Chapter 1

Data Analysis

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.

Table 1.1: 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	$> 40/$ > 25 GeV	$> 40/$ > 25 GeV	$> 40/$ > 25 GeV
SC $\lvert\eta\rvert$	< 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

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. 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 $W\gamma$ and W + jet 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. 1.3 and a presentation of the final results in Sec. 1.4.

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.

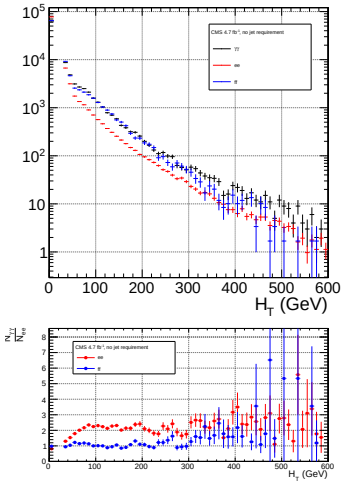
Normalization Finally, the control $\not\!\!E_T$ distribution is normalized in a region of

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.

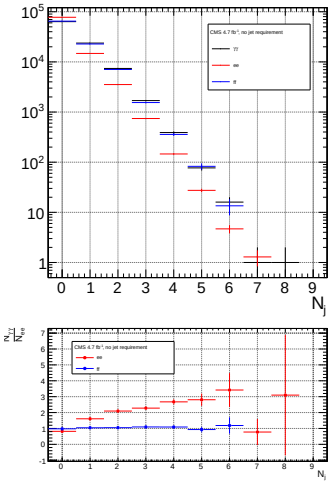
As explained in the beginning of this chapter, the *ff* sample is used as the primary QCD control sample, while the *ee* sample is used as a cross-check. Both samples have two well-measured EM objects per event, no real $\not\!\!E_T$, and similar hadronic activity to the $\gamma\gamma$ sample. Figure 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 1.2: Definition of hadronic jets. **Add a footnote describing the PF electron and PF muon definitions, with references.**

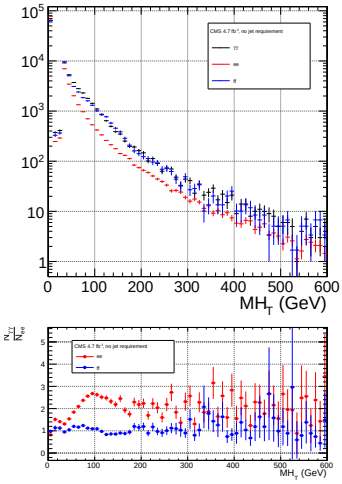
Variable	Cut
Algorithm	`L1FastL2L3Residual` corrected PF (cf. Sec. **??**)
p_T	> 30 GeV
$\|\eta\|$	< 5.0
Neutral hadronic energy fraction	< 0.99
Neutral electromagnetic energy fraction	< 0.99
Number of constituents	> 1
Charged hadronic energy	> 0.0 GeV if $\|\eta\| < 2.4$
Number of charged hadrons	> 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



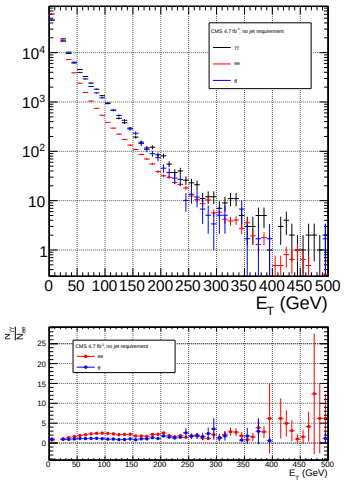
(a) H_T, defined as the scalar sum of corrected jet E_T for jets defined as in Table 1.2.



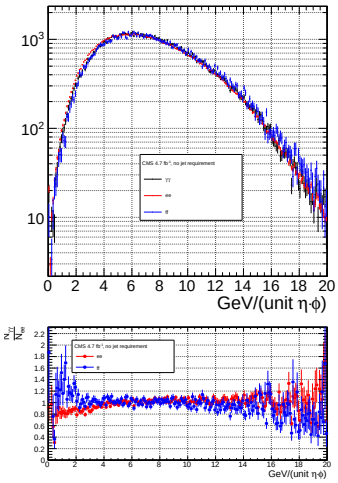
(b) Number of jets per event for jets defined as in Table 1.2.



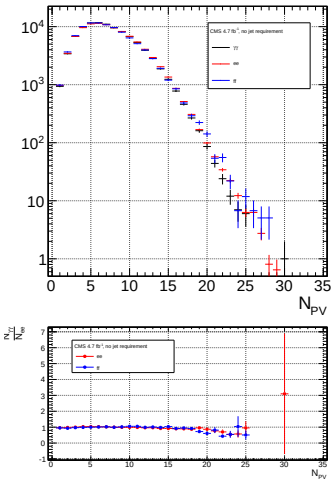
(c) $\not{H}_T$, defined as the magnitude of the negative vectorial sum of corrected jet E_T for jets defined as in Table 1.2.



(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, and ff samples. The ee and ff distributions are normalized to the number of events in the $\gamma\gamma$ distribution. Errors are statistical only.

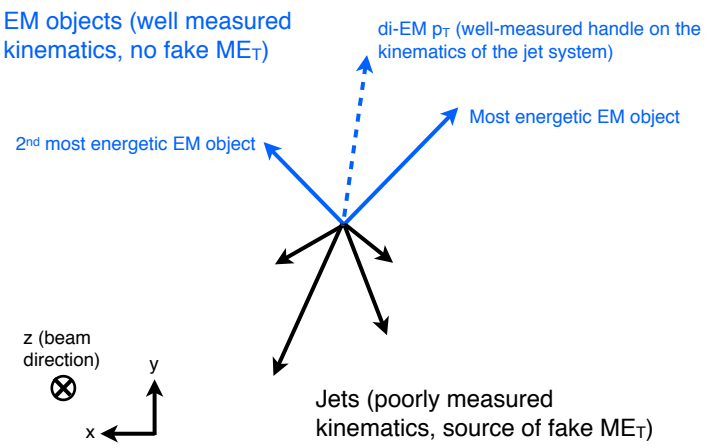


Figure 1.2: Cartoon showing the di-EM system in blue and the hadronic recoil in black. The di-EM p_T (dashed blue line) is used to reweight the control sample kinematic properties to match those of the candidate $\gamma\gamma$ sample.

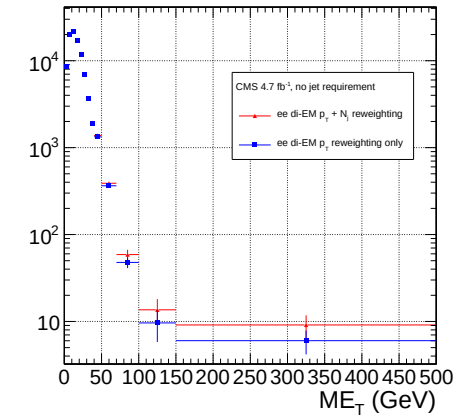
.

1.1.2 Reweighting

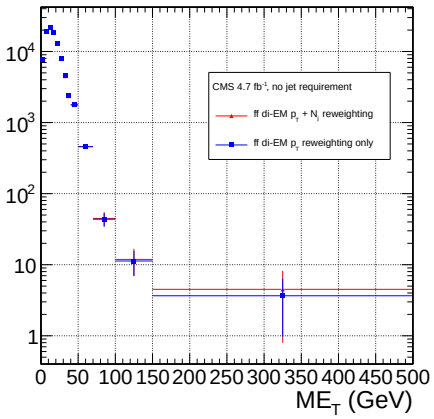
To reweight the control sample events to match the kinematics of the candidate sample events, a weight based on the p_T of the di-EM-object system and the number of jets in the event is used. As explained in Sec. 1.1.1, $\not\!\!E_T$ in the $\gamma\gamma$, ff, and ee samples is due to the poorly measured 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.

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 1.2, but with $|\eta|$ restricted to 2.6 (i.e. HF jets excluded). Figure 1.3 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.

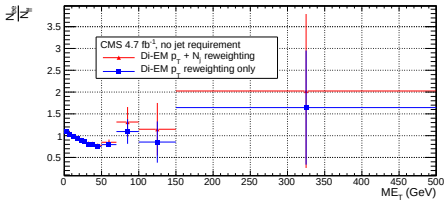
Although the electron and photon energies are well measured by the ECAL, the



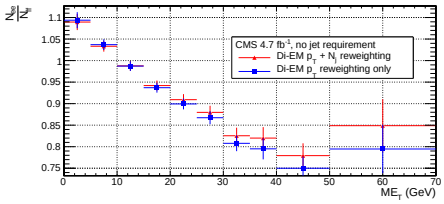
(a) ee $\not\!\!E_T$ spectra.



(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 1.3: $\not\!\!E_T$ spectra of the reweighted ee 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. 10) is used to calculate the di-EM p_T. The full normalization procedure is employed, along with ee sideband subtraction (discussed in Sec. **??**). Error bars include statistical, reweighting, and normalization error (see Sec. 1.3).

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. 1.4 and 1.5, 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. This leads to a modest improvement in the agreement between the *ee* and *ff* $\not\!\!E_T$ spectra, as shown in Figure 1.6.

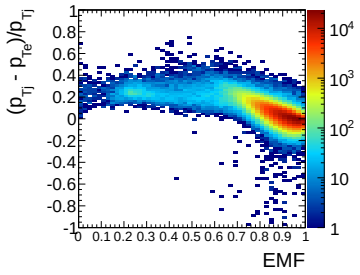
The control sample event weights are defined as

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

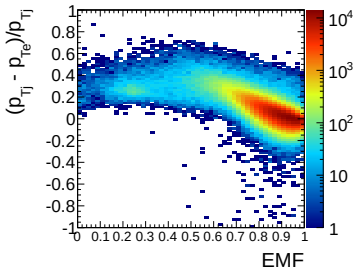
where i runs over the number of di-EM p_T bins, j runs over the number of jet bins, N_{control} is the total number of events in the control sample, $N_{\gamma\gamma}$ is the total number of events in the $\gamma\gamma$ sample, $N^{ij}_{\gamma\gamma}$ is the number of $\gamma\gamma$ events in the i^{th} 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 1.7.

1.1.3 Normalization

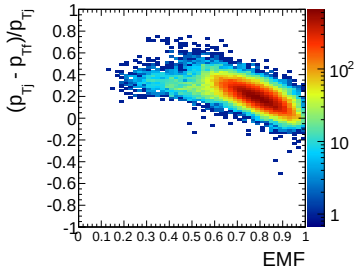
After reweighting, the $\not\!\!E_T$ distributions of the QCD control samples are normalized to the $\not\!\!E_T < 20$ GeV region of the candidate $\gamma\gamma$ $\not\!\!E_T$ spectrum, where signal contamination is low. The normalization factor is $(N^{\not\!\!E_T<20\text{GeV}}_{\gamma\gamma} - N^{\not\!\!E_T<20\text{GeV}}_{\text{electroweak}})/N^{\not\!\!E_T<20\text{GeV}}_{\text{control}}$, where



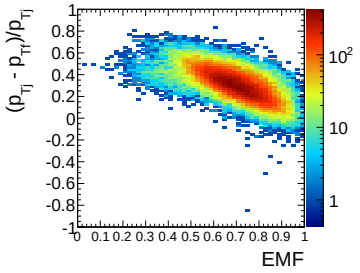
(a) Leading electron in ee events.



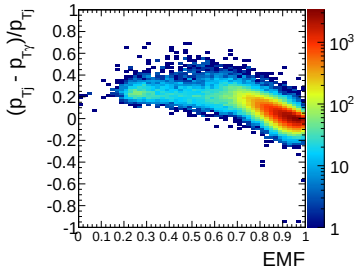
(b) Trailing electron in ee events.



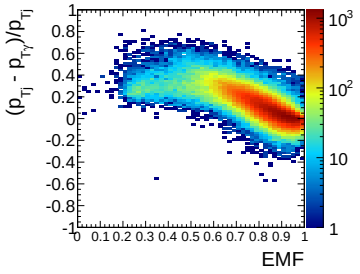
(c) Leading fake in ff events.



(d) Trailing fake in ff events.



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



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

Figure 1.4: 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.

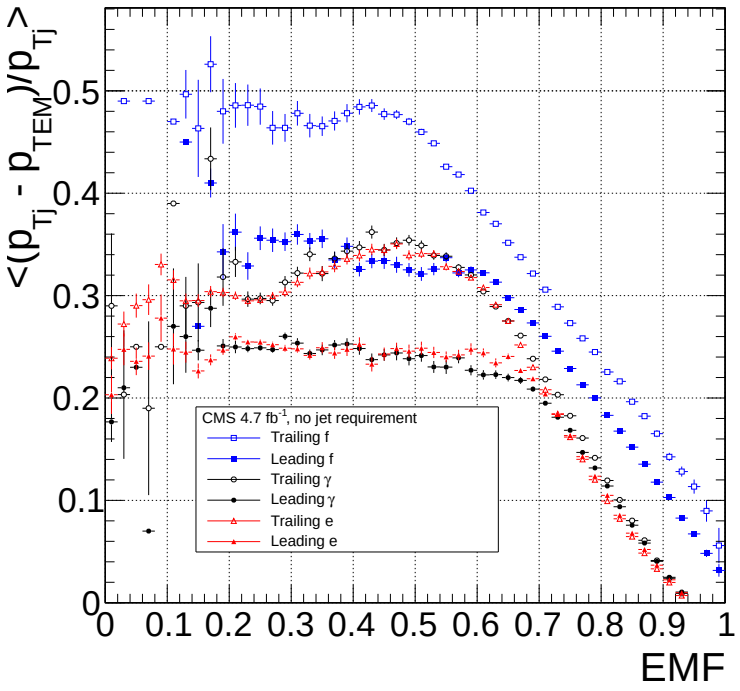
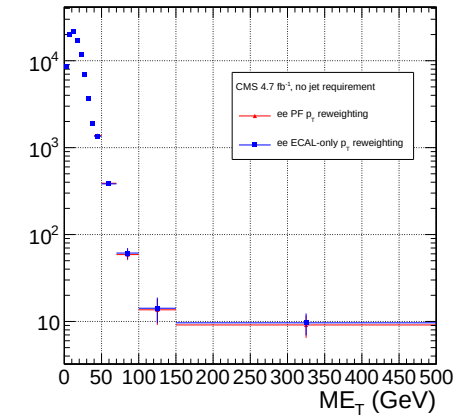
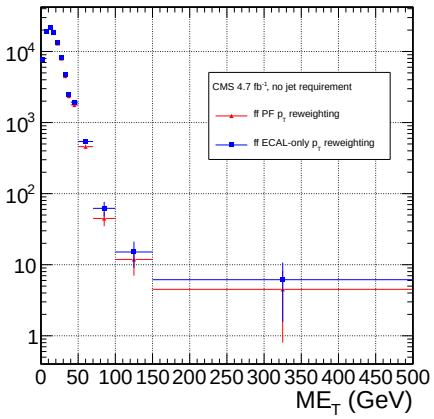


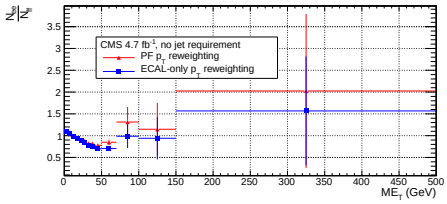
Figure 1.5: Average relative difference between the ECAL-only E_T measurement and the PF E_T measurement vs. EMF for the leading (filled marker) and trailing (open marker) electrons in ee events (red triangles), fakes in ff events (blue squares), and photons in $\gamma\gamma$ events (black circles). These are nothing more than profile histograms of Fig. 1.4. PF E_T is defined in the text. Error bars are statistical only.



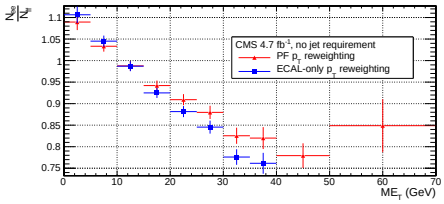
(a) ee $\not\!\!E_T$ spectra.



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

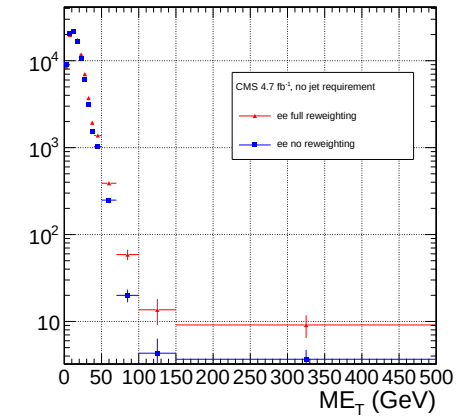


(c) Ratio of ee to ff $\not\!\!E_T$ spectra.

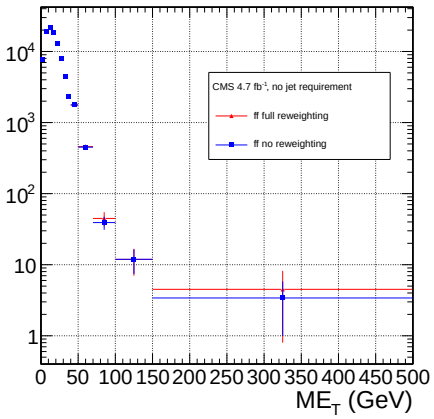


(d) Ratio of ee to ff $\not\!\!E_T$ spectra, zoomed x-axis.

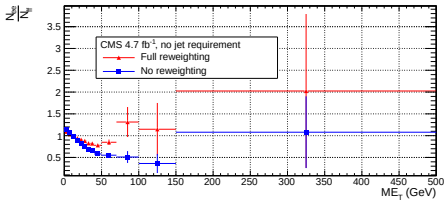
Figure 1.6: $\not\!\!E_T$ spectra of the reweighted ee and ff control samples. Blue squares indicate reweighting using the ECAL-only p_T estimate; red triangles indicate reweighting using the PF p_T estimate. The full reweighting and normalization procedure is employed, along with ee sideband subtraction (discussed in Sec. **??**). Error bars include statistical, reweighting, and normalization error (see Sec. 1.3).



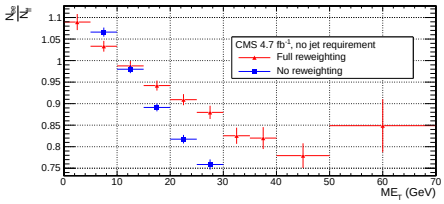
(a) ee $\not\!\!E_T$ spectra.



(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 ee 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. 10) is used to calculate the di-EM p_T. The full normalization procedure is employed, along with ee sideband subtraction (discussed in Sec. **??**). Error bars include statistical, reweighting (where appropriate), and normalization error (see Sec. 1.3).

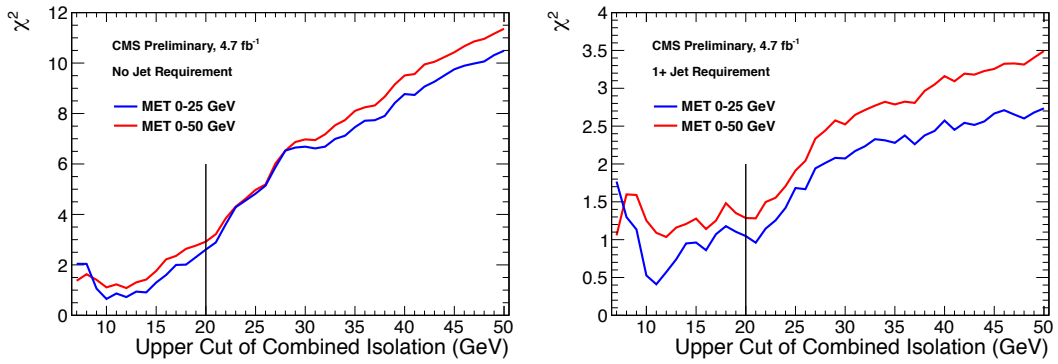


Figure 1.8: Neyman's χ^2 between the ff and $\gamma\gamma$ $\not\!\!E_T$ distributions, truncated at either 25 (red) or 50 (blue) GeV, vs. upper bound on fake combined isolation. The left plot includes all events; the right plot is for events with ≥ 1 jet similar to Table 1.2, but with $|\eta| \leq 2.6$ instead of 5.0, and the ΔR cleaning criteria applied to the two primary EM objects and all additional electrons, photons, and fake photons (instead of PF electrons and muons). The full reweighting and normalization procedure is employed in the $\not\!\!E_T$ calculation. Error bars include statistical, reweighting, and normalization error (see Sec. 1.3). Reprinted from Fig. 9 of ref. [**?**].

$N_{\text{electroweak}}^{\not\!\!E_T<20\text{GeV}}$ is the expected number of electroweak background events with $\not\!\!E_T < 20$ GeV (discussed in Section 1.2).

1.1.4 ff Control Sample

The upper bound on fake photon combined isolation guarantees that poorly isolated dijet events, with $\not\!\!E_T$ resolution dissimilar to the candidate diphoton events, do not enter the ff sample. The exact value of 20 GeV arises from a low-$\not\!\!E_T$ $ff/\gamma\gamma$ χ^2 optimization procedure [**?**]. Figure 1.8 shows the value of the Neyman's χ^2 between the ff and $\gamma\gamma$ $\not\!\!E_T$ 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 χ^2, while also being large enough that a sufficient number of ff events may be collected.

1.2 Modeling the Electroweak Background

1.3 Systematic Errors

1.4 Results