CMS Collaboration, *CERN-PH-EP 2011-102* (2011).
- [43] M. Cacciari, G.P. Salam, and G. Soyez, *JHEP* **0804** (2008) 063.
- [44] G. Salam, talk given at CERN Theory Institute: SM and BSM Physics at the LHC (2009).
- [45] C.W. Fabjan and R. Wigmans, *Rep. Prog. Phys.* **52** (1989) 1519.
- [46] <http://www-cdf.fnal.gov/physics/new/top/2004/jets/cdfpublic.html>  
(visited on 16 January 2011).
- [47] S. Chatrchyan et al. (CMS Collaboration), *JINST* **6** (2011) P09001.
- [48] CMS Collaboration, *CMS PAS JME-09-002* (2009).
- [49] T. Sj  strand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178** (2008) 852.
- [50] S. Chatrchyan et al. (CMS Collaboration), *JINST* **5** (2010) T03014.

- [51] J. P. Chou, S. Eno, S. Kunori, S. Sharma, and J. Wang, *CMS IN-2010/006* (2010).
- [52] Y. Chen, talk given at a meeting of the CMS JetMET group (2011).
- [53] R. Korzekwa et al., *iEEE Trans. Elec. Dev.* **38** (1991) 745.
- [54] G. Daskalakis, D. Evans, C.S. Hill, J. Jackson, P. Vanlaer, J. Berryhill, J. Haupt, D. Futyana, C. Seez, C. Timlin, and D. Wardrope, *CMS AN-2007/019* (2007).
- [55] F. James and M. Roos, *Comput. Phys. Commun.* **10** (1975) 343.
- [56] W. Verkerke and D.P. Kirkby, *CHEP-2003-MOLT007* (2003).
- [57] J.E. Gaiser, Ph.D. thesis, Stanford University (1982).
- [58] Information about all CMS datasets is available from the CMS Data Aggregation System (DAS), located at the URL <https://cmsweb.cern.ch/das/>.
- [59] S. Chatrchyan et al., *Phys. Rev. Lett.* **106** (2011) 211802.
- [60] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, *Phys. Rev.* **D78** (2008) 013004.
- [61] R. Field, talk given at LHC Physics Centre at CERN Minimum Bias and Underlying Event Working Group Meeting (2011).
- [62] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *JHEP* **0605** (2006) 026.
- [63] dummy
- [64] dummy