Chapter 1

Data Analysis

Update the int. lumi. everywhere.

The signature of GGM SUSY particle production in this search is an excess of two-photon events with high $\not{E}_T$. $\not{E}_T$ is reconstructed using the particle flow algorithm as described in Sec. **??**. Candidate two-photon events, as well as control events, are selected according to the offline object criteria presented in Secs. **??** and **??**, the event quality criteria in Sec. **??**, and the trigger requirements in Sec. **??**. These are summarized in Table 1.1.

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

- `/Photon/Run2011A-05Jul2011ReReco-ECAL-v1/AOD`
- `/Photon/Run2011A-05Aug2011-v1/AOD`
- `/Photon/Run2011A-03Oct2011-v1/AOD`
- `/Photon/Run2011B-PromptReco-v1/AOD`

The search strategy is to model the backgrounds to the GGM SUSY signal using $\not{E}_T$ shape templates derived from the control samples, and then to look for a high-$\not{E}_T$

excess above the estimated background in the $\gamma\gamma$ sample. There are two categories of backgrounds: QCD processes with no real $\not\!\!E_T$ and electroweak processes with real $\not\!\!E_T$ 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 $W \to e\nu$ 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 $\not\!\!E_T$ spectrum of the $\gamma\gamma$ search data sample overlaid on the $\not\!\!E_T$ 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 $\sigma_{i\eta i\eta}$ 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 $\not\!\!E_T$ 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 $\gamma\gamma/ff$ $\not{E}_T$ agreement for different values of the ff H/E cut in MC. Make the same plot in data for a restricted $\not{E}_T$ 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 $\not{E}_T$, 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 $W \to e\nu$ decays. The $e\gamma$ $\not{E}_T$ distribution is scaled by the electron→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. 1.1 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. **??**.

1.1 Modeling the QCD Background

1.1.1 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 $\not{E}_T$ 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

$\not\!\!E_T$ spectrum of the control sample is created, it is then normalized in the low-$\not\!\!E_T$ region, the assumption being that GGM SUSY does not predict a significant amount of events at low $\not\!\!E_T$. There are three aspects to this QCD background estimation procedure that bear highlighting:

Choice of control sample Since the underlying cause of $\not\!\!E_T$ 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 $\not\!\!E_T$ spectrum appears as it would if the control sample had the same kinematic properties as the candidate sample (i.e. particle p_T and η distributions, etc.). By choosing an appropriate control sample and reweighting it, the control $\not\!\!E_T$ distribution should now match both the hadronic activity and the kinematics of the candidate sample.

Normalization Finally, the control $\not\!\!E_T$ distribution is normalized in a region of low $\not\!\!E_T$, where contamination from the expected GGM SUSY signal is small. This implies an extrapolation of the low-$\not\!\!E_T$ QCD background prediction to the high-$\not\!\!E_T$ signal region.

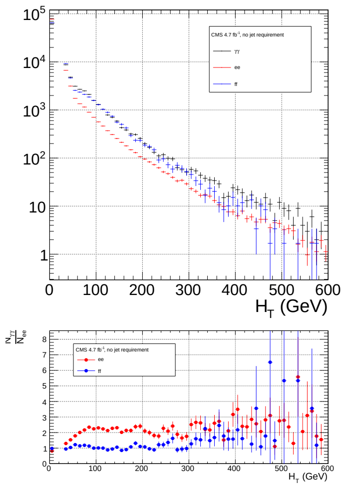
As explained in the beginning of this chapter, the *ff* sample is used as the primary QCD control sample, while the *ee* sample is used as a cross-check. Both samples have two well-measured EM objects per event, no real $\not\!\!E_T$, and similar hadronic activity to the $\gamma\gamma$ sample. Figure 1.1 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. 1.1a, 1.1b, 1.1c, and 1.1d. 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. 1.1.2).

Table 1.1: Selection criteria for $\gamma\gamma$, $e\gamma$, ee, and $f\!f$ events.

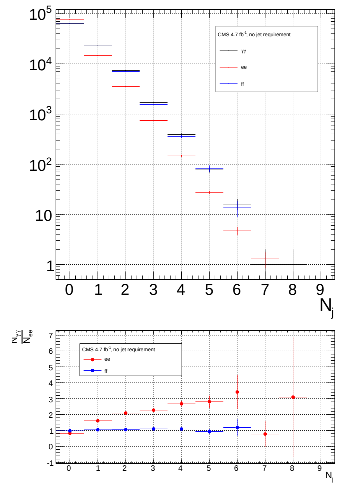
Variable	Cut			
	$\gamma\gamma$	$e\gamma$	ee	$f\!f$
HLT match	IsoVL	IsoVL	IsoVL	IsoVL || R9Id
E_T	> 40/ > 25 GeV	> 40/ > 25 GeV	> 40/ > 25 GeV	> 40/ > 25 GeV
SC $\|\eta\|$	< 1.4442	< 1.4442	< 1.4442	< 1.4442
H/E	< 0.05	< 0.05	< 0.05	< 0.05
$R9$	< 1	< 1	< 1	< 1
Pixel seed	No/No	Yes/No	Yes/Yes	No/No
I_{comb},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

Table 1.2: Definition of HB/HE/HF hadronic jets. **Add a footnote describing the PF electron and PF muon definitions, with references.**

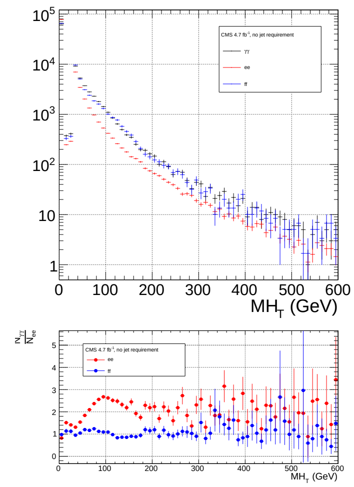
Variable	Cut
Algorithm	`L1FastL2L3Residual` corrected PF (cf. Sec. ??)
p_T	> 30 GeV
$\|\eta\|$	< 5.0
Neutral hadronic energy fraction	< 0.99
Neutral electromagnetic energy fraction	< 0.99
Number of constituents	> 1
Charged hadronic energy	> 0.0 GeV if $\|\eta\| < 2.4$
Number of charged hadrons	> 0 if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



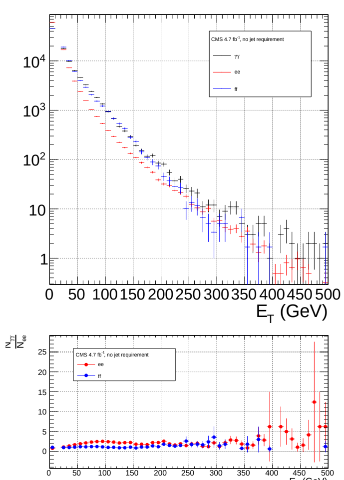
(a) H_T, defined as the scalar sum of corrected jet E_T for jets defined as in Table 1.2.



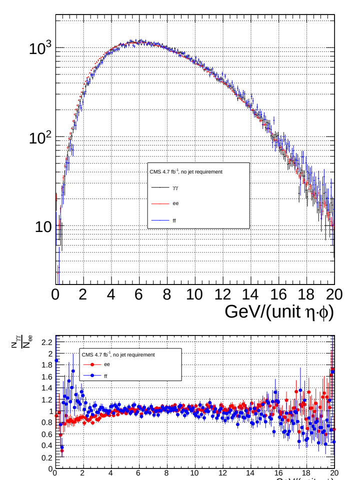
(b) Number of jets per event for jets defined as in Table 1.2.



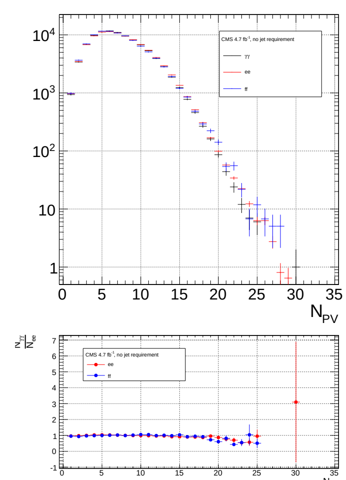
(c) $\not{H}_T$, defined as the magnitude of the negative vectorial sum of corrected jet E_T for jets defined as in Table 1.2.



(d) Corrected E_T for the jet with the largest corrected E_T per event, for jets defined as in Table 1.2 (excluding the p_T requirement).



(e) ρ (average pileup energy density in the calorimeters per unit $\eta \cdot \phi$, cf. Sec. **??**).



(f) Number of good reconstructed primary vertices per event according to the criteria of Sec. **??**.

Figure 1.1: Comparison of the shapes of some distributions relevant to hadronic activity between the $\gamma\gamma$, ee (81 GeV $\leq m_{ee} <$ 101 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.

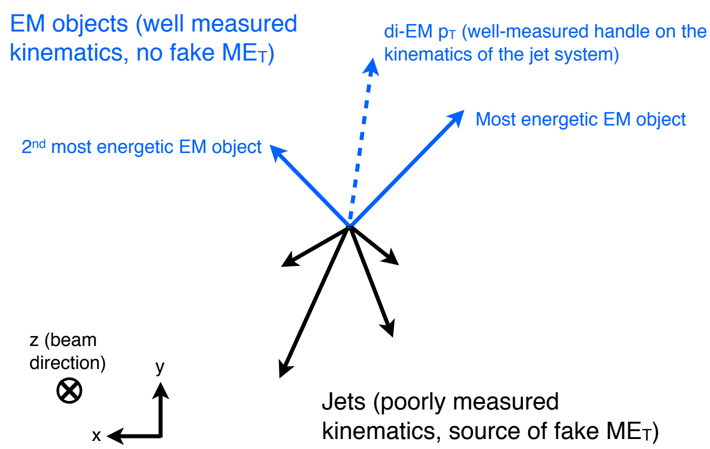
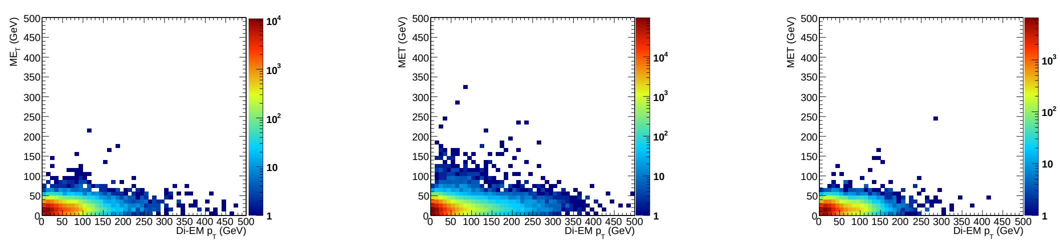


Figure 1.2: Cartoon showing the di-EM system in blue and the hadronic recoil in black. The di-EM p_T (dashed blue line) is used to reweight the control sample kinematic properties to match those of the candidate $\gamma\gamma$ sample.

.



(a) $\gamma\gamma$.

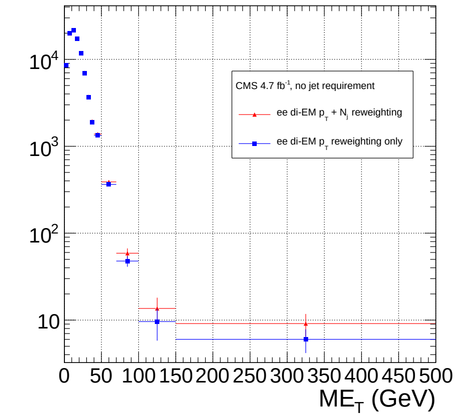
(b) ee.

(c) ff.

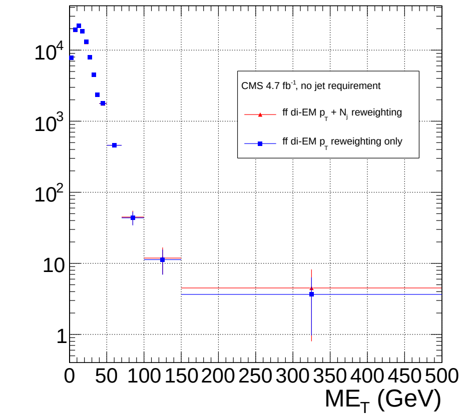
Figure 1.3: $\not{E}_T$ vs. di-EM p_T.

1.1.2 Reweighting

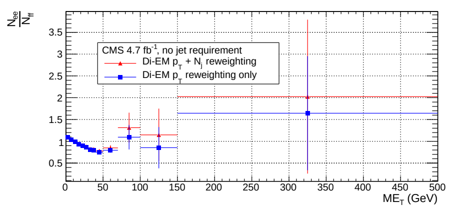
To reweight the control sample events to match the kinematics of the candidate sample events, a weight based on the p_T of the di-EM-object system and the number of jets in the event is used. As explained in Sec. 1.1.1, $\not{E}_T$ in the $\gamma\gamma$, ff, and ee samples is due to the poorly measured 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 $\not{E}_T$. The di-EM system is depicted in Figure 1.2. As shown in Figure 1.3, $\not{E}_T$ is largely uncorrelated with di-EM p_T, so there is little danger of reweighting away a true signal excess.



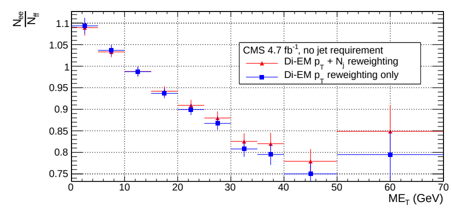
(a) ee $\not\!\!E_T$ spectra.



(b) ff $\not\!\!E_T$ spectra.



(c) Ratio of ee to ff $\not\!\!E_T$ spectra.



(d) Ratio of ee to ff $\not\!\!E_T$ spectra, zoomed x-axis.

Figure 1.4: $\not\!\!E_T$ spectra of the reweighted ee (81 GeV $\leq m_{\mathrm{ee}} <$ 101 GeV) and ff control samples. Blue squares indicate di-EM p_T reweighting only; red triangles indicate di-EM p_T + number of jets reweighting. PF p_T (cf. p. 11) is used to calculate the di-EM p_T. The full normalization procedure is employed, along with ee sideband subtraction (discussed in at the end of this section). Error bars include statistical, reweighting, and normalization error (see Sec. **??**).

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. 1.1). Jets are defined as in Table **??**. Figure 1.4 shows the effect of reweighting by number of jets per event, which is to increase(decrease) the tail of the ee(ff) $\not\!\!E_T$ spectrum.

Although the electron and photon energies are well measured by the ECAL, the ECAL-only measurement of the fake photon energy (cf. Sec **??**) 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. 1.5 and 1.6, 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.[1] This leads to a modest improvement in the agreement between the *ee* and *ff* $\not\!\!E_T$ spectra, as shown in Figure 1.7.

The control sample event weights are defined as

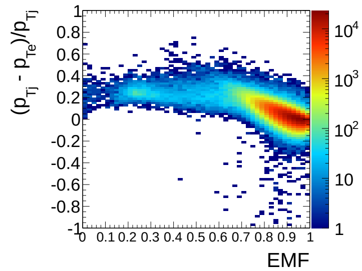
$$w_{ij} = \frac{N_{\text{control}}}{N_{\gamma\gamma}} \frac{N^{ij}_{\gamma\gamma}}{N^{ij}_{\text{control}}} \tag{1.1}$$

where i runs over the number of di-EM p_T bins, j runs over the number of jet bins, N_{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, $N^{ij}_{\gamma\gamma}$ is the number of $\gamma\gamma$ events in the i^{th} di-EM p_T bin and j^{th} jet bin, and N^{ij}_{control} is the number of control sample events in the i^{th} di-EM p_T bin and j^{th} jet bin. The effect of the reweighting is more significant for the *ee* sample than for the *ff* sample, as shown in Figure 1.8.

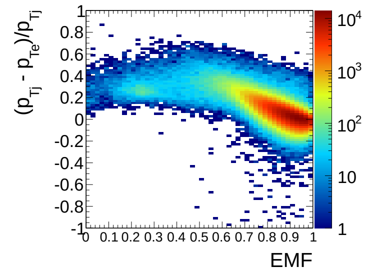
The *ee* sample contains a non-negligible background of $t\bar{t}$ events in which both W bosons decay to electrons. These events have significant real $\not\!\!E_T$ from the two neutrinos (unlike the $\gamma\gamma$ events), and therefore inflate the background estimate at high $\not\!\!E_T$. In order 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 GeV $\leq m_{\text{ee}} <$ 101 GeV, where m_{ee} is the

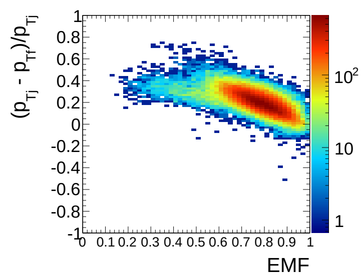
[1] In the few events ($\mathcal{O}(10^{-3})$) in which two PF jet objects, corresponding to the two electrons or fakes, are not found, the ECAL-only p_T is used.



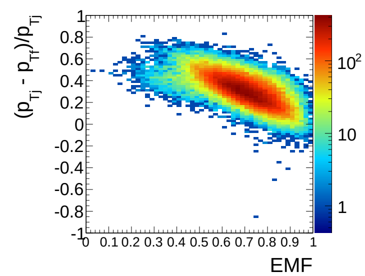
(a) Leading electron in ee events.



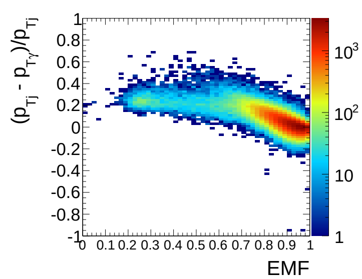
(b) Trailing electron in ee events.



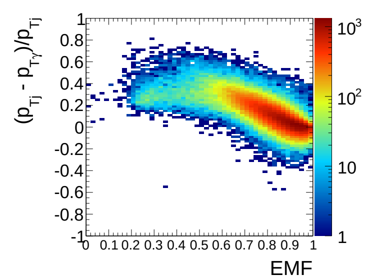
(c) Leading fake in ff events.



(d) Trailing fake in ff events.



(e) Leading photon in $\gamma\gamma$ events.



(f) Trailing photon in $\gamma\gamma$ events.

Figure 1.5: 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.

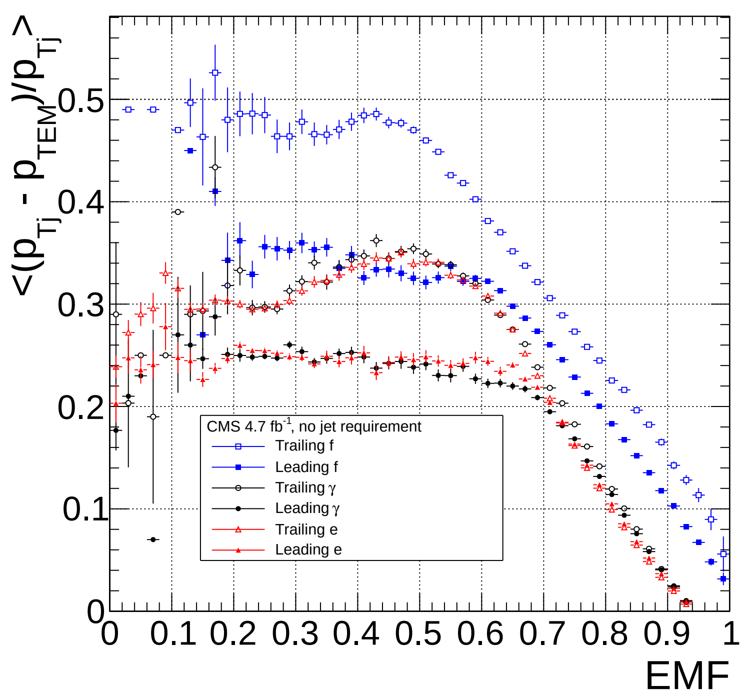
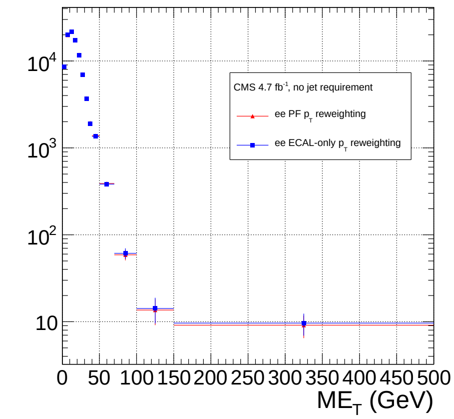
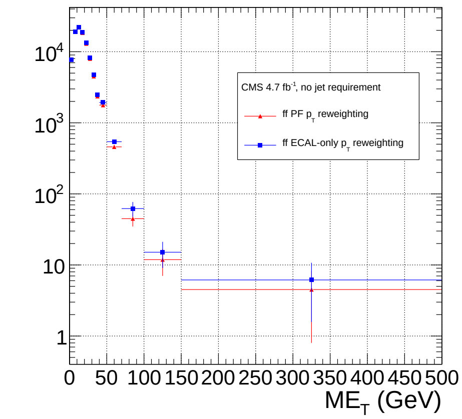


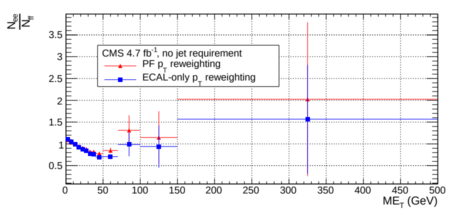
Figure 1.6: Average relative difference between the ECAL-only E_T measurement and the PF E_T measurement vs. EMF for the leading (filled marker) and trailing (open marker) electrons in ee events (red triangles), fakes in ff events (blue squares), and photons in $\gamma\gamma$ events (black circles). These are nothing more than profile histograms of Fig. 1.5. PF E_T is defined in the text. Error bars are statistical only.



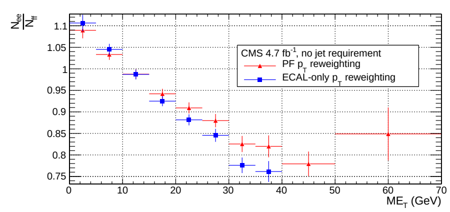
(a) ee $\not\!\!E_T$ spectra.



(b) ff $\not\!\!E_T$ spectra.

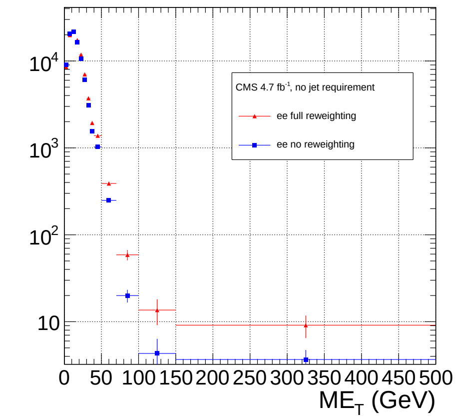


(c) Ratio of ee to ff $\not\!\!E_T$ spectra.

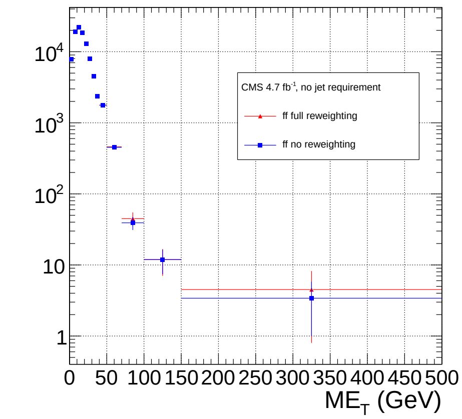


(d) Ratio of ee to ff $\not\!\!E_T$ spectra, zoomed x-axis.

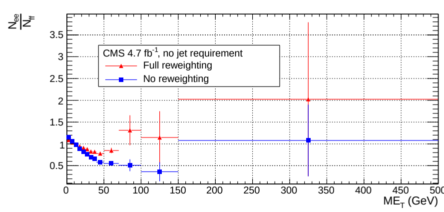
Figure 1.7: $\not\!\!E_T$ spectra of the reweighted ee (81 GeV $\leq m_{\text{ee}} <$ 101 GeV) and ff control samples. Blue squares indicate reweighting using the ECAL-only p_T estimate; red triangles indicate reweighting using the PF p_T estimate. The full reweighting and normalization procedure is employed, along with ee sideband subtraction (discussed at the end of this section). Error bars include statistical, reweighting, and normalization error (see Sec. **??**).



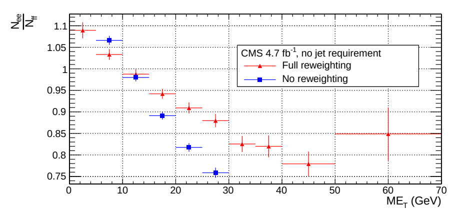
(a) *ee* $\not\!\!E_T$ spectra.



(b) *ff* $\not\!\!E_T$ spectra.



(c) Ratio of *ee* to *ff* $\not\!\!E_T$ spectra.



(d) Ratio of *ee* to *ff* $\not\!\!E_T$ spectra, zoomed x-axis.

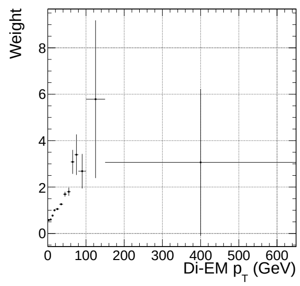
Figure 1.8: $\not\!\!E_T$ spectra of the *ee* (81 GeV $\leq m_{\text{ee}} <$ 101 GeV) and *ff* control samples. Red triangles indicate full di-EM p_T + number of jets reweighting; blue squares indicate no reweighting. PF p_T (cf. p. 11) is used to calculate the di-EM p_T. The full normalization procedure 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. **??**).

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 GeV $\leq m_{\mathrm{ee}} <$ 81 GeV and 101 GeV $\leq m_{\mathrm{ee}} <$ 111 GeV.

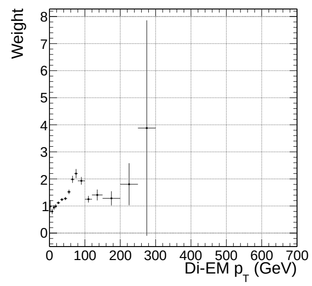
The full reweighting procedure is applied to the Z signal region and the two sideband regions independently. Only Z signal events are used in the calculation of the di-EM p_T weights for the Z signal region, and likewise only the events within a given sideband region are used in the calculation of the weights for that region. Assuming a constant $t\bar{t}$ background shape, the resulting reweighted sideband $\not{E}_T$ distributions are added together and subtracted from the reweighted Z signal $\not{E}_T$ distribution. The sideband subtracted Z signal $\not{E}_T$ distribution is then normalized as discussed in Secs. 1.1.1 and 1.1.3. The statistical and reweighting error from the sideband regions is propagated to the error on the final ee QCD $\not{E}_T$ 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 1.10 shows the $\not{E}_T$ spectra for the two sideband regions and the Z signal region after subtraction. The shapes of the spectra indicate that the high-$\not{E}_T$ $t\bar{t}$ tail, present in the sideband distributions, was successfully subtracted from the Z signal distribution.

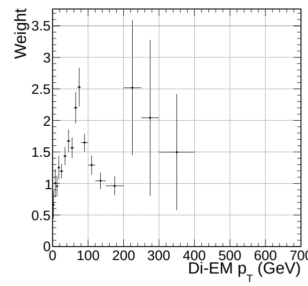
The ee (81 GeV $\leq m_{\mathrm{ee}} <$ 101 GeV), ff, and $\gamma\gamma$ di-EM p_T spectra for events with 0, 1, or ≥ 2 jets (as in Table **??**) are shown in Figure 1.11. Broad humps in the ff and $\gamma\gamma$ spectra are due to kinematic ΔR and p_T turn-ons that are suppressed in the ee sample due to the invariant mass cut. Figure 1.12 shows the weights applied to the ee (81 GeV $\leq m_{\mathrm{ee}} <$ 101 GeV) and ff $\not{E}_T$ spectra as a function of di-EM p_T and number of jets per event.



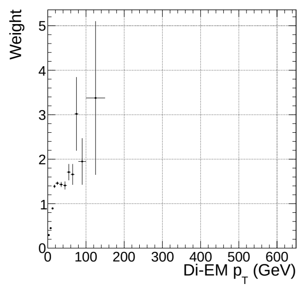
(a) 71 GeV $\leq m_{\mathrm{ee}} < 81$ GeV, 0 jets.



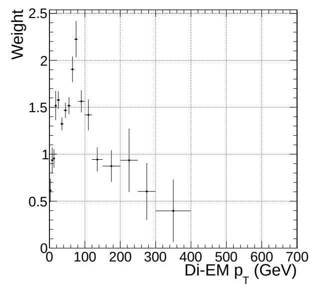
(b) 71 GeV $\leq m_{\mathrm{ee}} < 81$ GeV, 1 jet.



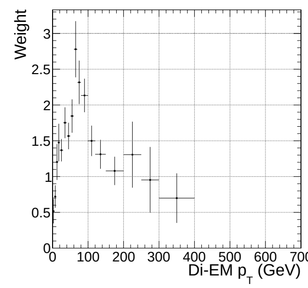
(c) 71 GeV $\leq m_{\mathrm{ee}} < 81$ GeV, ≥ 2 jets.



(d) 101 GeV $\leq m_{\mathrm{ee}} < 111$ GeV, 0 jets.



(e) 101 GeV $\leq m_{\mathrm{ee}} < 111$ GeV, 1 jet.



(f) 101 GeV $\leq m_{\mathrm{ee}} < 111$ GeV, ≥ 2 jets.

Figure 1.9: ee sideband di-EM p_T weights for events with 0, 1, or ≥ 2 jets (as in Table **??**). Errors are statistical only.

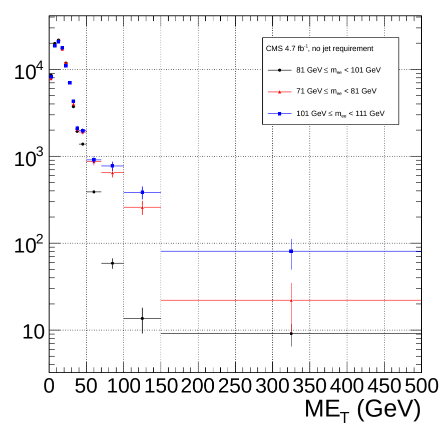
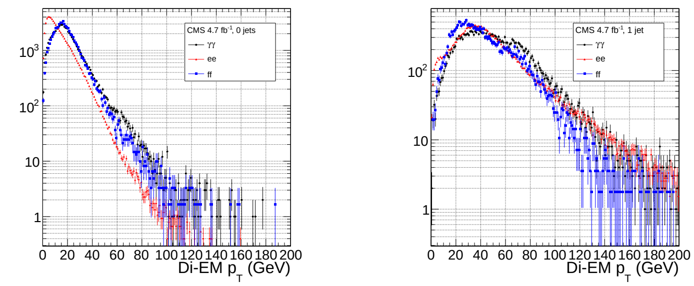
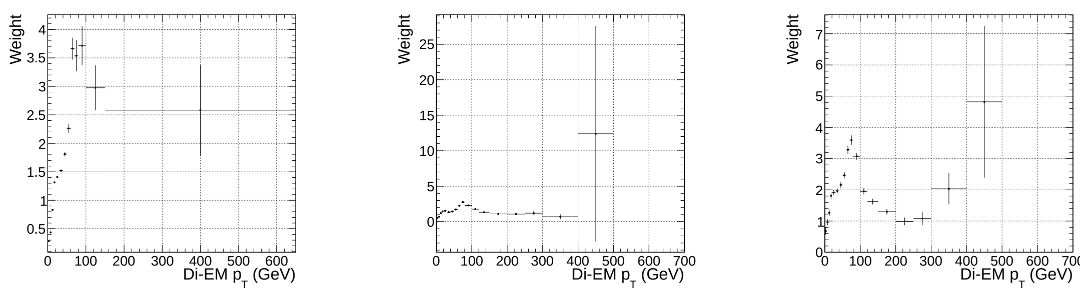


Figure 1.10: $\not\!\!E_T$ spectra of the ee sample for 71 GeV $\leq m_{\mathrm{ee}} < 81$ GeV (red triangles), 81 GeV $\leq m_{\mathrm{ee}} < 101$ GeV (black circles), and 101 GeV $\leq m_{\mathrm{ee}} < 111$ GeV (blue squares). The two sideband distributions (red and blue) and the Z signal distribution (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. **??**).

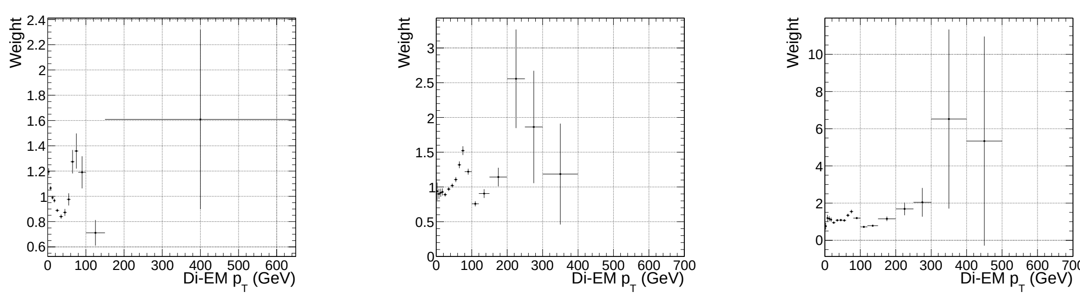


(a) 0 jets. (b) 1 jet. (c) ≥ 2 jets.

Figure 1.11: ee (81 GeV $\leq m_{\text{ee}} < 101$ GeV) (red triangles), *ff* (blue squares), and $\gamma\gamma$ (black circles) di-EM p_T spectra for events with 0, 1, or ≥ 2 jets (as in Table **??**). Errors are statistical only. **Zoom out the x-axis to show the full tail out to 500 GeV?**



(a) ee, 0 jets. (b) ee, 1 jet. (c) ee, ≥ 2 jets.



(d) *ff*, 0 jets. (e) *ff*, 1 jet. (f) *ff*, ≥ 2 jets.

Figure 1.12: ee (81 GeV $\leq m_{\text{ee}} < 101$ GeV) and *ff* di-EM p_T weights for events with 0, 1, or ≥ 2 jets (as in Table **??**). Errors are statistical only. **Zoom in the x-axis to hide large weights with large statistical errors?**

1.1.3 Normalization

After reweighting, the $\not{E}_T$ distributions of the QCD control samples are normalized to the $\not{E}_T < 20$ GeV region of the candidate $\gamma\gamma$ $\not{E}_T$ spectrum, where signal contamination is low. The normalization factor is $(N_{\gamma\gamma}^{\not{E}_T<20\text{GeV}} - N_{e\gamma}^{\not{E}_T<20\text{GeV}})/N_{\text{control}}^{\not{E}_T<20\text{GeV}}$, where $N_{e\gamma}^{\not{E}_T<20\text{GeV}}$ is the expected number of electroweak background events with $\not{E}_T < 20$ GeV (discussed in Section 1.2).

1.2 Modeling the Electroweak Background

$W\gamma$, W + 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-$\not{E}_T$ tail of the $\gamma\gamma$ $\not{E}_T$ spectrum. To estimate this background, the $e\gamma$ sample, which is enriched in $W \to e\nu$ decays, is scaled by $f_{e\to\gamma}/(1-f_{e\to\gamma})$, where $f_{e\to\gamma}$ is the rate at which electrons are misidentified as photons. The derivation of this scaling factor comes from the two equations

$$N_{e\gamma}^{W} = f_{e\to e} N_W \tag{1.2}$$

$$N_{\gamma\gamma}^{W} = (1 - f_{e\to e}) N_W \tag{1.3}$$

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

$$N^W_{\gamma\gamma} = \frac{f_{e\to\gamma}}{1 - f_{e\to\gamma}} N^W_{e\gamma} \tag{1.4}$$

$f_{e\to\gamma}$ 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

$$N^Z_{ee} = (1 - f_{e\to\gamma})^2 N_Z \tag{1.5}$$

$$N^Z_{e\gamma} = 2f_{e\to\gamma}(1 - f_{e\to\gamma}) N_Z \tag{1.6}$$

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

$$f_{e\to\gamma} = \frac{N^Z_{e\gamma}}{2N^Z_{ee} + N^Z_{e\gamma}} \tag{1.7}$$

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. **??**) 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 1.3 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 $f_{e\to\gamma}$ on the electron p_T and η is small. Applying a p_T- and η-dependent misidentification rate (with p_T and η 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

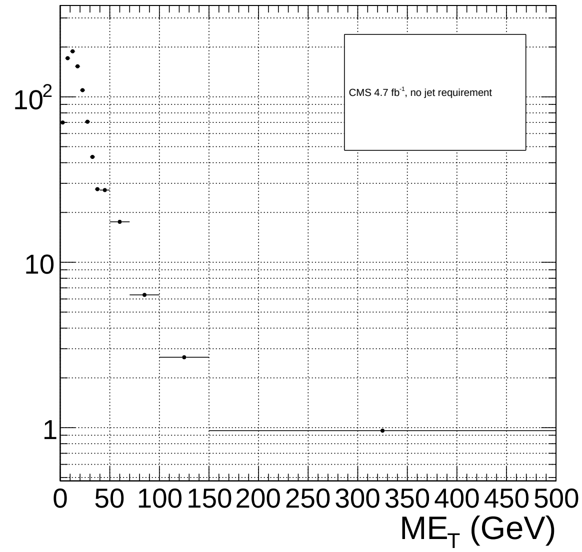


Figure 1.13: $\not\!\!E_T$ spectrum of the $e\gamma$ sample after scaling by $f_{e\to\gamma}$. The total error on $f_{e\to\gamma}$ is propagated to the total error on the electroweak background estimate. **How to properly treat the error when the same events are used in the $f_{e\to\gamma}$ calculation and in the $e\gamma$ sample? Replace with figure using latest $f_{e\to\gamma}$, and include error bars.**

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 $f_{e\to\gamma}$. **Check whether the p_T dependence or the shapes yield the larger error.** To assess the magnitude of this error, $f_{e\to\gamma}$ is recalculated using both linear and quadratic background shapes, and with a Crystal Ball + generated Z signal shape (as used in Sec. **??**). 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. 1.7, and the shape and p_T systematics discussed above, $f_{e\to\gamma}$ 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 1.13.

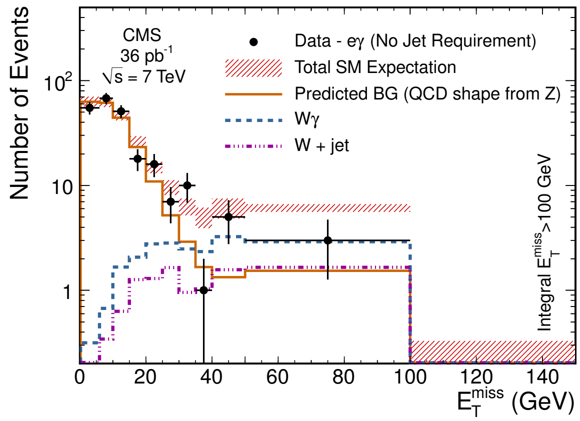


Figure 1.14: $\not\!\!E_T$ spectrum of the $e\gamma$ sample in 35 pb^{-1} of 2010 LHC data scaled by the 2010 measured $f_{e\to\gamma}$ (black dots), QCD and real Z predicted background from the 2010 ee sample (solid orange line), MC W + jet estimate (dash-dotted purple line), and MC $W\gamma$ estimate (dashed blue line). The total $e\gamma$ prediction (red band) is the sum of the ee, W + jet, and $W\gamma$ predictions. Reprinted from Fig. 2 of ref. [?].

In the 36 pb^{-1} version of this analysis [?], it was shown that the ee sample could accurately predict the QCD and real Z contribution to the $e\gamma$ sample at low $\not\!\!E_T$, and that the expectation from $W \to e\nu$ MC accounted for the remaining W contribution at high $\not\!\!E_T$. A plot of the $\not\!\!E_T$ distributions of the 2010 $e\gamma$ sample and the predicted components is shown in Figure 1.14. **Repeat for current selection?** This exercise helps to validate both the QCD and electroweak background prediction methods.

1.3 Errors on the Background Prediction

The statistical error on the final background estimate in a particular $\not\!\!E_T$ 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 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 $\not{E}_T$ bin. Since the distribution of the toy background estimates follows a Gaussian distribution in each $\not{E}_T$ 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 $\not{E}_T$ bin is given by

$$\tag{1.8}$$

where **Fill in at work.** For $\not{E}_T > 100$ 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* $\not{E}_T$ distribution towards lower values with respect to the *ff* and $\gamma\gamma$ samples, as shown in Figure 1.15. 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 $\not{E}_T$ $>$ 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 $e \to \gamma$ misidentification rate (see 1.2). For $\not{E}_T > 100$ GeV, the expected electroweak background is 3.6 $\pm$ **XXX** events, so this error amounts to **XXX**% of the total QCD

Table 1.3: 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.**

	Crystal Ball fit parameters				RooCMSShape fit parameters			
PDF	μ	σ	α	n	μ	α	β	γ
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	α	[50.0, 100.0]	0.065	0.048

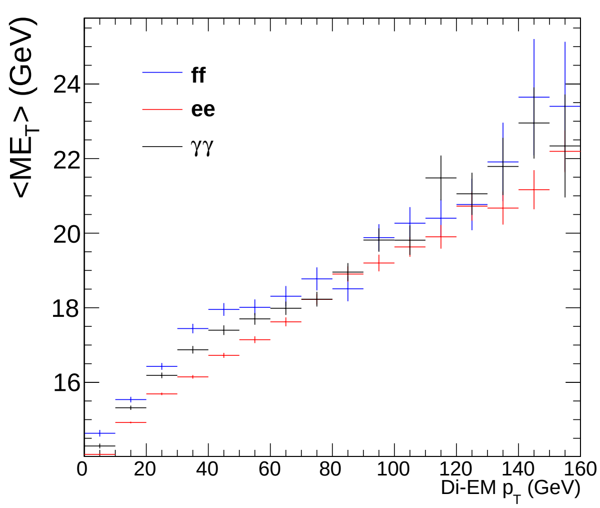


Figure 1.15: Average $\not{E}_T$ vs. di-EM p_T for the *ff* (blue), ee (red), and $\gamma\gamma$ (black) samples.

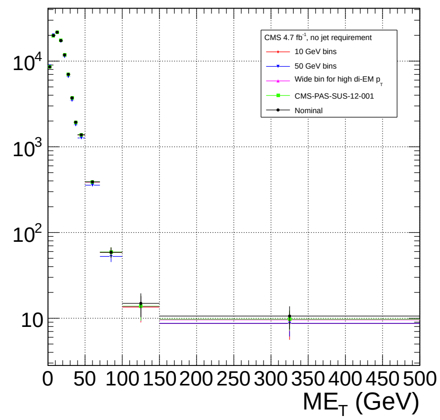
+ electroweak background estimate.

Finally, the assumption of a constant $t\bar{t}$ and W + jets background shape under 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. 1.1.2) 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 $\not\!\!E_T > 100$ GeV. **Is this the right way to do this?**

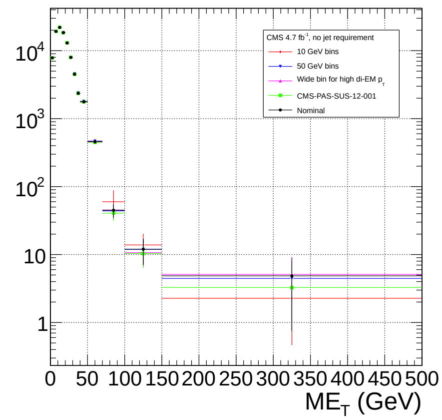
The uncertainty in how to define the (di-EM p_T, N_{j}) bins, especially at high di-EM p_T where the statistics are low, is covered by the 1000-toys procedure as long as the bins are not too coarse. This is shown in Figure **??**. If the bins are too coarse, the details of the shape of the di-EM p_T spectra are lost, and the reweighting has a smaller effect.

The use of uncorrected instead of corrected PF $\not\!\!E_T$ (see Sec. **??**) makes no difference in the agreement of the background predictions and the search sample in a control region at low $\not\!\!E_T$, as shown in Figure **??**. Since the control samples are derived from the same data as the search sample, any biases in the $\not\!\!E_T$ reconstruction due to jet energy scale are present equally in both samples.

Tables **??** and **??** list all the errors on the ee and $f\!f$ QCD background predictions, respectively, for the $\not\!\!E_T$ bins used in the search. Table **??** lists the errors on the 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 $f\!f$). In the final result, only the $f\!f$ QCD estimate is used.

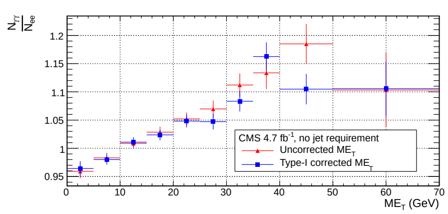


(a) ee.

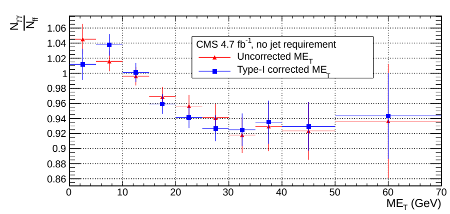


(b) ff.

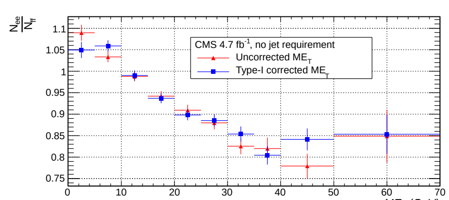
Figure 1.16: Comparison of $\not\!\!E_T$ distributions for five different di-EM p_T bin definitions: uniform bins of width 10 GeV (red diamonds); uniform bins of width 50 GeV (blue downward-pointing triangles); bins with lower edges {0.0, 5.0, 10.0, 15.0, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 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, 120.0, 150.0, 200.0, 700.0} 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} GeV for $\geq$1-jet events (green squares), i.e. the bins used in ref. [**?**]; and the nominal bin definitions shown in Fig. 1.12 (black circles).



(a) Ratio of $\gamma\gamma$ to ee $\not\!\!E_T$ distributions.



(b) Ratio of $\gamma\gamma$ to ff $\not\!\!E_T$ distributions.



(c) Ratio of ee to ff $\not\!\!E_T$ distributions.

Figure 1.17: Agreement between $\gamma\gamma$, ee, and ff samples for uncorrected (red triangles) and corrected (blue squares) $\not\!\!E_T$.

Table 1.4: 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)	≥100
Total	3.6	7.8	16	24	22
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	0.0015	0.0015	0.0016	0.0019	0.0018
$f_{e\rightarrow\gamma}$ (in norm. region)	0.0015	0.0015	0.0016	0.0019	0.0018
m_{ee} background shape	0	0	0	0	0

Table 1.5: 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)	≥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 $\not\!\!E_T$ 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 $\not\!\!E_T$ 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 1.6: 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)	≥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

1.4 Results

Table 1.7: 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)	≥100
Total (ee + $e\gamma$)	3.7	7.6	15	21	20
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	1.3	2.1	2.9	4	4.2
QCD	0.0014	0.0014	0.0015	0.0017	0.0015
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