Chapter 4

Data Sample and Selection Efficiencies

The candidate data sample was collected during the 1994-1995 run (referred to as run 1B, see Fig. 4.1), corresponding to an integrated luminosity L = 82.36 $\pm$ 4.40 pb^{-1}. This chapter will describe the event selection, the trigger, the acceptance and the background subtraction.

4.1 Event Selection

Because of some problems with the hit information of the central tracking chamber CDC in *RECO*, this analysis needs STA-type data(STA refers to standard data, including all the information) which are not available on disk. Only the micro-DST type data(DST refers to data summary type data, including compressed physics data; micro-DST is further compressed from DST) exists on disk. In order to get a data sample with one *RECO* version, a pre_candidate selection is performed on all 1B data. The resulting sample is re-reconstructed using one *RECO* version 12.20. The procedures are as follows:

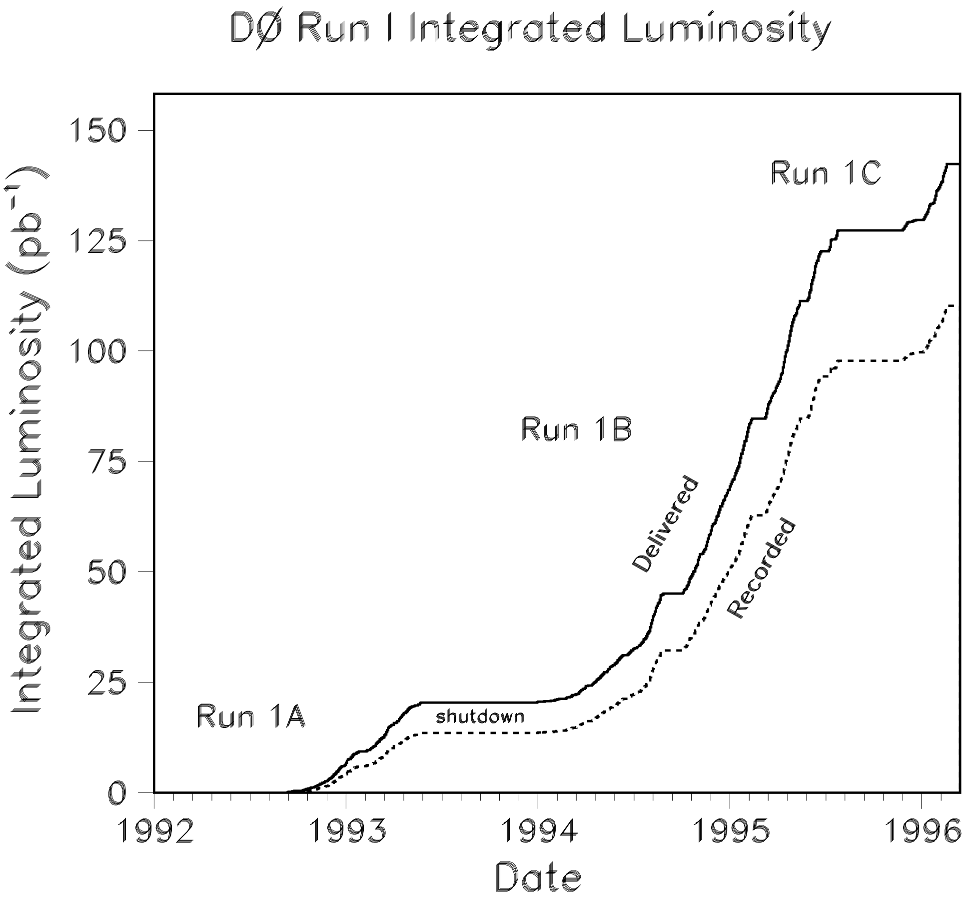

Figure 4.1: Accumulated Luminosity during Run 1

1. Ntuples are created using the information available on the disk-resident micro-DST data. A pre_candidate sample is selected by applying loose cuts:

Number of Photons(PPHO) ≥ 2
Each Photon satisfies the following cuts
Detector $\eta_{det} < 1.1$
$E_T > 10.0$ Gev
Isolation Energy between $R = 0.4$ and $R = 0.2 < 4.0$ Gev
Calorimeter Shower Hmatrix $\chi^2 < 250.$

2. A list is generated with run and event number of all the events which have passed the loose selection cuts. Their STA-type data are picked up by hand. Total number of pre_candidates is **3477**.

3. All the picked events are then re-reconstructed using the *RECO* version 12.20. The output sample is ready for analysis.

4.1.1 Trigger Requirements

Because of the huge EM jet background to photons and the limited rate DØ DAQ (DAQ refers to data acquisition system) can handle (≈ 2 Hz), rather sophisticated algorithms have been developed to trigger on high P_T photons efficiently while keeping the rate manageable.

Hardware Trigger, Level Ø and Level 1

The hardware trigger systems are described in Sections 2.6.1 and 2.6.2. For the double-photon sample, the level Ø scintillator counters are required to detect a $p\overline{p}$ collision. There must be no or very low main ring activity when the data are recorded. Two types of main ring vetoes are available [93]; one, called MRBS_LOSS, vetoes events during the main ring cycle from injection through transition, and the other, called MICRO_BLANK, vetoes the time when the main ring batches pass through the DØ detector in coincidence with a Tevatron bunch crossing. The double-photon trigger was only blanked when both MRBS_LOSS and MICRO_BLANK vetoes were satisfied.

After the level-Ø, the level 1 calorimeter trigger makes a fast sum over all cells within each trigger tower (its size being $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$). The EM energy sum of one trigger tower must be above the threshold $E_T^{thr} = 7.0$ GeV before it becomes a candidate. The number of candidate towers must be ≥ 2. The term for this level 1 configuration is *em_2_med*.

Software Filter Level 2

If an event has fired level 1, it is transferred to a large farm of processors, i.e. Level 2 (see Section 2.6.3). Sophisticated software algorithms are applied to the data in order to select interesting events at the rate ≈ 2 Hz. The selection requirements for double-photon candidates form a level-2 term called *em2_gis_gam*. Its prescription is described as follows:

Hardware Level Ø and Level 1 Trigger	
Level Ø	Level 1
detect one or more $p\overline{p}$ collision MRBS_LOSS.and.MICRO_BLANK	*em_2_med* num EM trigger towers ≥ 2 each trigger tower $E_T^{thr} > 7.0$ GeV
Software Level 2(L2) Filter	
em2_gis_gam num L2 EM clusters ≥ 2 each cluster $E_T^{thr} > 12.0$ GeV	
one L2 EM cluster	the 2nd L2 EM cluster
photon_shape isolation	photon_shape

Table 4.1: Trigger Requirements for the Direct Double-Photon Data Sample

- At level 2, an EM object starts from what has passed the level 1 EM trigger tower. The number of EM objects must be two or more.

- Both EM objects have to have $E_T^{thr} \geq 12\ GeV$.

- An isolation cone is built around each object with the core cone $R = 0.2$ and the isolation cone $R = 0.4$, where $R = \sqrt{\eta^2 + \phi^2}$. At least one object must satisfy the isolation requirement, which is defined as:

$$\frac{E(R=0.4) - E(R=0.2)}{E(R=0.2)} < 15\%$$

- Longitudinal and transverse photon shape cuts are required for both EM objects, where the photon_shape cuts are as follows:

1. The fraction of the energy in the third EM layor(EM3) within 3×3 readout towers is required to satisfy $0.1<$ EM3 <0.9.
2. The fraction of the energy in the first hadronic layor(FH1) within 3×3 readout towers is required to satisfy FH1 $<$ cut_offset
3. Transverse shower quantities as energy weighted shower radia, SIGMA3 and SIGMA5, are calculated for EM3 cells within 3×3 and 5×5 windows. They are required to satisfy

   $$\text{SIGMA5 - SIGMA3} \leq \text{cut\_offset}$$

   where the cut_offset is tuned for several energy bins and eta bins.

All the cuts listed above are tuned to be $\geq 98\%$ efficient for testbeam electrons. The trigger requirements are summarized in Table 4.1.

4.1.2 Offline Cuts

After events pass the Level 2 filters, they are further selected by offline cuts. Using the nearest-neighbor(NN)algorithm, preliminary photon(PPHO) and electron(PELC) candidates are found (see Section 3.4.1). The number of PPHOs must be ≥ 2. Each PPHO is required to satisfy fudicial cuts, kinematic cuts and photon ID cuts, as shown in Table 4.2. The number of candidates found after cuts is **333**.

Offline Selection Cuts for Each PPHO
fudicial cuts
$\|\eta_{det}\| < 1.0, \quad \|\eta\| < 1.0$ $\frac{min(\|\phi_\gamma - \phi^i_{crack}\|,\|\phi^{i+1}_{crack} - \phi_\gamma\|)^a}{2\pi/32} > 0.05$
kinematic cuts
$E^1_T > 14.0$ GeV, $E^2_T > 13.0$ GeV
photon ID cuts (see Section 3.4.2)
Hmatrix $\chi^2 < 100$ $E_T^{iso-4} < 2.0\ GeV$ .and. $E_T^{iso-7} < 7.0\ GeV$ RHCDCW < 0.5, NHCDC3D < 1

Table 4.2: Offline Selection Requirements for Double-photon sample

[a] ϕ_{crack} refers to the cracks between the calorimeter modules, see Section 4.2.3

4.2 Geometric Acceptance

Theoretical predictions usually assume perfect detectors with no crack and edge losses, whereas the measured results, using a particular apparatus, always have these effects folded in. When one wants to compare experimental measurements to theoretical predictions, these effects have to be corrected in measurements, or in the theoretical calculations, so that one is comparing apples to apples. The measured data are usually corrected for the losses due to the apparatus defects; a quantity, called *Geometric Acceptance* (A), is defined to account for them:

$$\mathcal{A} = \frac{\text{Number of Events Detected by the Detectors}}{\text{Number of Events Incident onto the Detectors}} \tag{4.1}$$

4.2.1 Event Generator: PYTHIA

To describe the topology of direct double-photon events reasonably well, we chose PYTHIA [94, 95] as the Monte Carlo event generator. The physics processes contributing in PYTHIA are:

$$q\,q \;\rightarrow\; \gamma\gamma \tag{4.2}$$

$$g\,g \;\rightarrow\; \gamma\gamma \tag{4.3}$$

where the cross sections for the processes (4.2) and (4.3) are at the order of $\alpha_{EM}^2\alpha_s^0$ and $\alpha_{EM}^2\alpha_s^2$ respectively. The PDF used is CTEQ2M, both initial and final state radiation are turned on, and the emerging quarks and gluons are allowed to fragment after the hard collision.

Later, we will discuss more a complete NLO QCD calculation [35] and a resummed calculation, RESBOS [28]. As far as the ϕ and η distributions and their correlations are concerned, PYTHIA provides a close description of what the resummed calculation(RESBOS) does. The difference in the acceptance between PYTHIA and RESBOS are $\leq 0.5\%$.

In order to enrich events in the high E_T and central rapidity region, phase space cuts, $E_T < 7.$ GeV and -3 $< \eta <$ 3, are required during event generation. After the events are generated, the z coordinate of the hard collision vertex is smeared with a gaussian of 25 cm width. Two million events were generated accordingly.

4.2.2 Detector Response Simulation

In this analysis we are only concerned with photons, and the detector response of photons can be modeled using testbeam electrons and decay electrons from $W \rightarrow e\nu$ and $Z \rightarrow e^+e^-$.

The energy resolution of the central EM calorimeter (CCEM) is parametrized as [66]:

$$\frac{\sigma_E}{E} = 0.015 \oplus \frac{0.13}{\sqrt{E_T}} \oplus \frac{0.4}{E} \tag{4.4}$$

where the sampling term $0.13/\sqrt{E_T}$ was measured in the Load2 testbeam [91], the constant term $0.015^{+0.006}_{-0.015}$ was determined directly from the observed width of the resonance $Z \rightarrow e^+e^-$, and the noise term $0.4/E$ was extracted from a full simulation including calorimeter uranium and electronic noise and underlying events contribution.

For photons, the centroids of their showers and the collision vertices determine their travel direction. Using a detailed Plate-level Monte Carlo simulation [96], the position resolutions of the shower centroid of single electrons (the same for photons) are measured as [97]:

$$\delta(r\phi) \approx 0.26 \text{ cm}, \qquad \delta(z) \approx 0.21 \text{ cm}$$

Since the (x,y) positions of the collision vertices were stable within a few percent of a centimeter during run I, the azimuthal angular resolution is dominated by the resolution of the shower centroid. Taking into account the radius of EM3, 91.6 cm, the resolution $\delta\phi$ is:

$$\delta\phi = 0.003 \; radians \tag{4.5}$$

When comparing the Monte Carlo hit position z_{MC} with the z_{cal} measured by the calorimeter, the resolution $\delta(z_{cal})$ was found to be dependent upon both the z_{cal} position and the polar angle θ_{cal}. It is parametrized as:

$$\delta(z_{cal}) = (p_1 + p_2 \times |\theta_{cal}|) + (p_3 + p_4 \times |\theta_{cal}|) \times |z_{cal}| \tag{4.6}$$

where the parameters $p_1 = 0.33183\,\mathrm{cm}$, $p_2 = 0.52281 \times 10^{-2}\,\mathrm{cm}/degree$, $p_3 = 0.41968 \times 10^{-3}$, and $p_4 = 0.75496 \times 10^{-4}/degree$.

4.2.3 Geometric Acceptance, $\mathcal{A}$

The Geometric Acceptance($\mathcal{A}$) is defined as Eqn. (4.1). Two main factors contribute to the loss that some events fail to be detected in the central calorimeter (CC). One is the cracks between CC EM modules. There are 32 EM modules arranged in a ϕ ring in total; the crack regions between neighboring modules are non-instrumented and are cut out by requiring the ϕ distance between the photon hit and the nearby cracks to be more than 5% of the module size. For the double-photon data sample, the ϕ distribution of each photon is uniform, and, assuming the two photons are loosely correlated in ϕ, the fraction of events surviving after the ϕ crack cuts $\approx (90\%)^2 \approx 80\%$.

The other factor is the vertex smearing. The CC calorimeter physics η spans between -1.2 and 1.2 from the nominal vertex position $z_{vert} = 0.0$. When the collision vertex z moves around the nominal position $z = 0.0$, the CC effective η span changes so that some photons travel out of the CC coverage and cannot be detected. This results in losing $\approx 15\%$ of double-photon events.

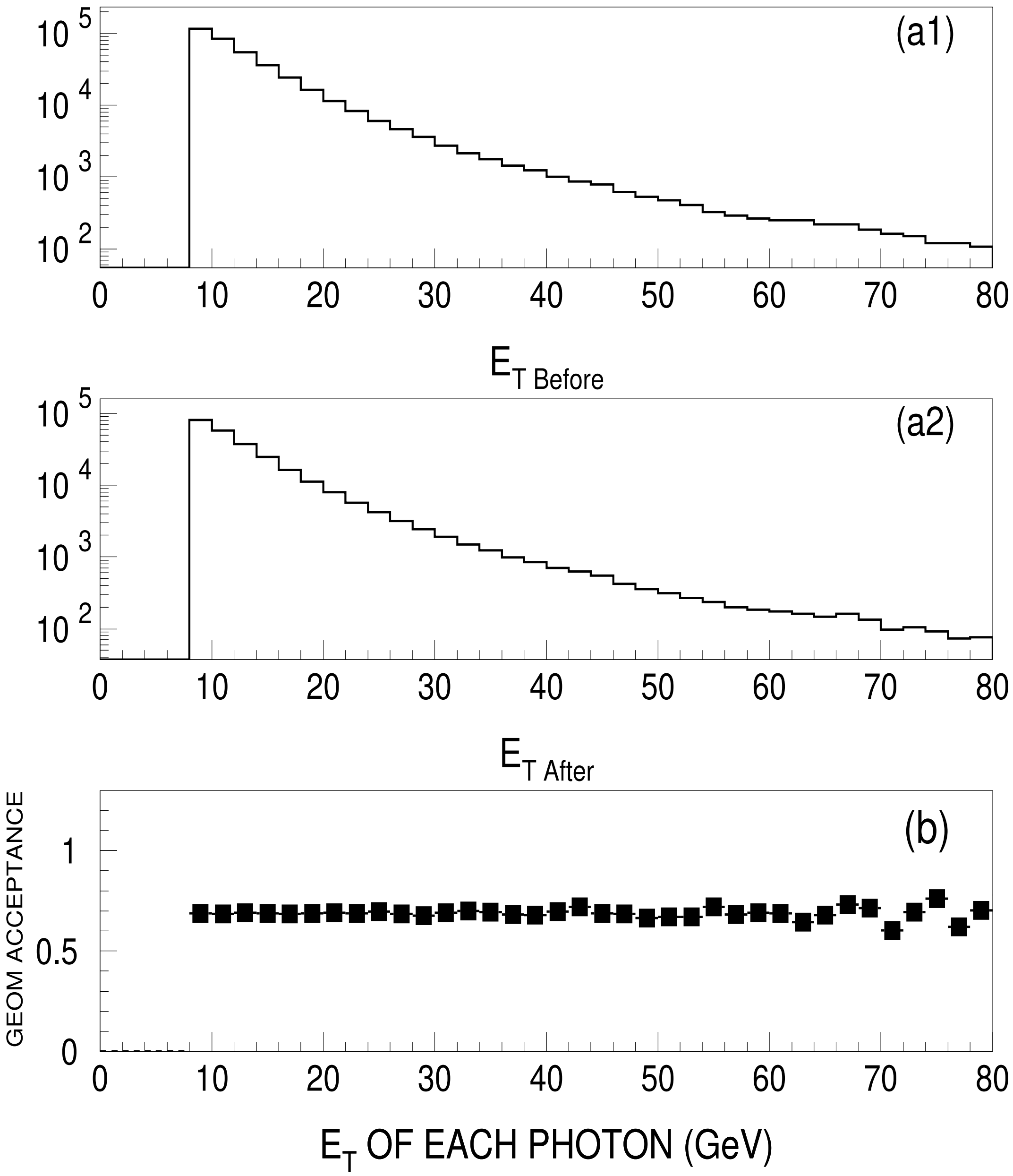

Figure 4.2: Illustration of the determination of the Geometric Acceptance. The distributions of E_T of each photon before (a1) and after (a2) cuts. The ratio of plot (a2) over plot (a1) bin-by-bin is what is shown in plot (b), which is the differential Geometric Acceptance vs.E_T.

	Geometric Acceptance
E_T^{γ}	0.687 ± 0.001
$P_T^{\gamma\gamma}$	0.688 ± 0.002
$M^{\gamma\gamma}$	0.672 ± 0.003
$\Delta\phi^{\gamma\gamma}$	0.688 ± 0.003

Table 4.3: Geometric Acceptance vs.various kinematic quantities

Combining the two factors, the geometric acceptance of double-photon events is:

$$\mathcal{A} \;\approx\; 80\% \times (1 - 15\%) \;\approx\; 68\%$$

For a precise determination, 2 million Monte Carlo events are used, which are generated as described in Section 4.2.1 and fast smeared using the resolution parameters given in Section 4.2.2. The acceptance value in each bin of a variable of interest is calculated as:

$$\mathcal{A}/bin \;=\; \frac{\text{Events/bin when } |\eta^{\gamma 1,\gamma 2}| < 1.0, |\eta_{\text{det}}^{\gamma 1,\gamma 2}| < 1.0, \phi_{\text{crack}}^{\gamma 1,\gamma 2} \text{ cut}}{\text{Events/bin when } |\eta^{\gamma 1,\gamma 2}| < 1.0} \tag{4.7}$$

Fig. 4.3 shows the dependence of the Geometric Acceptance ($\mathcal{A}$) upon E_T^{γ}, $P_T^{\gamma\gamma}$, $M^{\gamma\gamma}$ and $\Delta\phi^{\gamma\gamma}$. They are found to be essentially independent of each variable.(see Table 4.3)

4.3 Efficiencies

Before getting into the final candidate sample, all the events are required to pass the selections including the trigger and the offline cuts. To measure the

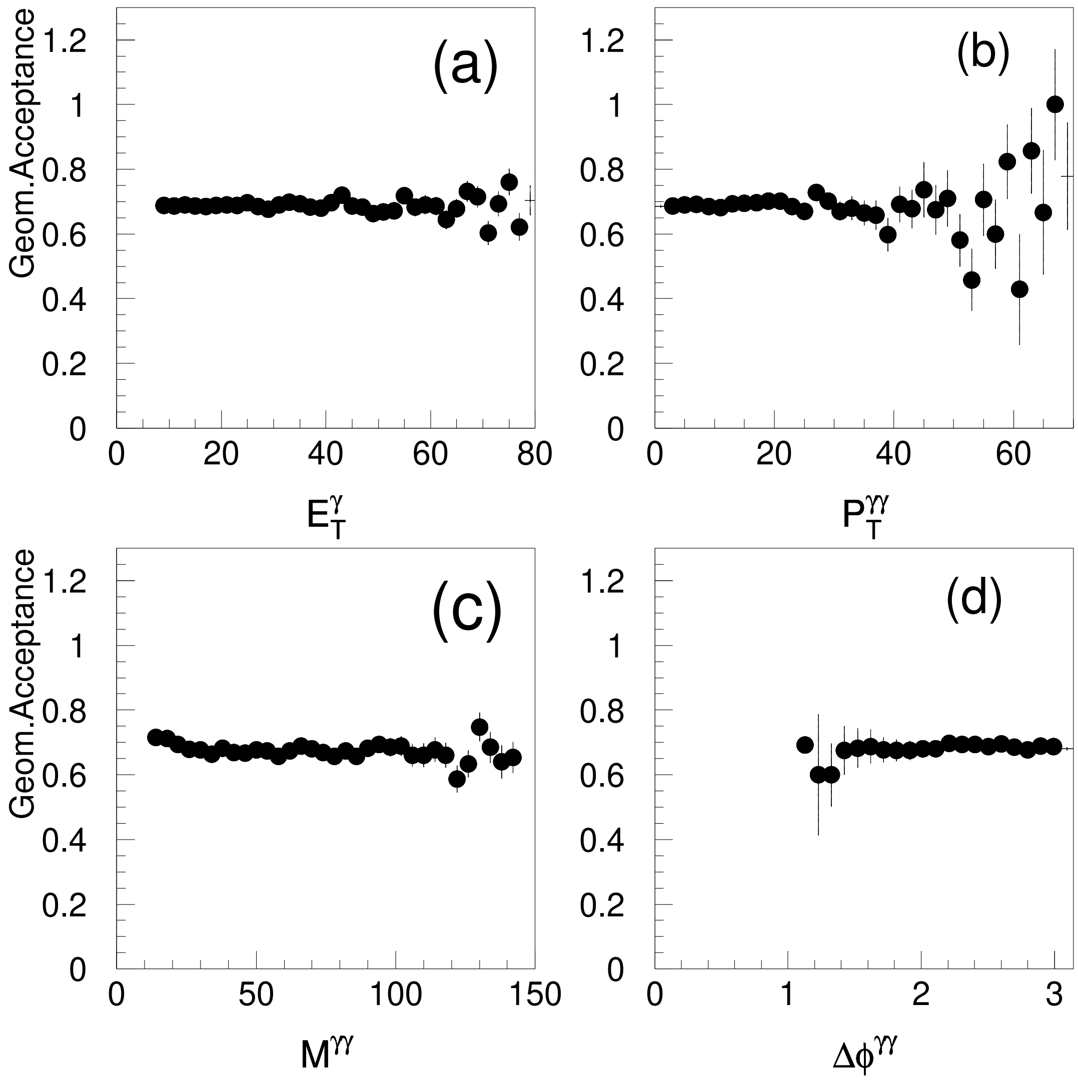

Figure 4.3: Differential Geometric Acceptance vs. (a)E_T^{γ} of each photon, (b)$P_T^{\gamma\gamma} = |\vec{E}_T^{\gamma 1} + \vec{E}_T^{\gamma 2}|$ of the double-photon system, (c)$M^{\gamma\gamma}$ of the photon pair, and (d)$\Delta\phi = |\phi^{\gamma 1} - \phi^{\gamma 2}|$ of the photon pair. They are nearly constant in all cases.

production cross section, one must understand the efficiencies of each selection step.

4.3.1 Trigger Efficiencies

Level Ø Efficiency

The Level Ø requirement for double-photon events is that the level Ø counters must detect $p\overline{p}$ collisions and yield a fast-z information (see Table 4.1). To study its efficiency, a sample of W $\rightarrow e\nu$ was chosen by requiring one Level-2 filter *EM1_EISTRKCC_MS* having fired; this does not require Level-Ø confirmation(for trigger version 10.0 and later) [56]. Measuring the fraction of events which have fast-z information, the Level-Ø efficiency for requiring fast-z is (see [56])

$$\epsilon_{L\emptyset} \;=\; 0.98 \pm 0.01 \tag{4.8}$$

Double-photon Trigger Efficiency

The double-photon candidate sample requires that both the Level-1 trigger *em_2_med* and the Level-2 filter *em2_gis_gam* are satisfied (see Table 4.1 for descriptions). To measure the overall trigger efficiency, we use one "unbiased" electron from $Z \rightarrow ee$ decays as a diagnostic sample and send these electrons through the trigger simulation.

The diagnostic sample is collected by selecting an inclusive $Z \rightarrow e^+e^-$ sample as follows:

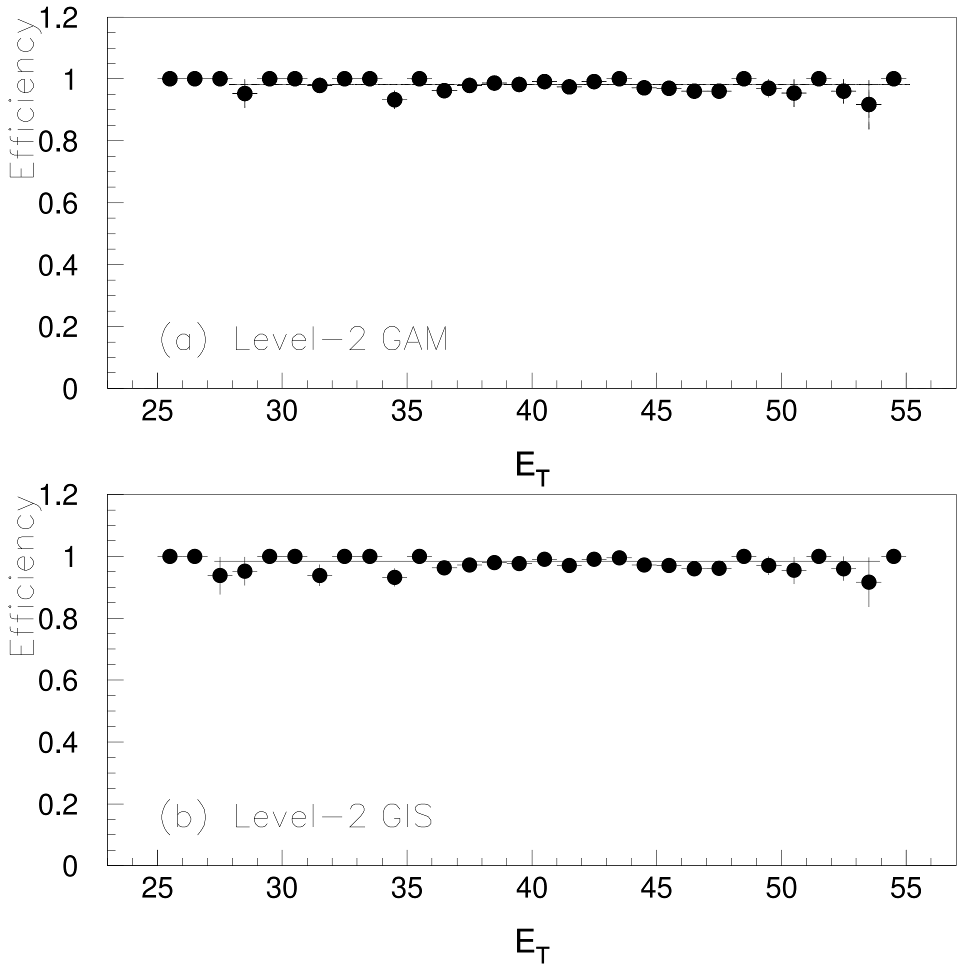

Figure 4.4: Trigger Efficiency Plots (a) and (b) for Level-2 *GAM* and *GIS* configurations. The E_T range shown is far above the threshold of 12 GeV.

- Requiring at least two high E_T electrons or one electron and one photon both having $E_T > 20$ GeV.
- Requiring both EM objects to be in the central region, $|\eta^{1,2}| < 1.0$, $|\eta^{1,2}_{det}| < 1.0$, and being away from ϕ cracks.
- Requiring both EM objects to satisfy electron or photon ID cuts:

$$\chi_1^2 < 100, \qquad \chi_2^2 < 100.$$
$$\text{isol\_frac}_1 < 15\%, \ \ \text{isol\_frac}_2 < 20\%$$

  Their pair mass must be in the resonance window $75 < M_{ee} < 95$ GeV.
- Requiring the Level-2 filter *em2_eis_esc* to have been fired.

A trigger simulation is setup with some special Level-2 filters: *EIS, ELE, ESC, GIS, GAM, ISO*, where *EIS* means electron_shape and isolation, *ELE* means electron_shape only, *ESC* means no shape and isolation cuts, *GIS* means photon_shape and isolation, *GAM* means photon_shape only, and *ISO* means isolation only.

The diagnostic sample has passed the level-2 requirement, *em2_eis_esc*. If both EM objects have passed the Level-2 *EIS* requirement, either of them may have passed the Level-2 *ESC* requirement. They are considered unbiased and both of them get into the "unbiased" electron sample. If one object satisfied the Level-2 *EIS* requirements and the other failed, the failed one must have passed through the *ESC* and therefore gets into the "unbiased" electron sample. The "unbiased" electron is hence put through the trigger simulation, and the resulting trigger efficiencies for the Level-2 terms *GIS* and *GAM* are measured as:

$$\epsilon_{\text{trig}} = \frac{\text{Num of Electrons passed the trigger in the “unbiased'' sample}}{\text{Num of all Electrons in the “unbiased'' sample}} \tag{4.9}$$

Fig. 4.4 shows the trigger efficiencies for *GIS* and *GAM* respectively. The efficiencies are constant over the E_T range $26 < E_T < 55$; the fitted values are

$$\epsilon_{\text{trig}}^{GIS} = 0.983 \pm 0.003, \qquad \epsilon_{\text{trig}}^{GAM} = 0.987 \pm 0.003$$

Because of the kinematics of $Z \to e^+e^-$ decays, the E_T of the electrons populate the range between 25 and 55 GeV. For the double-photon data sample, photons have E_T down to 13 GeV. To understand the low E_T ($13 < E_T < 30$) trigger behavior, data from special GAMMA runs with a lower threshold of $E_T^{thres} = 6\ GeV$ were used for efficiency studies. The special run trigger requires the *GAM* in Level-2 term, and the threshold (6 GeV) effect in the data is negligible for $E_T > 12$ GeV (the effects on the efficiencies are $< 0.5\%$). The set of data is run through the same trigger simulation as above; the measured efficiencies $\epsilon_{\text{spec}}^{GIS}, \epsilon_{\text{spec}}^{GAM}$ are corrected by the efficiency $\epsilon_{\text{GAM}} = 0.987$ of the *GAM* requirement in the data set. The trigger efficiency is therefore determined in each E_T bin as:

$$\epsilon_{\text{trig}}^{GIS} = \epsilon_{\text{spec}}^{GIS} \times 0.987$$
$$\epsilon_{\text{trig}}^{GAM} = \epsilon_{\text{spec}}^{GAM} \times 0.987$$

Fig. 4.5 shows the combined trigger turn-on curves for *GAM* and *GIS*, with the threshold $E_T^{thr} = 12$ GeV for a range of $12 < E_T < 60$. The curves can be

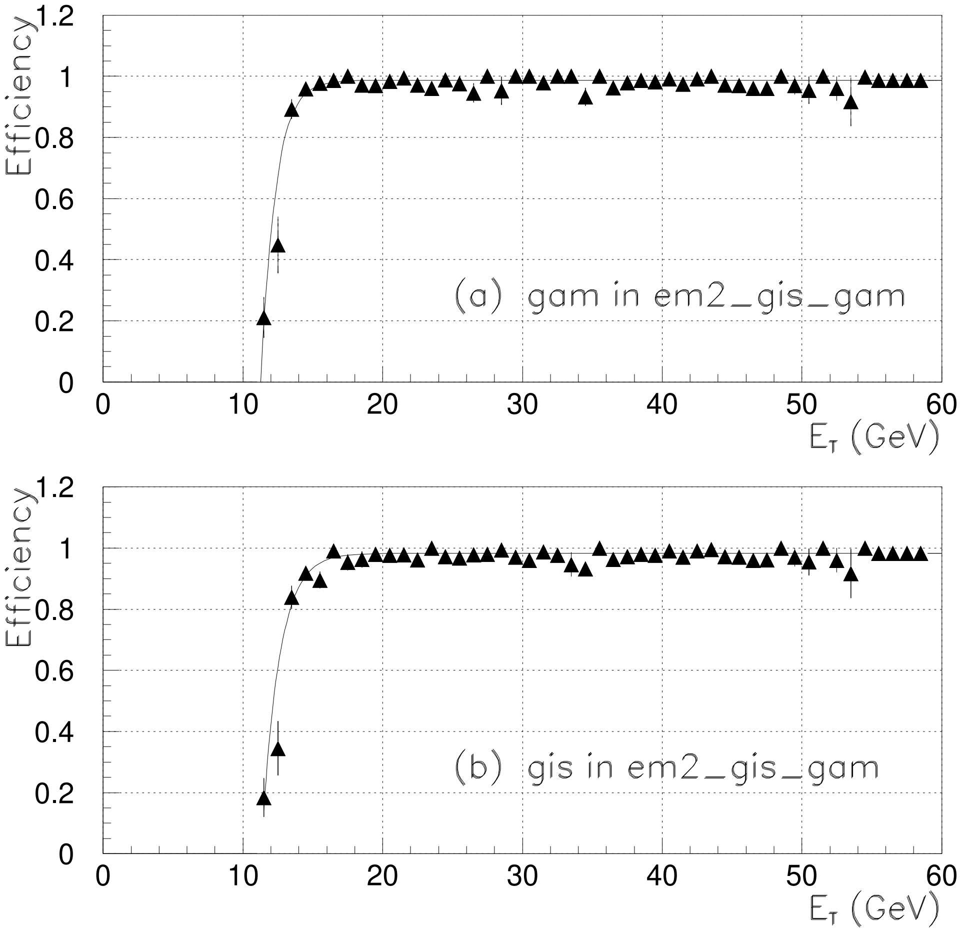

Figure 4.5: The trigger efficiency curves for (a) *GAM* and (b) *GIS* with a threshold of $E_T^{thr} = 12$ GeV.

fitted as the following functions:

$$\epsilon_{\rm trig}^{GIS}(E_T) = 0.9831 \times (1.0 - e^{9.178-0.812E_T}) \tag{4.10}$$

$$\epsilon_{\rm trig}^{GAM}(E_T) = 0.9874 \times (1.0 - e^{10.890-0.965E_T}) \tag{4.11}$$

4.3.2 Emulated Photons

After events pass the trigger requirements, they are run through the *RECO* program and preliminary photon candidates are filled into the ZEBRA [78] bank PPHO. These PPHOs are further selected using fudicial cuts and photon ID cuts (Table 4.2 gives a complete list of the cuts). The offline selection efficiency ϵ_{offl} includes all the steps starting from *RECO*.

Before embarking on efficiencies, it is useful to introduce the "emulated photon" sample. The idea is that if we choose an average area of the calorimeter, away from jets and real objects, build an emulated photon and construct a road associated with it. We can learn a lot about the tracking chamber response from underlying particles and noise by looking at the hits and tracks within the road of the emulated photon. The procedures are as follows:

- Selecting a tight $Z \to e^+e^-$ sample similar to the diagnostic sample in Section 4.3:

  1. Both PELCs, $E_T^1 > 20, \quad E_T^2 > 20$ GeV.
  2. $|\eta^1| < 1.0$ or $|\eta^1| > 1.5, \quad |\eta^2| < 1.0$ or $|\eta^2| > 1.5$.
  3. $\chi_1^2 < 100, \quad \chi_2^2 < 100$.
  4. $\text{Isol\_frac}_1 < 15\%, \quad \text{Isol\_frac}_2 < 15\%$.
  5. $70 < M_{ee} < 100$ GeV for pair mass.

- Determining the bisector of two electrons as $\eta = \frac{1}{2}(\eta_1 + \eta_2)$, $\phi = \frac{1}{2}(\phi_1 + \phi_2)$. Using the bisector as the shower centroid of an emulated photon, we build a core cone (R=0.2) and a isolation cone (R=0.4) and fill their E and E_T by summing over all towers. A tracking road is also constructed between the centroid and the vertex, and hits and tracks are found within the road.

The constructed emulated photon sample will be used for several studies later.

4.3.3 Offline Efficiencies

Section 3.4.1 describes how *RECO* programs find preliminary photon candidates (PPHOs). Since double photons have a high E_T ($>$ 12 GeV) and are usually isolated, they can always be found by the DØ nearest-neighbor algorithm and satisfy the E_T cut of 1.5 GeV. The EM fraction cut($>$ 90%) is not so easy to verify in efficiency. At the test beam load2, single electron beams were used to test the CC calorimeter. Using the same algorithm as in DØ, the EM fraction cut (defined as $\mathrm{EM_{frac}} < 90\%$) is 100% efficient up to 75 GeV and is $\approx$ 1% inefficient for 100 GeV electrons. Before applying the results at DØ, one has to consider the differences between the test beam and DØ. The sampling fraction at the test beam is different from that used for the run 1B data, the test beam has less upstream material than DØ does, and the test beam electrons only come from vertex z=0 while the event vertex at DØ has a wide gaussian distribution with RMS = 25 cm. Taking into account these differences, the EM fraction cut at DØ remains $>$ 99% efficient for photons up to E_T = 75 GeV, based on Monte Carlo simulation.

A further cut, which requires the fraction of the energy of the highest-energy tower over the whole cluster energy must be $>$ 40%, is studied using

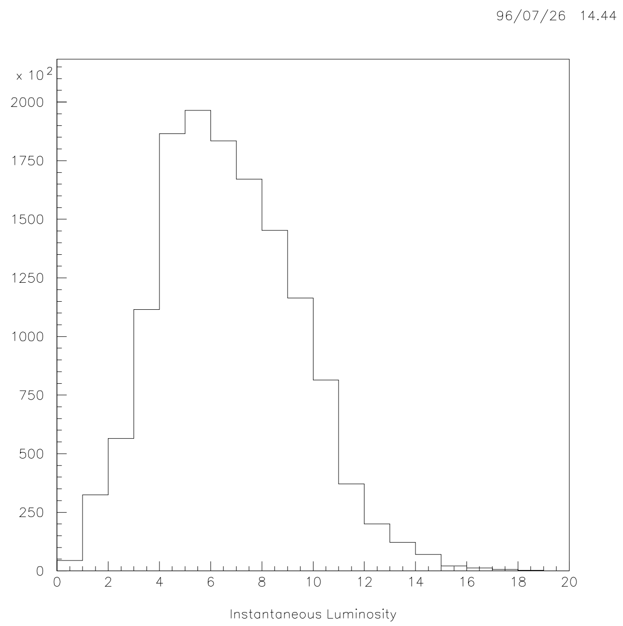

Figure 4.6: The instantaneous luminosity L spectrum for run 1B, in the unit of $10^{30} cm^{-2} s^{-1}$.

both the test beam and the Plate level Monte Carlo electrons (their vertex being smeared as a gaussian of RMS = 25 cm). No events fail the cut.

No Tracks in the Road

The preliminary EM objects are considered as PPHOs or PELCs, depending on whether any tracks can be found in the road between the EM objects and the vertex. What is the efficiency for true photons to satisfy the no-track requirement? Two effects will make true photons fail, listed below:

- True photons may convert when they pass through the material before the central drift chamber (CDC). The amount of upstream material is $\approx 0.1 X_0$ (X_0 being *radiation length*) in the central region, including the beryllium beam pipe, the VTX chamber and the TRD. The conversion probability (p_{conv}) is expected to be ≈ 0.1 for a photon to convert into e^+e^- pair before it reaches the CDC. A Monte Carlo single photon sample was generated and run through the DØ detector simulation. It confirmed that the fraction of converted photons $\approx 10\%$, i.e. the efficiency for photons not to convert is

$$\epsilon_{\text{dont\_convert}} \approx 0.90 \tag{4.12}$$

- Even if a true photon has not converted, some charged tracks from underlying particles may happen to fall in the tracking road of the photon. This loss is measured using the emulated photon sample, with an additional requirement that the E_T in the isolation cone (R=0.4 - R=0.2) $<$ 2 GeV. This isolation cut makes sure that the emulated photons have the same environment as the true photons. The efficiency for the emulated photons not to have tracks in the road is

$$\epsilon_{\text{no\_underlying\_tracks}} = 0.917 \pm 0.001 \tag{4.13}$$

The two efficiencies are uncorrelated, and the overall efficiency for true photons not to have tracks in their road is

$$\begin{aligned} \epsilon_{\text{no\_tracks}} &= \epsilon_{\text{dont\_convert}} \times \epsilon_{\text{no\_underlying\_tracks}} \qquad (4.14) \\ &\approx 0.82 \pm 0.01 \end{aligned}$$

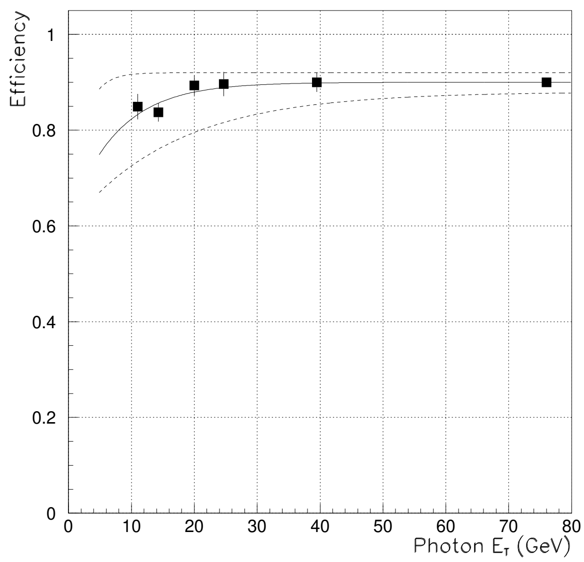

Figure 4.7: Efficiency Curve of the cuts $\chi^2 < 100$, $E_T^{iso-4} < 2.0$ GeV for photons. The dashed lines show the range of uncertainty.

Calorimeter χ^2 and Isolation Cuts

Further tight photon ID cuts are applied to the data sample(see Section 3.4.2). To understand the efficiencies of χ^2 and E_T^{isol} cuts, we generated a single photon Monte Carlo sample and put it through Plate level Geant simulation [96]. The photons are then overlaid with four sets of minibias events which were collected at four different instantaneous luminosities, covering the range in the fig. 4.6). Here the minibias events indicate the events that are collected when the level-Ø detects hard collisions. The minibias events include the contributions from both underlying particles and electronics and uranium noise. Due to strong correlation, the shape of the combined efficiency of the two cuts,

$$\chi^2 < 100, \quad E_T^{iso-4} < 2.0 \text{ GeV} ,$$

is measured using the single photon MC sample. The normalization is determined by using a clean $Z \to e^+e^-$ sample. Four different curves were obtained corresponding to the four instantaneous luminosities, and they are weighted according to the Figure 4.6. The resulting efficiency curve as a function of E_T^γ is shown in Fig. 4.7.

An additional isolation cut $E_T^{iso-7} < 7.0$ GeV is applied to the data to reduce the jet background. This is a cut more on the environment around the photon rather than the photon itself. The efficiency is therefore measured by studying the E_T flow of underlying particles. The emulated photon sample (see Section 4.3.2) describes a region away from any high E_T objects, and so the E_T distributions in the core cone R=0.2, the isolation cones R=(0.4 - 0.2) and R=(0.7 - 0.2) provide a spectrum for the E_T flow from underlying particles (see Fig. 4.8).

The EM showers from electrons and photons are very narrow transversely, with nearly all the energy being contained in the core cone R=0.2. Fig. 4.9 shows the E_T leakage for 15, 20, 25 GeV testbeam electrons. Even with underlying particles, the isolation cut $E_T^{iso-4} < 2$ GeV is $> 90\%$ efficient. Using the distributions in Fig. 4.8, we measured the ratio of events satisfying both $E_T^{iso-4} < 2.0$ and $E_T^{iso-7} < 7.0$ GeV over those satisfying $E_T^{iso-4} < 2.0$ GeV. The resulting efficiency value is:

$$\epsilon_{\mathrm{isol7}} = 0.967 \pm 0.010 \qquad (4.15)$$

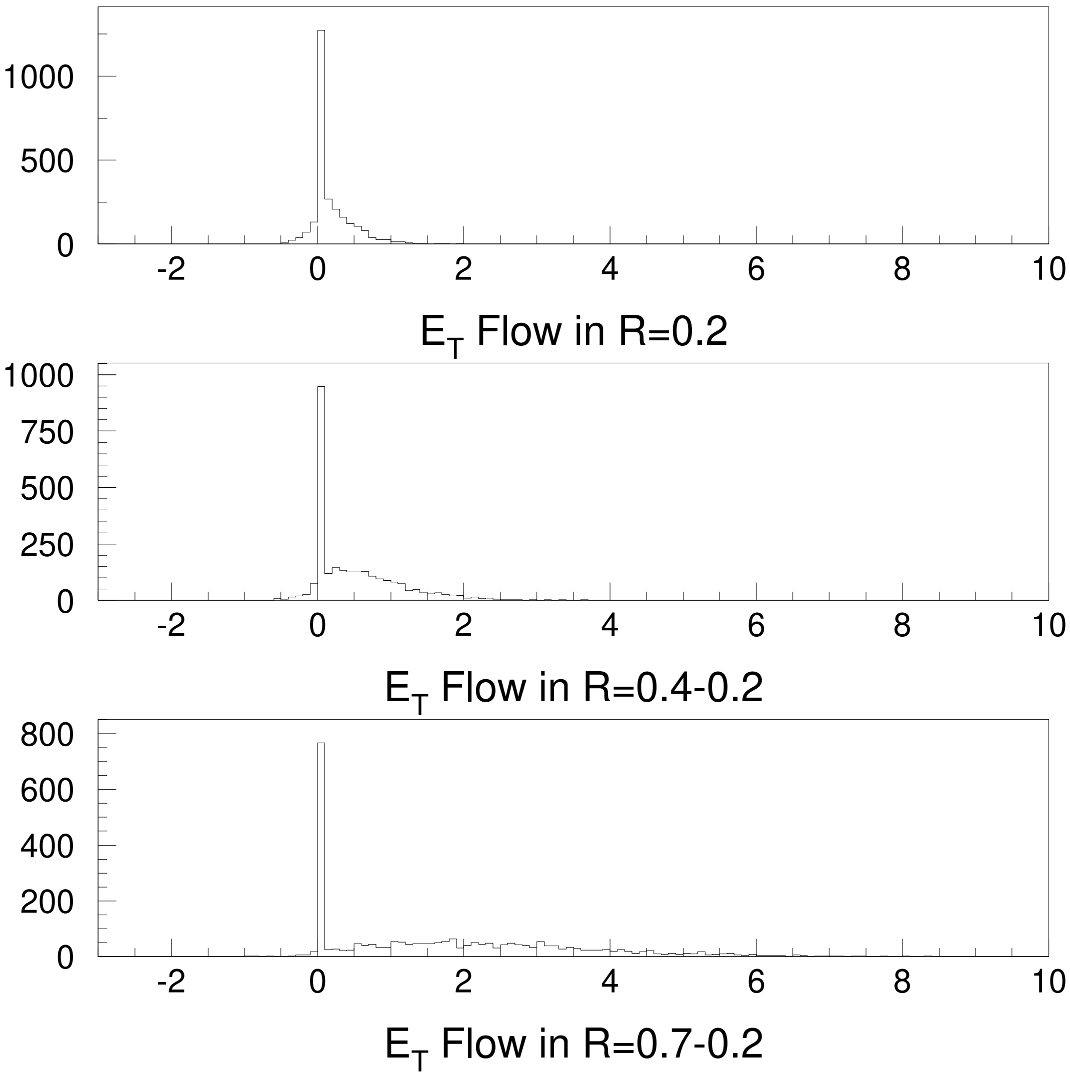

Figure 4.8: E_T flow spectrum of the underlying particles in $Z \to e^+e^-$ events, with the core cone R=0.2, the isolation cones (R=0.4 - R=0.2) and (R=0.7 - R=0.2)

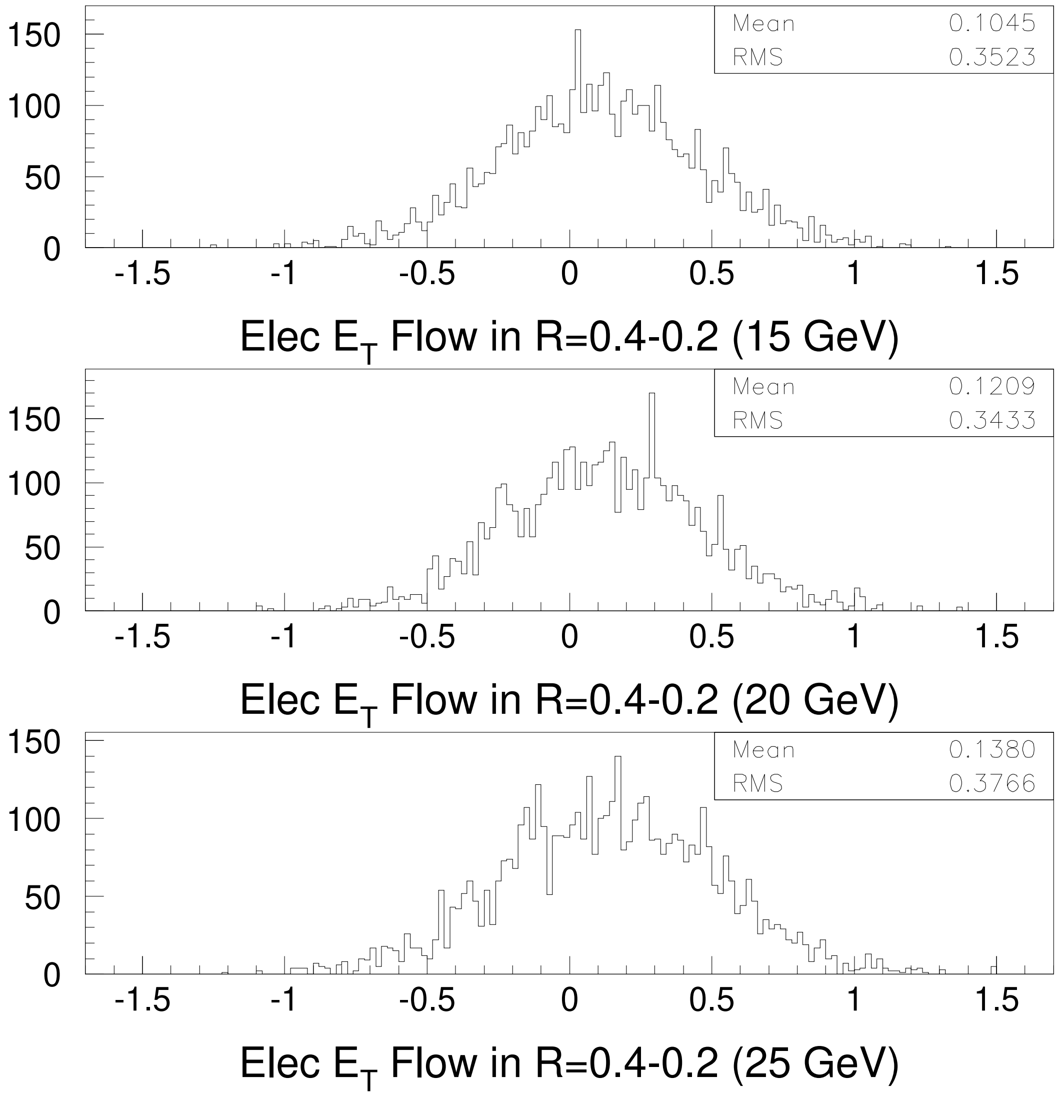

Figure 4.9: E_T flow spectrum in the isolation cone (R=0.4 - R=0.2) for test-beam electrons with energies E = 15, 20, 25 GeV. Very small fractions of energy leak out of the core cone, and this leakage is nearly energy independent.

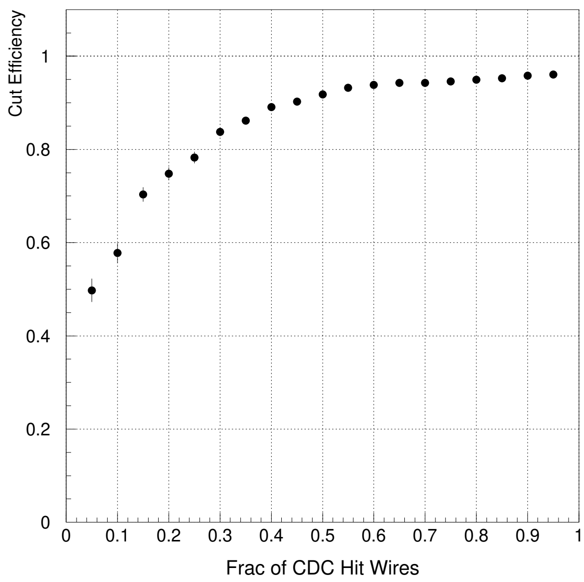

Figure 4.10: Efficiency versus the cut of the fraction of CDC hit wires when the number of 3D hits is required to be 0.

Hits_in_road Cuts

The efficiency of the Hits_in_Road cuts (see Section 3.4.2 for definition) is determined using the emulated photon sample (Section 4.3.2). Requiring NHCDC3D < 1 and varying the RHCDCW cut for the no-track emulated photons, the efficiency curve is shown in Fig. 4.10. The chosen cut is **RHCDCW** $<$ **0.5**, and its efficiency is

$$\epsilon_{\mathrm{hits\_in\_road}} = 0.918 \pm 0.007 \tag{4.16}$$

Tracking Efficiency in CDC

Before applying the hits_in_road cuts, some electrons from Z decay enter the double-photon sample, as shown in Fig. 3.4. To estimate the contamination from Z $\rightarrow$ ee, we will measure the tracking efficiency in the CDC and the rejection factor of the hits_in_road cuts for electrons when tracks are not found in the road.

A sample of PELC+PELC and PELC+PPHO from Z's is selected for the tracking efficiency study. For each PELC or PPHO, the requirements are

- the level_1 em_2_med and level_2 ele_2_high are fired
- $|\eta| < 1.1$, ϕ_crack_cut, $E_T > 20.$ GeV
- $\chi^2 < 100$, isol_frac $< 15\%$

The sample is further divided into one with only one vertex and the other with any number of vertices. The Z invariant mass spectrum is modelled using a relativistic Breit-Wigner shape with an exponential factor which accounts for the uncertainty of the Z production cross section [92], being further convoluted with a gaussian response function derived from the detector resolution on an event by event basis. A likelihood function $\mathcal{L}$ is constructed for each measurement of the mass m$_i$ and its resolution σ_i:

$$\mathcal{L}(m_i, \sigma_i; M_Z, \Gamma_Z) = \int dm\left(\frac{e^{-\beta m} m^2}{(m^2 - M_Z^2)^2 + \frac{m^4 \Gamma^2}{M_Z^2}} + Ae^{-Bm}\right)\frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(m-m_i)^2}{2\sigma_i^2}} \tag{4.17}$$

where M_Z and Γ_Z are the input mass and width of Z from LEP results, A and B are parameters to fit the background, and m is the best fitted mass of Z.

By maximizing $\ln(\mathcal{L})$, A and B can be determined and so the amount of background is measured for both PELC+PELC and PELC+PPHO. Let us call T to be the CDC tracking efficiency, and N$_{ee}$ and N$_{e\gamma}$ to be the number

	one vertex	any number vertices
tracking efficiency in CDC	$(91.4 \pm 2.1)\%$	$(89.8 \pm 1.9)\%$

Table 4.4: Single electron tracking efficiency in CDC

of true ee pairs and true $e\gamma$ pairs within the mass window $80 < M < 95$ GeV. After subtracting the backgrounds, the relation holds as:

$$\frac{N_{ee}}{N_{e\gamma}} = \frac{T \times T}{2 \times T \times (1 - T)} \tag{4.18}$$

The tracking efficiencies in the CDC are measured for the two samples respectively, one with only one event vertex and the other having any number of vertices. The results are summarized in Table 4.4.

Z electrons in the double-photon sample

When an electron fails to have tracks reconstructed in its road, it can be effectively rejected by applying hits_in_road cuts. Using the above sample with both PELC+PELC and PELC+PPHO from Z, the rejection factor is measured as $\mathbf{F_{reject}} = \mathbf{6.6 \pm 5.0}$, namely as the total number of PPHOs over the number of the PPHOs passing the hits_in_road cuts.

Taking the measured $Z \to e^+e^-$ cross section $\sigma_Z B(Z \to e^+e^-) = 0.218$ nb [98], the estimated number of Z events in the double-photon sample is:

$$\begin{aligned} N_{\mathrm{Z-backgr}} &= L \times \sigma_Z B(Z \to e^+e^-) \times A_Z \times \epsilon_Z \times (1-T)^2 \\ &\approx 82.36 \times 218. \times 0.36 \times 0.9 \times (1-0.9)^2 \end{aligned}$$

$$N_{\mathrm{Z-backgr}} \approx 58 \tag{4.19}$$

Looking at the mass plot of the photon pair in Fig. 3.4 and assuming a linear background around the Z resonance on a log scale, the background is estimated as $\approx 51 \pm 8$. It is consistent with the estimation in eqn. 4.19. After the hits_in_road cuts, the number of Z remaining in the photon sample is approximately 8 ± 6.

Overall Efficiency

Taking into account of all the cuts which select our photon candidates, the overall efficiency is equal to:

$$\begin{aligned} \epsilon_{\text{offline}} &= \epsilon_{\text{no\_tracks}} \\ &\times \epsilon_{\text{hits\_in\_road}} \\ &\times \epsilon_{\chi^2\text{\_and\_isol4}} \\ &\times \epsilon_{\text{isol7}} \end{aligned} \tag{4.20}$$

4.4 Background Subtraction

The background to direct photons consists of some jets fragmenting into one leading energy π^0, or η, or ω and some soft particles around it. Using the calorimeter shower profile, the two photons from high energy π^0 (or η, or ω) travel nearly collinearly (with a very small opening angle) and so they are not distinguishable from directly produced photons on an event-by-event basis. But looking at the longitudinal energy deposition, especially the energy deposited in the first EM layer(EM1), the nearly-collinear two photons are likely to deposit more energy than a single photon. We can model the distribution of -ln(EM1/ETOT) for true photons and fakes(fakes refer to the EM jets which are indistinguishable from the photons). Their distributions are shown in Fig. 4.11. The data are composed of both true and fake photons, and each fraction can be determined by fitting the distribution of -ln(EM1/ETOT) in the data as a sum of that of true and fake photons.

4.4.1 Discriminant Distributions of True and Fake Photon Samples

To model the discriminant -ln(EM1/ETOT) of photons, a large Monte Carlo sample of single photons is generated and run through plate level detector simulation. It is further smeared by electronic and uranium noise. The plots (a1) and (b1) in Fig. 4.11 show the EM1 fraction and -ln(EM1/ETOT) distributions for photons.

For the fake photon sample, a large jet Monte Carlo sample is generated by using PYTHIA. It is filtered through several steps as below [99]:

- qq, qg, or gg are allowed to collide, and the emerging partons after the hard collision are allowed to fragment and hadronize.
- A seed list ordered in E_T is made by looping through all π^0, η, and γ with $E_T > 5$ GeV.
- The number of total charged tracks, N_{trk}, is counted in the core cone R=0.2 around each seed.

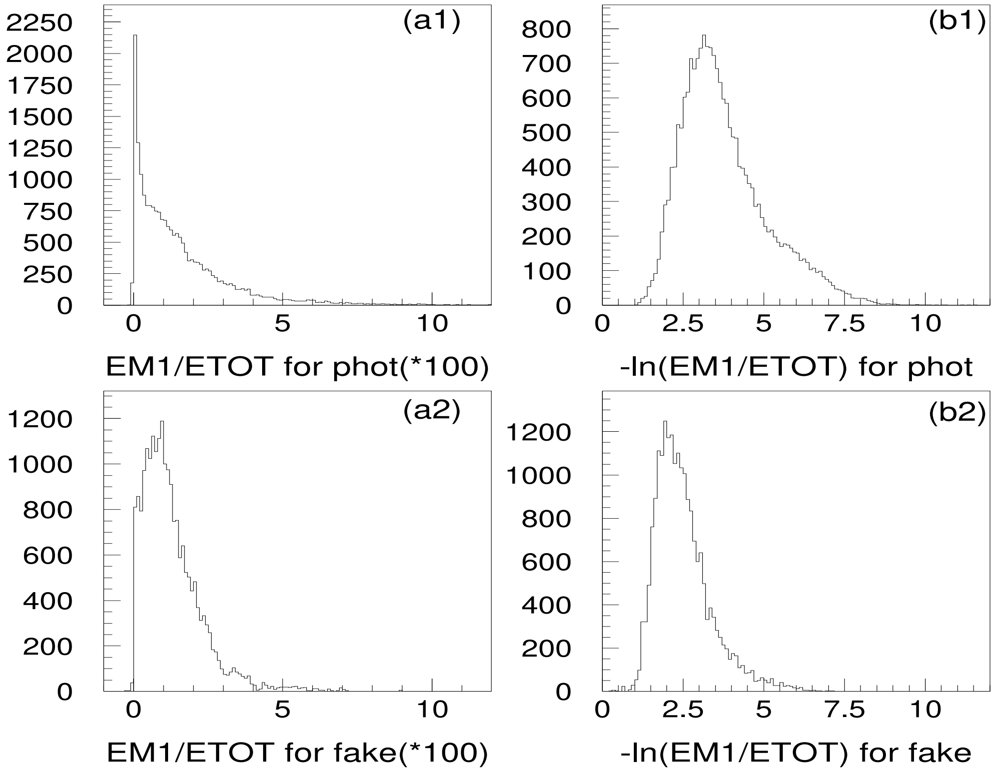

Figure 4.11: Plots (a1) and (a2) show the distributions of the fraction of EM1 over total energy for Monte Carlo photon and fake, and plots (b1) and (b2) describe the distributions of -ln(EM1/ETOT) correspondingly.

- An isolation cone between R=0.4 and R=0.2 is built around each seed particle, within which the E_T sum and E_T^{iso} of all particles are calculated.

- One seed becomes a candidate only when:

$$N_{trk} = 0, \quad E_T^{iso} < 2 \text{ GeV}, \quad \frac{E_T(\text{R=0.2})}{E_T(\text{parent parton})} > 0.6$$

- All candidate seeds are run through the detector simulation and then are superimposed with zero-bias events (including electronic and uranian noise)

- The final fake photon sample is further selected using the same photon ID cuts as the data.

The distributions of the EM1 fraction and -ln(EM1/ETOT) of the fake photons are shown in the plots (a2) and (b2) in 4.11.

The modeling of the (EM1/ETOT) distribution for Z electrons was done as well, and was compared with Z electron data. As shown in Fig. 4.12, the Monte Carlo reproduces the data reasonably well.

4.4.2 Constructing Discriminant Distributions for Double Photons

The double-photon events are assumed to be made up of three possible sources: photon-photon, photon-fake, and fake-fake. We construct the product,

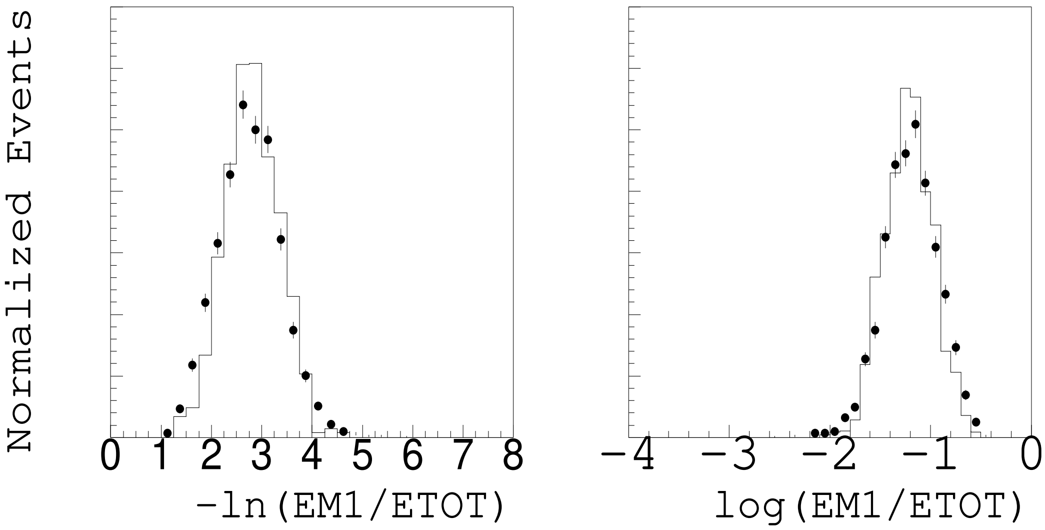

Figure 4.12: -ln(EM1/ETOT)(left) and Log(EM1/ETOT)(right) distributions for Z electrons from data (points) and Monte Carlo (histograms), CC only.

$$(-\ln(EM1/ETOT))_{\gamma 1} \times (-\ln(EM1/ETOT))_{\gamma 2}$$

as the event discriminant. The distributions of the discriminants are shown in Fig. 4.13. Random numbers are generated according to the histograms of photons and fakes. The random numbers are multiplied to make products of three combinations, phot*phot, phot*fake and fake*fake, whose distributions are shown in Fig. 4.13.

4.4.3 Fitting Double-photon purity $\alpha^{\gamma\gamma}$

The distribution of the double-photon event discriminant can be filled into a histogram of n bins, resulting in a set of numbers d_1, $d_2 ... d_n$. Let us fill the histograms of the discriminant of the modeled phot*phot, phot*fake and fake*fake in the same bins, and let $f_i(\alpha^{\gamma\gamma}, \alpha^{\gamma f}, \alpha^{ff})$ be the expected number of

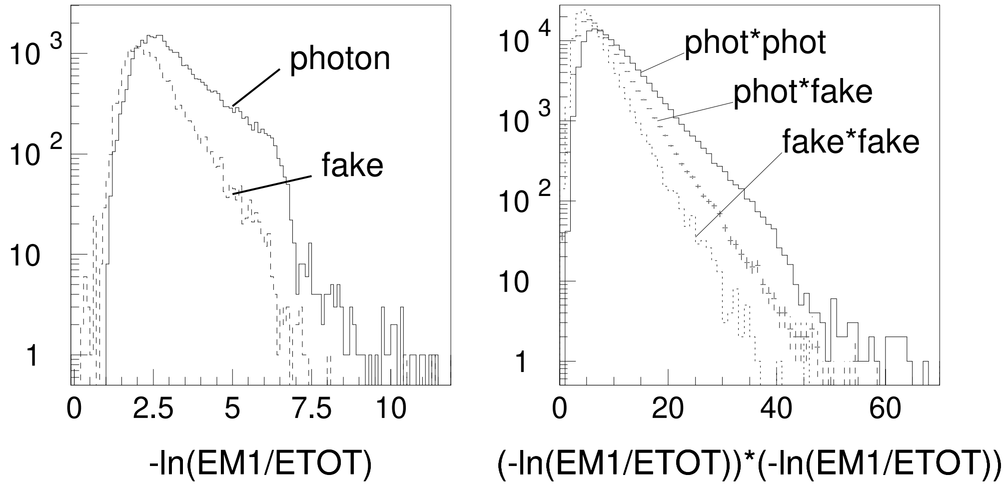

Figure 4.13: The distributions of -ln(EM1/ETOT) for modeled photons and fakes are shown left, and right are the product distributions for phot*phot, phot*fake and fake*fake.

events in the bin i, where $\alpha^{\gamma\gamma}$, $\alpha^{\gamma f}$ and α^{ff} are the fractions of the respective sources. $f_i(\alpha^{\gamma\gamma}, \alpha^{\gamma f}, \alpha^{ff})$ can be expressed as:

$$f_i(\alpha^{\gamma\gamma}, \alpha^{\gamma f}, \alpha^{ff}) = N_D(\alpha^{\gamma\gamma}\frac{a_i^{\gamma\gamma}}{N^{\gamma\gamma}} + \alpha^{\gamma f}\frac{a_i^{\gamma f}}{N^{\gamma f}} + \alpha^{ff}\frac{a_i^{ff}}{N^{ff}}) \tag{4.21}$$

where N_D, $N^{\gamma\gamma}$, $N^{\gamma f}$ and N^{ff} are the total number of events in the samples of data, phot*phot, phot*fake and fake*fake respectively, and $a_i^{\gamma\gamma}$, $a_i^{\gamma f}$ and a_i^{ff} are the number of events in the ith bin from the three modeled sources. The fractions of the three sources have to sum up to unity:

$$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1 \tag{4.22}$$

Let d_i to be the measured number of events in the ith bin from the data sample, the probability distribution for d_i is a Poisson(where f_i is the expected number):

$$e^{-f_i}\frac{f_i^{d_i}}{d_i!} \tag{4.23}$$

The three sources, phot*phot, phot*fake and fake*fake, are modeled using Monte Carlo, and only have finite number of events. Let A_i be the expected number of events for the measured a_i; a_i follows a Binomial distribution. It can be approximated as a Poisson when A_i is much smaller than the total number of events in the sample:

$$e^{-A_i}\frac{A_i^{a_i}}{a_i!} \tag{4.24}$$

The overall log likelihood, including the Poisson distributions for data and the modeled three sources, can therefore be expressed as:

$$\ln\mathcal{L} = \sum_{i=1}^{n} d_i \ln f_i - f_i + \sum_{i=1}^{n}\sum_{j}^{\gamma\gamma,\gamma f,ff} a_i^j \ln A_i^j - A_i^j \tag{4.25}$$

By maximizing $\ln\mathcal{L}$ with respect to $\alpha^{\gamma\gamma}$, $\alpha^{\gamma f}$, α^{ff} and $A_i^{\gamma\gamma}$, $A_i^{\gamma f}$, A_i^{ff}, we can determine the fractions of the three sources. A detailed description of the approach, *Fitting using Finite Monte Carlo Samples*, can be found elsewhere [100, 101].

Due to their physics meaning, the three fraction parameters $\alpha^{\gamma\gamma}$, $\alpha^{\gamma f}$, α^{ff} can only have the value between 0 and 1, and their sum has to be equal to 1. To avoid any boundary anomaly, a transformation on the α parameters is performed, giving the new parameters θs as follows:

$$\alpha^{\gamma\gamma} = \sin^2(\theta^{\gamma\gamma}) \tag{4.26}$$

$$\alpha^{\gamma f} = \sin^2(\theta^{\gamma f}) \tag{4.27}$$

$$\alpha^{ff} = \sin^2(\theta^{ff}) \tag{4.28}$$

The θ parameters can take any value between -π and π to guarantee the α parameters falling in their physical range.

Subsample	$E_T^1 < \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 > \mathbf{20.}$
Num Evts	113	139	81
$< E_T^1 >$	17.26 GeV	25.50 GeV	34.04 GeV
$< E_T^2 >$	15.85 GeV	16.93 GeV	28.71 GeV
$\theta^{\gamma\gamma}$	$0.76^{+0.26}_{-0.20}$	$0.90^{+0.16}_{-0.28}$	$0.79^{+0.11}_{-0.14}$
$\theta^{\gamma f}$	$0.81^{+0.20}_{-0.49}$	$0.33^{+0.35}_{-0.87}$	$0.00^{+0.49}_{-0.49}$
θ^{ff}	$0.00^{+0.48}_{-0.48}$	$0.56^{+0.15}_{-0.36}$	$0.78^{+0.11}_{-0.14}$
$\alpha^{\gamma\gamma}$	$0.47^{+0.25}_{-0.20}$	$0.61^{+0.14}_{-0.28}$	$0.51^{+0.11}_{-0.14}$
$\alpha^{\gamma f}$	$0.53^{+0.20}_{-0.43}$	$0.10^{+0.28}_{-0.10}$	$0.00^{+0.22}_{at\ limit}$
α^{ff}	$0.00^{+0.21}_{at\ limit}$	$0.28^{+0.14}_{-0.24}$	$0.49^{+0.11}_{-0.14}$

Table 4.5: Summary of Double-photon Purity Fitting Results for three regions.

Assumption	$E_T^1 < \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 > \mathbf{20.}$
$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1$ $\alpha^{\gamma f} = 0$	$0.72^{+0.18}_{-0.19}$	$0.66^{+0.15}_{-0.16}$	$0.57^{+0.17}_{-0.17}$
$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1$ $\alpha^{\gamma f} = \alpha^{ff}$	$0.63^{+0.19}_{-0.19}$	$0.55^{+0.17}_{-0.17}$	$0.46^{+0.22}_{-0.22}$
$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1$ $\alpha^{ff} = 0$	$0.46^{+0.27}_{-0.25}$	$0.37^{+0.23}_{-0.23}$	$0.31^{+0.33}_{-0.28}$

Table 4.6: Fitted results for three assumed cases.

The double-photon data sample is divided into three subsamples. The log-likelihood fitting is performed for each subsample. The resulting fraction of the true photon*photon events in the data sample is called $\alpha^{\gamma\gamma}$, which will be denoted as the $\gamma\gamma$ purity sometime. The results of these fits are summarized in Table 4.5, where the errors are the statistical errors of the fit.

We notice the strong decrease of one background fraction, $\alpha^{\gamma f}$, with increasing P_T, accompanied by a strong increase of the other background, α^{ff}. This indicates an instability of the fit with two independent parameters. To estimate this effect, we performed some tests by reducing this two-parameter fit into a one-parameter fit. Three cases are considered, listed as follows:

1. assume one background source, phot*fake, is negligible.

   $\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1 \quad \text{and} \quad \alpha^{\gamma f} = 0$

2. assume two background sources, phot*fake and fake*fake, are equal.

Subsample	$E_T^1 < \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 > \mathbf{20.}$
$\alpha^{\gamma\gamma}\vert_{\text{test}}$	0.53 ± 0.17	0.49 ± 0.15	0.46 ± 0.15

Table 4.7: Likelihood weighted $\alpha^{\gamma\gamma}|_{\text{test}}$ values for three subsamples.

$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1 \quad \text{and} \quad \alpha^{\gamma f} = \alpha^{ff}$

3. assume the other background source, fake*fake, is negligible.

$\alpha^{\gamma\gamma} + \alpha^{\gamma f} + \alpha^{ff} = 1 \quad \text{and} \quad \alpha^{ff} = 0$

The resulting fits are summarized in Table 4.6. The log likelihood as a function of any two of the three parameters $\alpha_{\gamma\gamma}$, $\alpha_{\gamma j}$ and α_{jj} above shows a strong correlation between the two parameters. Effectively varying the paramters within their full range but using the likelihood function $\mathcal{L}$ as a weight we calculate for each E_T range the average $\gamma\gamma$ purity and its variance, called $\alpha^{\gamma\gamma}|_{\text{test}}$ and $\sigma^2(\alpha^{\gamma\gamma}|_{\text{test}})$, as follows,

$$\alpha^{\gamma\gamma}|_{\text{test}} = \int \alpha^{\gamma\gamma}\mathcal{L}d\alpha^{\gamma\gamma}d\alpha^{\gamma f} \tag{4.29}$$

Its associated uncertainty, called $\sigma^2(\alpha^{\gamma\gamma}|_{\text{test}})$, is equal to

$$\sigma^2(\alpha^{\gamma\gamma}|_{\text{test}}) = \int (\alpha^{\gamma\gamma} - \alpha^{\gamma\gamma}_{\text{test}})^2\mathcal{L}d\alpha^{\gamma\gamma}d\alpha^{\gamma f} \tag{4.30}$$

The resulting values are listed in Table 4.7. Within the fitting statistical errors, we see no evidence for an E_T dependence of the $\gamma\gamma$ purity $\alpha_{\gamma\gamma}$. The fitted fraction $\alpha^{\gamma\gamma}$s in Table 4.5 are consistent with the test results in Table 4.7 within the fitting uncertainties.

Even though the MonteCarlo (EM1/ETOT) distribution has been compared with Z electrons, and has shown good agreement (see Fig. 4.12). The

Subsample	$E_T^1 < \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 < \mathbf{20.}$	$E_T^1 > \mathbf{20.}$ $E_T^2 > \mathbf{20.}$
$\frac{\delta(\alpha^{\gamma\gamma})}{\alpha^{\gamma\gamma}}$	42%	52%	22%

Table 4.8: The total errors for $\alpha^{\gamma\gamma}$.

(EM1/ETOT) distribution for electrons from Z decay occupies the range from 0.01 to 0.1, (see Fig. 4.12), whereas the (EM1/ETOT) distribution for photons has a sizable fraction below 0.01 (see Fig. 4.11). In view of the good agreement of the data and Monte Carlo distributions in Fig. 4.12, we assume that the Monte Carlo distribution extrapolates the EM1/ETOT distribution well into the region below 0.01. Later on, we will refer the (EM1/ETOT) distribution as the discriminant distribution.

To estimate the uncertainties resulting from the shapes of the discriminants, each bin of the discriminant distribution is fluctuated by one standard deviation randomly up or down. They are then multiplied to form the products of three combinations, phot*phot, phot*fake and fake*fake, and the resulting histograms are smoothed. The log likelihood fitting is performed once more on data with the three new discriminant distributions. The biggest change in the fitted result $\alpha_{\gamma\gamma}$ is ≈ 25 %. The final errors are calculated by adding this 25% and the errors above in quadrature, as listed in Table 4.8. The phot*phot fraction $\alpha^{\gamma\gamma}$, sometimes called purity, will be used in the cross section determination in Chapter 6.