CDF/ANAL/EXOTIC/CDFR/8078
Version 2.0
March 17, 2008

Study of Conversion Detection Efficiency for the High-p_T Like-Sign Dilepton Analysis to Search for the Wh Production

Toru Okusawa, Yoshihiro Seiya, Takayuki Wakisaka, and Kazuhiro Yamamoto

Osaka City University, Japan

Abstract

This note describes a study on the efficiency of photon-conversion identification based on the conventional two-track separation. The efficiency is utilized for estimating residual photon-conversion backgrounds in our fermiophobic higgs and high-mass Standard-Model higgs searches using the events of high-p_T isolated like-sign dileptons. The base efficiency is defined with respect to the sample of conversions with one leg passing our electron selection criteria except the isolation cut and conversion veto. In the measurement, the efficiency is broken down to two parts. One is the efficiency that is dependent on the p_T of partner tracks and is defined to be relative to the plateau at higher p_T regions. The other is the absolute efficiency on the plateau. We find that the p_T dependent efficiency reaches its plateau at $p_T \simeq 1$ GeV/c and the absolute plateau efficiency is $0.935 \pm 0.006(\text{stat}) \pm 0.010(\text{syst})$. Using these results, we estimate the conversion detection efficiency as a function of p_T of parent photon. A further breakdown of the plateau efficiency is tried, in which the track-finding efficiency is estimated to be $0.975 \pm 0.007(\text{stat})$, while the efficiency of the identification criteria is $0.957 \pm 0.007(\text{stat})$. We also estimate the probability of accidental track association to an isolated electron as a result of applying the conversion identification algorithm and find it to be $(1.5 \pm 1.0)\%$. Considerations of realistic application of the efficiency to estimate residual conversion backgrounds are given in the latter part of the note including an example for the like-sign dielectron events.

1 Introduction

Our main physics motivation is to search for the fermiophobic higgs boson and the high-mass Standard-Model (SM) higgs boson using like-sign (LS) dilepton events produced by the following process:

$$qq' \to W^{\pm}h \to W^{\pm}W^{*}W^{*} \to \ell^{\pm}\ell^{\pm} + X\ .$$

Although the LS requirement is quite effective to suppress QCD and known electroweak processes, fake-lepton backgrounds[1] still remain at a considerable level in the events of our signature. In our previous background studies [1] using data processed with the offline version 4, we separated the contributions of residual photon-conversions which survived our conversion veto from the rest of the fake-lepton backgrounds by knowing the conversion detection efficiency and the number of identified conversions. This note represents an update of the estimation of conversion detection efficiency mainly to use new data processed with the offline version `5.3.1` (the `0d` samples) and also to discuss several new points in the estimation of residual conversion backgrounds.

2 Photon-conversion detection efficiency

Once we know the detection efficiency, $\varepsilon_{\rm con}$, we can calculate a ratio of the residual conversions to the identified conversions, $R_{\rm res}$, by

$$R_{\rm res} = \frac{1-\varepsilon_{\rm con}}{\varepsilon_{\rm con}} \,. \qquad (1)$$

The amount of residual conversions in samples of our interest can be basically estimated by multiplying $R_{\rm res}$ to the number of identified conversions. The detection efficiency is defined by taking the denominator to be the conversions with one leg passing our electron selection criteria [2] except the isolation cut and conversion veto, which we refer to as the seed electron. The reason for turning off the isolation cut is rather technical. We use fake-event Monte Carlo (MC) samples created by generating single γ or π^0 to compare conversions with those in the real data, and the isolation variable in such fake-events is not usable in this respect. In §10, we will discuss how to handle the situation when we impose the isolation cut in the actual LS dilepton analysis.

The base efficiency is estimated as a function of partner-track p_T and decomposed into two factors in our analysis as

$$\varepsilon_{\rm con}(p_T) = \varepsilon_{\rm rel}(p_T) \cdot \varepsilon_{\rm pla} \,, \qquad (2)$$

where $\varepsilon_{\rm rel}(p_T)$ is the p_T dependent efficiency covering low-p_T regions and defined relative to the efficiency on the plateau, and $\varepsilon_{\rm pla}$ is the absolute efficiency on the plateau. These efficiencies could be further broken down to the track-finding efficiency and the efficiency of the identification[2] criteria given by

$$|\delta_{xy}| < 0.2 \text{ cm} \quad \text{and} \quad |\Delta \cot\theta| < 0.04 \,, \qquad (3)$$

where $|\delta_{xy}|$ is the distance of the closest approach points between a pair of tracks of conversion candidates. To clarify, our definition of conversions are those for which the

[1] We use "fake lepton" as a general word including non-prompt leptons such as those from photon conversions or from heavy-flavor decays, as well as literal fake leptons.

[2] We are using a modified version of `ElectronUser/ConversionFinder.cc` to get additional information from the function.

finder function returns the flag equal to 1, and the so-called tridents which have a 3rd track with the same charge as the seed electron are not included since they are likely to be prompt electrons.

The inefficiency of detecting conversions is dominated by that of reconstructing low momentum partner-tracks and manifests itself in $\varepsilon_{\rm rel}(p_T)$. It is not negligible even for high E_T seed-electrons because the kinematics of conversions relatively favors asymmetric energy sharing as shown in Fig. 1 which is a theoretical calculation of the relative probability of energy sharing as a function of fractional electron (e^-) energy [3]. Its asymmetric nature enhances the residual conversion events.

Our first goal is to measure the base efficiency, $\varepsilon_{\rm rel}(p_T)$, by comparing observed p_T spectrum of identified conversions to a perfect, or generator-level, MC spectrum. It turns out, however, that direct uses of this efficiency would not be useful in actual applications because it includes very low values at low p_T. As we can easily imagine from the form of the residual ratio, $(1-\varepsilon)/\varepsilon$, too-low efficiencies in a certain parametrization results in a significant loss of statitical power of predicting residual conversions, namely, we are forced to estimate the most needed part of residual conversions from the identified conversions with the least statistics. In order to avoid this problem, we transform the base efficiency to a function of different parameter, and we choose the parent-photon p_T for such parametrization. The base efficiency as a function of partner-track p_T is convoluted with the perfect p_T spectra of the MC for a given p_T bin of the parent photon to get $\varepsilon^{\gamma}_{\rm rel}(p_T)$. More details are explained in §5.

The next is to measure the plateau efficiency, $\varepsilon_{\rm pla}$, by using conversions in the real data identified with a method independent from our nominal identification algorithm.

Finally, the $\varepsilon^{\gamma}_{\rm con}(p_T)$ is obtained by multiplying this absolute efficiency to the relative efficiency:

$$\varepsilon^{\gamma}_{\rm con}(p_T) = \varepsilon^{\gamma}_{\rm rel}(p_T) \cdot \varepsilon_{\rm pla}\,. \tag{4}$$

3 Data samples

3.1 Real data

We use the `bhel0d` dataset for the most part of this study. We apply a pre-selection including bad-run removal using a list of the version 10.0a with "no Si" and "CMX ignored" options, and a trigger bit selection of `ELECTRON_CENTRAL_18_v*`. The sample corresponds to an integrated luminosity of $355.075 \times 1.019 = 361.8$ pb^{-1}. From this sample, we select events with at least one good electron satisfying our high-p_T electron selection ($E_T > 20$ GeV and $p_T > 10$ GeV/c) except the isolation cut and conversion veto. A summary of the statistics is given in Table 1.

3.2 Conversion Monte Carlo samples

There are four MC-samples used in this study. The two of them are made by generating events of single π^0 using the `FAKE_EVENT` module: one with a flat p_T spectrum and the other with a certain p_T slope. The rest of the samples are made by generating single-γ events instead of π^0. The generation parameters are listed in Table 2. The "Good" in the table means that there is at least one high-p_T good electron in the event. Simulations of π^0 decays and suceeding conversions of γ themselves are left to the CDF detector simulation package, `cdfSim`, of the offline version `5.3.3`. The simulated probability of conversion is about 18% per photon independent of its p_T.

Since p_T spectra of partner tracks are correlated with p_T of parent photons, we need to match MC slopes with those we observe in the data. We note here that the slopes in Table 2 are not necessarily the optimized ones but the values we just pick up for reasons nothing more than historical. In the MC tuning, we try to describe conversions found in the data with a mixture of the π^0 and γ MCs with optimized slopes. In order to improve the statistical power of the MCs in the optimization, the "flat samples" (the samples 1 and 3 in Table 2) are used to optimize the slopes, while the samples generated with the slopes in the table are combined appropriately, after re-weighting with the tuned slopes, to find out the optimum relative ratio γ/π^0.

4 Monte Carlo tuning

Tuning the MC samples is done by using identified conversions collected with an additional requirement of partner-track $p_T > 2$ GeV/c which is expected to be high enough to avoid non-plateau region. The slope parameter α is optimized by matching the weighted p_T spectra of reconstructed photon with what is observed in the real data, where the photon p_T is formed by the scalar-p_T sum of the seed-electron track and the partner track. On the other hand, we use E/p distributions to constrain the mixture of the π^0 and γ. The actual optimization of the three parameters, two slopes for the π^0 and γ, and the relative ratio γ/π^0, is carried out by an iterative way. Namely, 1) find a relative ratio for a given set of two slopes; 2) by adding the π^0 and γ MCs with the obtained ratio, we look for the best set of two slopes that describes photon p_T of the data using the χ^2; 3) go back to the first step and iterate. The optimized parameters[3] are

$$\alpha_{\pi^0} = 5.6 \pm 0.5\,, \qquad \alpha_\gamma = 4.9 \pm 0.5\,, \qquad \gamma/\pi^0 = 0.66 \pm 0.03\,. \tag{5}$$

Figure 2 shows the optimized E/p distribution compared to the data. The reduced χ^2 as a function of slope parameters is shown in Fig. 3. Figure 4 shows a comparison of reconstructed photon-p_T spectra between the MC and the data. We also demonstrate that our MC sample simulates the conversion points reasonably well in Fig. 5, and that the detailed p_T and E_T spectra of seed electrons are in good agreement with the data as shown in Figs. 6 and 7. The jump at 50 GeV seen in Fig. 7 is due to the E/p cut turned off at this E_T. From these figures, we say that our MC events reproduce observed conversions well in a wide kinematical range.

[3] The iteration converged in a couple of times.

5 Relative efficiency in low p_T region

Figure 8 compares the p_T spectrum of conversion partner-tracks observed in the real data with the one obtained at the generator level of MC, where the MC spectrum is fit to the data in the region of $2 < p_T < 5$ GeV/c. We clearly see losses of tracks in very low-p_T region. The efficiency relative to the plateau, $\varepsilon_{\rm rel}(p_T)$, is obtained by dividing the observed partner-track distribution by the fitted MC distribution, and is shown in Fig. 9, where the uncertainties in the figure are only statistical[4]. The efficiency reaches its plateau at $p_T \simeq 1$ GeV/c. Figure 9 also shows the MC prediction of the efficiency. The MC performance is actually good in reproducing the p_T dependence of the conversion-detection efficiency.

We investigate systematic uncertainties of the relative efficiency coming from the MC tuning parameters. This is done by shifting the parameters by the quoted uncertainties given in Eq. (5), and the result is summarized in Fig. 10 in terms of relative uncertainties. The uncertainty is dominated by MC statistics and at a level of 5%. The dependence on the MC tuning is small; only the slope dependence becomes non-negligible at very low p_T. The acuracy has been greatly improved compared to our previous analysis [1]. The main reason is that we used the high-p_T inclusive electron sample, which is a high statistics sample, to collect conversions.

The relative efficiency as a function of partner-track p_T is an important and the most basic efficiency, but it is not useful in actual applications. Residual conversions result from low efficiencies of detecting conversions, which in turn means that the statistical power of identified conversions is low as well. If we use the parametrization of the partner-track p_T, we would suffer from it to a considerable extent because there is an extremely low efficient point, and have to predict the most contributing residual conversions from the least statistics of identified conversions. A solution is to choose a different parametrization so that low and high efficiency regions are reasonably mixed and smeared out.

One might think that natual choices would be the E_T or p_T of seed electrons because the seed electron is a physics object we identify primarily in physics analyses. However, we found that such parametrizations had sample-to-sample depenences through different photon-p_T spectra. As an example, we consider the parametrization using the seed-electron E_T.

Before proceeding to an investigation of parametrization, we note that the transformation to a different parametrization is done by using only the base relative efficiency which has been just measured and the generator-level MC information regarding the partner-track p_T spectrum. The p_T cut on the seed-electron no longer has to be high in this situation because the generated MC particles cover low-p_T regions down to 3 GeV/c as given in Table 2. The kinematical cut in the actual LS dilepton analysis is asymmetric [2], where the low-p_T cut is defined to be $E_T > 6$ GeV and $p_T > 6$ GeV/c, and the primary estimation of residual conversion backgrounds in the LS dilepton analysis will be carried out for this low-p_T side of dilepton pairs. Given this, we mainly

[4]We are not using the binomial statistics but simple error propagation for ratios.

discuss the relative efficiency with different parametrizations by imposing the low-p_T cut in what follows.

For getting the relative efficiency as a function of seed-electron E_T, we convolute the base efficiency as a function of partner-track p_T with the generator-level spectrum of partner-track p_T in a given seed-electron E_T bin. In this case, the partner-track p_T spectrum is inevitably dependent on the parent photon-p_T spectrum which has basically a sample depenence. In fact, a comparison of photon-p_T spectra between the MC that is tuned by the inclusive electron data and LS lepton + conversion events[5] shown in Fig. 11 demonstrates such a dependence. We see that a harder p_T spectrum is favored in the lepton + conversion events. Our analysis shows that the p_T spectrum of $\propto p_T^{-2.5}$ of the γ MC describes the observed spectrum better, which should be compared to $p_T^{-4.9}$ tuned by the inclusive electron sample. The relative efficiency as a function of electron E_T for each case is shown in Fig. 12. There is a $\simeq 15\%$ difference at lower E_T regions, which translates to about a 40% change in terms of the residual ratio as calculated from

$$\frac{\Delta R_{\mathrm{res}}}{R_{\mathrm{res}}} = -\frac{1}{1-\varepsilon}\frac{\Delta\varepsilon}{\varepsilon} \, . \tag{6}$$

We could measure the slopes for samples of interest but it is not always possible due to, for example, limited statistics.

In order to obtain more stable efficiencies over different samples, we choose to use the parent-photon p_T itself to parametrize them. The convolution is expressed by

$$\varepsilon_{\mathrm{rel}}^{\gamma}(p_T) = \frac{\displaystyle\int f(\tilde{p}_T|p_T)\cdot\varepsilon_{\mathrm{rel}}(\tilde{p}_T)d\tilde{p}_T}{\displaystyle\int f(\tilde{p}_T|p_T)d\tilde{p}_T} \, , \tag{7}$$

where $f(\tilde{p}_T|p_T)$ represents the true (i.e. generator level) p_T spectrum of partner tracks for a given p_T of the parent photon. In actual calculation, the $\varepsilon_{\mathrm{rel}}(p_T)$ for the tracks with $p_T > 2$ GeV/c (normalization region) is set to 1. The obtained relative efficiency as a function of photon p_T is shown in Fig. 13, where uncertainties are statistical only. The drop of the efficiency for the p_T less than 15 GeV/c is a threshold effect and it does not mean that the intrinsic efficiency is low in this p_T region. In the figure, efficiencies from γ and π^0 MCs are separately shown, and they are consistent with each other. This gives us an advantage that the application is not senstive to the composition of γ and π^0 in samples. Related to this, we show a comparison of E/p distributions for conversions in the LS lepton+conversion events, the γ MC, and the π^0 MC in Fig. 14, where the MC slope-parameters are tuned by the LS lepton+conversion events themselves and the MC distributions are normalized to the data. We comment that the conversions found in the LS lepton+conversion events are more likely to be originating from photons, not from π^0. In fact, if we try to fit the data with the two

[5] The LS lepton + conversion events used here were collected from the whole high-p_T inclusive lepton datasets from the `0d` to the `0j` (up to the period 12) and corresponds to an integrated luminosity of about 2 fb^{-1} [2].

MC distributions, a fitter returns answers of negligible contributions from the π^0 MC. We use the efficiency obtained by using the γ MC from now on.

We mention here one complication in the calculation of statistical uncertainties included in Fig. 13. For a given p_T bin of the parent photon, the denominator shows a certain p_T distribution of partner tracks. This denominator is a sum of weights related to the MC slope. The numerator accepts another weight which is the relative efficiency $\varepsilon_{\rm rel}(p_T)$. The efficiency for a given p_T bin is thus symbolically written as

$$\varepsilon = \frac{\sum w'_i w_i x_i}{\sum w_i x_i}, \tag{8}$$

where x_i is actually 1 (event) for all i. The statistical uncertainty of ε is basically calculated by propagating the uncertainty of x_i, namely $\Delta x_i = 1$:

$$\Delta\varepsilon = \frac{\sum(w'_i w_i - \varepsilon w_i)\Delta x_i}{\sum w_i x_i} = \frac{\sum(w'_i w_i - \varepsilon w_i)}{D}\,, \tag{9}$$

where the $D = \sum w_i x_i$ is the integral of entries in the denominator. From this equation, we see that there arises a cross term proportioanl to $w' w^2$ when we compute $\langle(\Delta\varepsilon)^2\rangle$. We thus need to accumulate the square of weights of the form $\sqrt{w'}w$ in order to calculate the statistical uncertainty properly. As the systematic uncertainty of the relative efficiency as a function of photon p_T, we consider the input relative efficiency. Shifting them by $\pm\sigma$, we look at resulting changes of the final efficiency. A summary of the uncertainty calculations is shown in Fig. 15, where the total uncertainty is obtained by adding all the uncertainties in quadrature[6].

6 Absolute efficiency on the plateau

We need to identify conversions with a method independent from the nominal identification algorithm to measure the absolute efficiency on the plateau region. By the same way as the previous analysis, we use hit information of the CES strips for this purpose. It is based on a distinct feature of the conversion pair that the electron and the positron have approximately the same z position at any radius. When the electron and the positron separate enough so that they reach different ϕ wedges as illustrated in Fig. 16, we can measure each z position of the electron and the positron on the CES plane, and will find that the z position difference, Δz, has a peak around zero in the "right" side (RS) wedge as expected from the charge combination. We refer to this identification of conversions as the CES method. In our analysis, we look for the highest energy CES cluster in the nearest three towers along the η direction of the wedges next to the seed electron (see Fig. 16).

[6]Strictly speaking, the statistical uncertainty of the relative efficiency as a function of photon p_T and the statistical uncertainty of the input relative efficiency are correlated because we are using the same MC samples, but we simply add them in quadrature.

We go back to our initial electron sample (1195056 events) as described in §3.1 to measure the plateau efficiency. In order to make the geometrical configuration of conversions unambiguous, we here impose a set of "fiducial" cuts on the electrons, which essentially selects only the electrons located in the central region inside a tower. One requirement is that the local x coordinate of the extrapolated track at the CES plane satisfy $|x| < 15$ cm. Another cut is for the z direction to ensure that the partner track of a conversion is hitting the tower with the same η-index as the seed-electron tower. The appropriate definition of the z fiducial is obtained by looking again at identified conversions, but this time, only those with the η indices being different between the seed-electron tower and the tower the partner track is pointing. Figure 17 shows a distribution of the un-signed detector η for the seed electrons with such mismatching tower indices, where the partner-track p_T is required to be larger than 1 GeV/c, that is, in the plateau region. We avoid the η regions seen as peaks in this figure. Then, it is natural and relevant to require the CES cluster that we have found belong to the tower with the same η-index as the electron, i.e. exactly next to the electron tower and in the RS wedge. The last of our "fiducial" cuts is to require the number of tracks associated to the electron be just one in order to suppress the conversions with the partner track passing through the same tower as the seed electron, which better ensures the configuration we desire. Finally, as in the previous analysis, we demand a certain energy deposition in the RS tower to purify the sample: $E_{\rm strip}/p > 0.5$ and $E_{\rm CEM}/p > 0.7$, where the track momentum p is calculated from the assumed p_T of 2 GeV/c and the z position of the CES cluster. It is probably good to remember that the requirements introduced above are applied only to the information of the electrons and the RS tower, none is explicitly required on anything of partner tracks.

A demonstration of the Δz peaks for the RS and the wrong side (WS) wedge is given in Fig. 18 in which we see a clear separation between what we observe in the RS and in the WS. The bremsstrahlung emitted off an electron could produce a false Δz peak in the RS tower but the probability of such cases is confirmed to be very small, less than 0.1%, by using a large MC sample of single electrons.

The distribution is fit with Gaussian plus constant. The absolute efficiency is measured by looking at the reduction of the fitted Gaussian functions before and after our nominal conversion identification algorithm. The result is shown in Fig. 19. For records, the ratio of the background to the signal (Gaussian) part is about 20% and 10% before and after conversion identification, respectively. The efficiency is obtained to be $\varepsilon_{\rm pla} = 0.935 \pm 0.006$. Instead of using the fitted Gaussian functions, we also try to count the excess events over the fitted background, which gives $\simeq 2\%$ larger counts. The efficiency in this case is, however, still stable and found to be 0.94 ± 0.01. The contamination of low-momentum partner-tracks with $p_T < 1$ GeV/c in the Δz peak is checked using the sample after the conversion tagging cut and found to be small ($< 1\%$). We do not observe any large systematics in the E_T dependence but just fluctuations due to statistics. From these, we believe that the systematic uncertainty is about 1% (absolute) at most, thus we quote

$$\varepsilon_{\rm pla} = 0.935 \pm 0.006({\rm stat}) \pm 0.010({\rm syst})\,. \tag{10}$$

By performing the same analysis for the MC data, we obtain a MC prediction for the plateau efficiency which is found to be 0.966 ± 0.009.

7 Overall efficiency and residual-conversion ratio

We calculate the overall conversion-detection efficiency by $\varepsilon^{\gamma}_{\rm con}(p_T) = \varepsilon^{\gamma}_{\rm rel}(p_T) \cdot \varepsilon_{\rm pla}$. The result is shown in Fig. 20. We see that the $\varepsilon^{\gamma}_{\rm con}(p_T)$ increases from 30% to 90% as the photon p_T increases. Figure 21 shows the $R_{\rm res}$ as a function of photon p_T.

Since the $\varepsilon^{\gamma}_{\rm con}(p_T)$ is obtained by merely rescaling $\varepsilon^{\gamma}_{\rm rel}(p_T)$, the relative uncertainty from MC statistics is the same as that of $\varepsilon^{\gamma}_{\rm rel}(p_T)$. We then propagate it according to Eq. (1) to the uncertainty of the $R_{\rm res}$. The other uncertainty of the $\varepsilon^{\gamma}_{\rm rel}(p_T)$ is separately propagated to the uncertainty of the $R_{\rm res}$. The systematic uncertainty originating from the plateau efficiency is obtained by a similar propagation which is straightforward. A summary of the relative uncertainties of the $R_{\rm res}$ is given in Fig. 22. Figure 23 show the result for the case of the high-p_T cut to be applied to conversions in the high-p_T side of dileptons.

8 Breakdown of the efficiency on the plateau

It is further informative to get some ideas on the breakdown of the absolute efficiency on the plateau as a product of the track-finding efficiency and the efficiency of the identification criteria:

$$\varepsilon_{\rm pla} = \varepsilon_t \cdot \varepsilon_c\,, \tag{11}$$

where ε_t represents the track-finding efficiency and ε_c is the efficiency of the identification criteria. We use the same CES method for this estimation.

The sample (Δz peak) is exactly the same one we use for measuring the absolute efficiency on the plateau. The track search is performed by looking for the opposite-sign (OS) track that gives the minimum of

$$D \equiv \sqrt{(\delta_{xy})^2 + (\Delta \cot\theta)^2}\,. \tag{12}$$

If this track is pointing the tower where the CES strip cluster is found, it is declared as a track is found. Counting the Δz peak is the same as in the previous section. The obtained efficiency is

$$\varepsilon_t = 0.975 \pm 0.007\,. \tag{13}$$

It is straightforward to estimate the identification cut efficiency using the above conversion candidates with the partner track found. The result is

$$\varepsilon_c = 0.957 \pm 0.007\,. \tag{14}$$

The product of the two efficiencies is $\varepsilon_t \cdot \varepsilon_c = 0.93 \pm 0.01$ which is consistent with the direct measurement, Eq. (10). The MC predictions are 0.991 ± 0.006 and 0.971 ± 0.006 for the track finding efficiency and the efficiency of identification criteria, respectively.

9 Accidental track association

The conversion identification algorithm could accidentally pick up a track for an electron that is not originating from a conversion process. The probability of this accidental track association, or an over-efficiency in other words, serves us a correction that might be necessary in estimating residual-conversion backgrounds.

The official conversion finder only returns a flag of its final judgement regarding the conversion identification. If you look into more details in the identification process, we find some electrons that have multiple partner-tracks satisfying the tagging criteria. One can infer the level of the accidental track association by looking at the multiplicity of such tracks. Table 3 shows the number of electrons as a function of partner-track multiplicities, where the electrons are required to satisfy our calorimeter isolation[7] cut ($\mathrm{ISO}^{\mathrm{cal}}_{0.4} < 2$ GeV). In the table, we divide electrons into conversions and non-conversions. Note that the non-conversions do not have partner-tracks with the charge opposite to the seed-electron since they are identified as conversions otherwise, and the conversions do not have partner-tracks with the LS charge since they are identified as tridents otherwise. The ratios shown in the table are calculated sequentially. The OS track multiplicity of 1 for the conversion electrons is, of course, at a high probability because the electrons are already classified as conversions. We see that the ratios are rather stable at higher multiplicities and can conjecture that the over-efficiency is about 1–2%.

Another simpler way is to apply the conversion identification algrorithm to muons and to look at the fraction with "partner" tracks. We use the `bhmu0d` dataset (361.8 pb^{-1}) for this exercise, where muons are required to pass our high-p_T muon selection cuts [2]. From 220611 good muons, we find 3792 muons with partner tracks, thus, $(1.72 \pm 0.03)\%$. Based on these investigations, we conclude that the probability of accidental track association due to application of the conversion identification algorithm can be quoted as

$$p_{\mathrm{acc}} = (1.5 \pm 1.0)\% \,, \tag{15}$$

for isolated electrons.

We now look over how to use the probability of accidental track association. Suppose that we find N_e good electrons after applying the conversion veto. It is expressed using the true number of electrons N_e^* and the number of residual conversions N_{rc} as

$$N_e = (1 - p_{\mathrm{acc}})N_e^* + N_{rc} \,. \tag{16}$$

On the other hand, the number of conversions can be expressed by

$$N_c = N_c^* + p_{\mathrm{acc}} N_e^* \,. \tag{17}$$

[7]See [2] for the definition of our calorimeter isolation.

Then, the level of real-electron contamination in conversions, thus the level of over-estimates of residual conversions is given by

$$\frac{p_{\rm acc}N_e^*}{N_c} = \frac{p_{\rm acc}}{1-p_{\rm acc}}\frac{N_e - N_{rc}}{N_c} = \frac{p_{\rm acc}}{1-p_{\rm acc}}\left(\frac{N_e}{N_c} - R_{\rm res}\right) \simeq p_{\rm acc}\left(\frac{N_e}{N_c} - R_{\rm res}\right) . \quad (18)$$

For example, it is about 1% if the number of conversions is comparable to the number of good electrons and the residual ratio is 0.2 (corresponding to the photon p_T of about 15 GeV/c).

10 Application of residual-conversion ratio

We have obtained the primary result, the residual ratio as a function of photon p_T, which we aim at in this note. In this section, we consider a couple of issues when we face in realistic applications.

For a given conversion, we reconstruct the parent-photon p_T from the seed-electron and its partner track, then obtain a residual ratio. By summing these ratios, or weights, over all the conversions found in a sample of interest, we obtain an expectation of the total number of residual conversions. However, we usually need the kinematical information of expected residual conversions such as p_T, while we only know the p_T of parent photons because the residual ratios is parametrized by the photon p_T. A handle of predicting kinematical information of residual conversions is the fact that they are most likely to have very low partner-track p_T, less than 1 GeV/c, as seen from the base relative efficiency shown in Fig 9. We, thus, approximate the partner-track p_T of residual conversions, expected from a given conversion, to be (0.5 ± 0.5) GeV/c, and re-assign the difference from the original partner-track p_T to the seed-electron p_T:

$${p_T}^{(\rm part)} \rightarrow 0.5 \text{ GeV}/c, \qquad {p_T}^{(\rm seed)} \rightarrow ({p_T}^{(\rm seed)} + {p_T}^{(\rm part)} - 0.5) \text{ GeV}/c\,. \quad (19)$$

By this way, we can predict the kinematical information of residual conversions with an uncertainty of $\Delta p_T \simeq 0.5$ GeV/c. For the conversions that already have partner-track p_T less than 1 GeV/c, we do not perform this p_T re-assignment.

Yet we have a problem due to the fact that the plateau region is not fully efficient. It means that conversions with partner-track p_T of large values can still become residual conversions at small probabilities. Performing the above p_T re-assignment all the time is, thus, not really accurate, but using the original p_T is also a valid thing to do reflecting the overall inefficiency of conversion detection. Let us call residual conversions with the partner-track p_T less than 1 GeV/c "proper" and those with p_T larger than 1 GeV/c "non-proper". More precise predictions are then given by splitting residual conversions into the proper and non-proper part. The fraction of proper residual-conversions for each photon p_T bin, the "splitting" factor, can be constructed from the MC by weighting the generator-level spectrum of partner-track p_T with $1 - \varepsilon_{\rm con}(p_T)$, where the $\varepsilon_{\rm con}(p_T)$ is the conversion detection efficiency as a function of parter-track

p_T as given in Eq. (2). The result is shown in Fig 24. We form a function through fitting to the plot for the γ MC, which is given by

$$\min(1, f_L), \qquad f_L \equiv (2.6 \pm 0.1) \times p_T^{-(0.501\pm0.008)}, \tag{20}$$

where the L reminds us that the parametrization corresponds to the low-p_T cut. A similar fraction for the case of high-p_T cut is given in Fig. 25. The parametrized function is

$$\min(1, f_H), \qquad f_H \equiv (5.0 \pm 0.1) \times p_T^{-(0.663\pm0.009)}. \tag{21}$$

The procedure for predicting residual conversions now includes one more step after obtaining the residual ratio w: we further get a splitting factor f using the functions above. Then the weight corresponding to the kinematics given by the p_T re-assignment is wf, while the weight $w(1-f)$ is applied when the original seed-electron p_T is used as the p_T of predicted residual conversion.

Finally, we consider another detail regarding the isolation cut. We impose a calorimeter isolation cut in the actual LS analysis. The residual conversions, therefore, must pass this cut too. But our conversion studies so far are carried out without applying the isolation cut (except in §9) because the MC samples of fake-events are not sufficient to simulate such information. What we need here is to introduce effects of the isolation cut specific to residual conversions that are not common to identified conversions. As we expect, residual conversions tend to pass the isolation cut more than identified conversions because the partner-track activity is less harmful. Our approach to introduce such effects is to subtract partner-track p_T from the original isolation value, then apply the nominal isolation cut. This subtraction is coordinated with the prediction of kinematic information for expected residual conversions just described in the previous paragraphs, i.e. when the partner-track p_T is assumed to be 0.5 GeV/c, the difference from the original value is subtracted from the isolation, and when the residual conversion is considered to be non-proper, we do not do anything to the isolation but just apply the cut as usual. Further we require the following conditions to trigger the subtraction in order to pick up only the case where the original isolation value is certainly affected by the partner track:

- the partner track is pointing a different wedge from the one of the seed electron;
- the partner track is within the isolation cone radius of 0.4;
- the partner track is passing through the local x coordinate at the CES plane in the region $|x| < 21$ cm to avoid ϕ cracks.

The contribution of partner-track p_T to the isolation is checked by looking at the ratio between the two (just like E/p for electron clusters but for the isolation cone in this case) and shown in Fig. 26. We confirm a clear peak around unity in the plot. The left side of the peak can be considered to show detector-resolution effects, and the tail in the right side can be attributed to the effects from activities around the conversions, i.e. what the isolation is meant to measure. We argue that this scheme would be capable of respecting sample dependences of the isolation, if any, because what we are trying to subtract is the activity only from the conversion partner-track, while underlying events are intact.

11 Example estimation of residual conversions

As an example, we show results of residual-conversion estimation for LS dielectron events in four datasets, `bhel0d`, `bhel0h`, `bhel0i`, and `bhel0j` (up to the period 12), which corresponds to an integrated luminosity of $\simeq 2$ fb^{-1}.

In our initial LS dilepton analysis, we examined the selected events on the 2-dimensional plane of p_{T2} vs. p_{T12}, where the p_{T2} is the transverse momentum of the track of the 2nd lepton (low-p_T side of dilepton) and the p_{T12} is the transverse momentum of the dilepton system calculated using the track momenta [4]. On this plane, we introduced 4 regions as listed in Table 4. We now pick up the region 2 as an example and show numbers obtained without vetoing Z events.

Table 5 lists the estimated residual conversions for several cases of analysis schemes. The variation of method details is cumulatively applied. The 2nd row is the case that we do not perform subtraction of partner-track p_T from the isolation. Since this makes the probability of passing the isolation cut smaller, the number gets down by about 30%. If the application of splitting factors is stopped, which corresponds to the 3rd row, the partner-track p_T is always assumed to be 0.5 GeV/c. There is about a 5% change. The 4th row is the case when the p_T re-assignment is not performed, thus the partner-track p_T is kept at the original value, which means the p_{T2} (seed-electron p_T) stays at a lower value compared to the re-assigned p_T. The last row shows the result when we use the residual ratios as a function of seed-electron E_T instead of photon p_T. As discussed in §5, the conversion detection efficiency is underestimated to lead to larger values of residual conversions. The effect is about $24.1/15.6 \simeq 55\%$.

Following the discussion in §9, we check the level of over-estimation of residual conversions due to accidental track association by the conversion identification algorithm. The number of good LS dielectron events in the region 2 is 69 events, while the number of conversions used to predict the 19.6 events is 79. The effect of real-electron contamination in the predicted residual conversions is therefore estimated to be $1.5\% \times (69 - 19.6)/79 \simeq 1\%$.

It would be interesting to compare the data-driven estimation (19.6 ± 2.9), which has just been presented, with MC predictions[8]. After the same event selection for MC samples as the real data, we look at `OBSP` information to check if a given electron is coming from a photon conversion and separate the contribution of residual conversions. Table 6 shows the result, where we also analyzed WZ, ZZ, and $t\bar{t}$ to find them neglible, thus omitted from the table. With the quoted uncertainties, they are barely consistent with each other, but the MC predictions might be underestimated, which we will keep

[8] Relevant MC scale-factors [5] are all applied. More detailed studies on the backgrounds of LS dilepton events including other sources such as fake-leptons and physics processes are given in [6].

our eyes on as more data are collected in the course of the Run II.

12 Stability over different datasets

The conversion detection efficiency and residual ratio have been measured using the `bhel0d` dataset in the previous sections. We here check the stability of these numbers over the four datasets. Figure 27 compares the basic relative efficiencies as a function of partner-track p_T between four different datasets. A structure is common to all the datasets because the MC data used to obtain the efficiency are the same. We do not observe any significant changes among the datasets. The plateau efficiencies for all the datasets are listed in Table 7. The plateau efficiency is stable as well.

13 Requiring silicon hits

It was known from the previous analysis [1] that the largest backgrounds in the LS dilepton events are residual conversions. In order to improve the signal to background ratio, we will eventually impose a cut requiring Si hits on lepton tracks. Correspondingly, we investigate the case of analyzing conversions with the seed-electron track satisfying our Si-hit cut

$$\text{the number of Si hits} \geq 3\,. \tag{22}$$

The results and disussions are mostly the same as what we described in the previous sections, thus we only briefly show the main points.

Figure 28 shows the base relative efficiency as a function of partner-track p_T. We see that the efficiency rise is slightly shifted toward the low p_T direction. The plateau efficiency is estimated to be

$$0.95 \pm 0.01\ (\text{stat}) \pm 0.01\ (\text{syst})\,, \tag{23}$$

which is virtually the same as the result without the Si cut. The overall detection efficiency and residual ratio are then shown in Figs. 29 and 30, respectively, for the case of the low-p_T cut. The relative uncertainty of the R_{res} is about 25% (not shown). The splitting functions are obtained to be

$$f_L = (3.5 \pm 0.2) \times p_T^{-(0.67\pm0.02)}\,, \tag{24}$$

and

$$f_H = (10.3 \pm 1.3) \times p_T^{-(0.94\pm0.04)}\,. \tag{25}$$

As the photon p_T increases, the factors decrease a little faster than those in the previous section, reflecting the slight extension of the plateau region of the detection efficiency toward the lower p_T.

14 Conclusions

We measured the conversion-detection efficiency using a high-p_T inclusive electron sample and Monte Carlo samples. The efficiency as a function of partner-track p_T reaches its plateau at $p_T \simeq 1$ GeV/c with the plateau efficiency of $\varepsilon_{\text{pla}} = 0.935 \pm 0.006(\text{stat}) \pm 0.010(\text{syst})$. We converted these results to the efficiency as a function of photon p_T, which is about 30% at 6 GeV/c and reaches 90% at very high p_T, above 40 GeV/c or so. Using it, we obtained the residual-conversion ratio which is about 2.5 at 6 GeV/c and decreases to $\simeq 0.1$ at high p_T. The relative uncertainty of the residual ratio is about 15% for the p_T region we considered. An attempt was made to separately estimate the track-finding efficiency and the efficiency of the conversion identification criteria themselves for conversions with high partner-track p_T, and found 0.975 ± 0.007 and 0.957 ± 0.007, respectively. We did not find any significant problems of the Monte Carlo: the efficiencies are in good agreement with the data to a few %. We also estimated the probability of accidental track association as a result of applying the conversion identification algorithm, which was found to be $(1.5 \pm 1.0)\%$ for isolated electrons. Some issues in realistic application were considered and the solutions were presented. An example of estimating residual conversion backgrounds was given for like-sign dielectron events using four datasets from the `bhel0d` to the `bhel0j` (up to the period 12) corresponding to an integrated luminosity of 2 fb^{-1}. The long-term stability of efficiencies was confirmed using the same four datasets. Finally, the result when we required silicon hits for seed-electron tracks was included envisaging the use of such cuts to reduce residual conversion backgrounds in like-sign dilepton events.

Table 1: Event reduction of the real data.

	Events
Input (`bhel0d`)	26499561
Bad-run removal	19242849
Trigger (`ELECTRON_CENTRAL_18_v*`)	13820276
$\lvert z_{\rm pv}\rvert < 60$ cm	12872227
High p_T good electrons ≥ 1	1195056

Table 2: Monte Carlo samples.

Sample	Particle	min-p_T (GeV/c)	max-p_T (GeV/c)	Slope α (${p_T}^{-\alpha}$)	max-$\lvert\eta\rvert$	Generated (events)	Good (events)
1.	π^0	3	150	0 (flat)	1.3	1520000	155651
2.	π^0	15	150	6.3	1.3	1600000	72500
3.	γ	3	150	0 (flat)	1.3	1580000	123470
4.	γ	15	150	4.8	1.3	1580000	112935

Table 3: The number of electrons for different multiplicities of partner tracks passing the conversion identification.

	Non-conversions		Conversions	
Multiplicity	(Like sign)	Ratio (%)	(Opposite sign)	Ratio (%)
All	331262		257305	
1	7423	(2.24 ± 0.03)	251943	(97.9 ± 0.3)
2	175	(2.4 ± 0.2)	5248	(2.08 ± 0.03)
3	3	(2 ± 1)	113	(2.2 ± 0.2)
4	0		1	(0.9 ± 0.9)
5	0		0	

Table 4: Definition of the regions on the p_{T2} vs. p_{T12} plane.

Region	Cuts
1 (Upper left)	$p_{T12} < 35$ GeV/c and $p_{T2} \geq 18$ GeV/c
2 (Lower left)	$p_{T12} < 35$ GeV/c and $p_{T2} < 18$ GeV/c
3 (Lower right)	$p_{T12} \geq 35$ GeV/c and $p_{T2} < 18$ GeV/c
4 (Upper right)	$p_{T12} \geq 35$ GeV/c and $p_{T2} \geq 18$ GeV/c

Table 5: Residual conversions estimated for like-sign dielectron events in the region $p_{T12} < 35$ GeV/c and $p_{T2} < 18$ GeV/c for several varieties of method details as described in the texts. The data were collected from high-p_T inclusive electron datasets, corresponding to an integrated luminosity of 2 fb^{-1}, without vetoing Z events.

Variation of method details	Residual conversions
Nominal	19.6 ± 2.7(stat) ± 1.1(syst)
+ No p_T subtraction from isolation	14.3 ± 2.3(stat) ± 0.8(syst)
+ No splitting factor	13.6 ± 2.3(stat) ± 0.8(syst)
+ No p_T re-assignment	15.6 ± 2.4(stat) ± 0.8(syst)
+ Use a function of seed-electron E_T	24.1 ± 3.5(stat) ± 1.1(syst)

Table 6: Monte Carlo prediction of residual conversions for the like-sign dielectron events in the region $p_{T12} < 35$ GeV/c and $p_{T2} < 18$ GeV/c. Relevant scale factors are all applied and the normalization corresponds to an integrated luminosity of 2 fb^{-1}.

Process	Residual conversions (statistical uncertainty only)
$W \rightarrow e\nu$	4.2 ± 2.4
$Z \rightarrow ee$	6.6 ± 2.1
$Z \rightarrow \tau\tau$	0.77 ± 0.29
WW	0.02 ± 0.01
Total	11.6 ± 3.2

Table 7: Plateau efficiencies for the four datasets. The uncertainties are statistical.

Dataset	$\varepsilon_{\mathrm{pla}}$
`bhel0d`	0.935 ± 0.006
`bhel0h`	0.945 ± 0.007
`bhel0i`	0.937 ± 0.006
`bhel0j`	0.933 ± 0.006

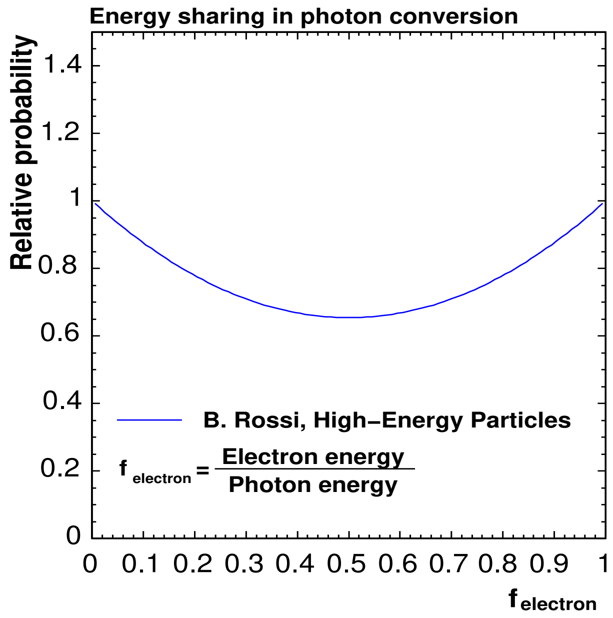


Figure 1: A theoretical calculation of the relative probability of energy sharing in the photon conversion as a function of fractional electron (e^-) energy [3].

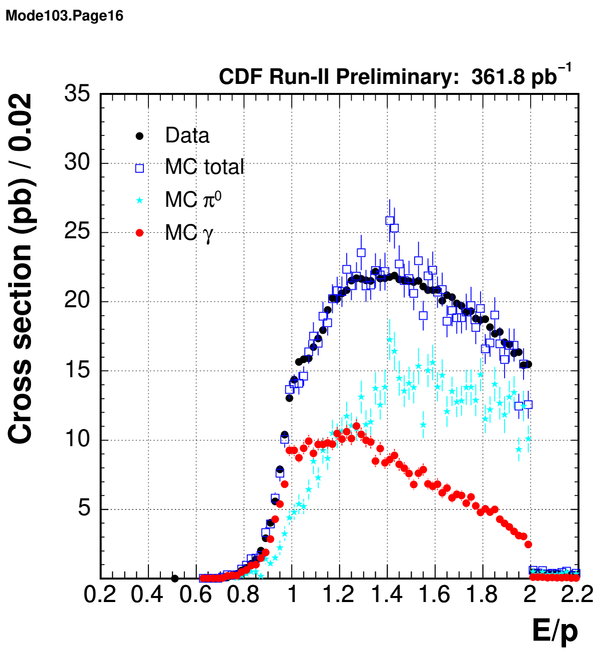


Figure 2: Distribution of E/p for the seed electron of conversions.

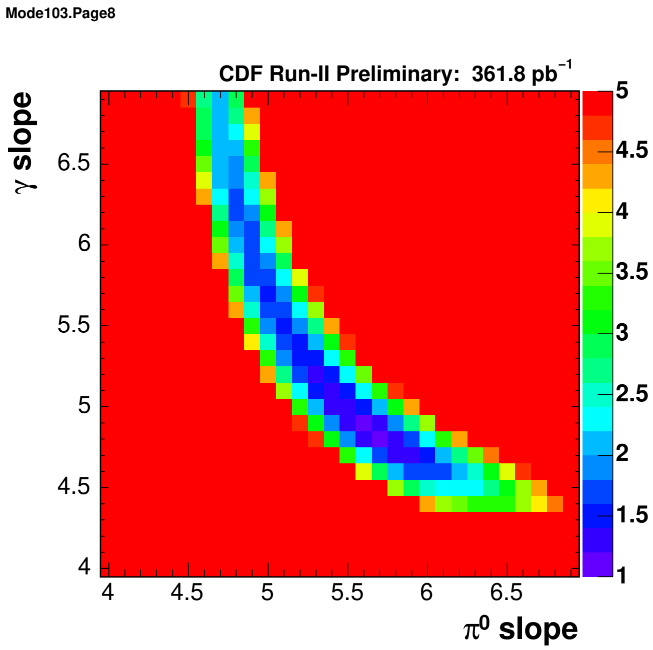


Figure 3: Reduced χ^2 between the data and the Monte Carlo distributions of reconstructed photon-p_T as a function of Monte Carlo slope-parameters.

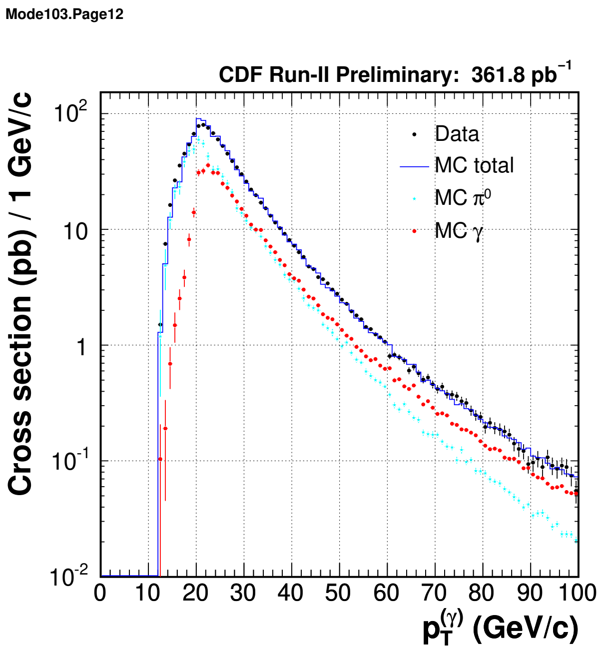


Figure 4: The reconstructed photon-p_T spectrum.

Mode107.Page7

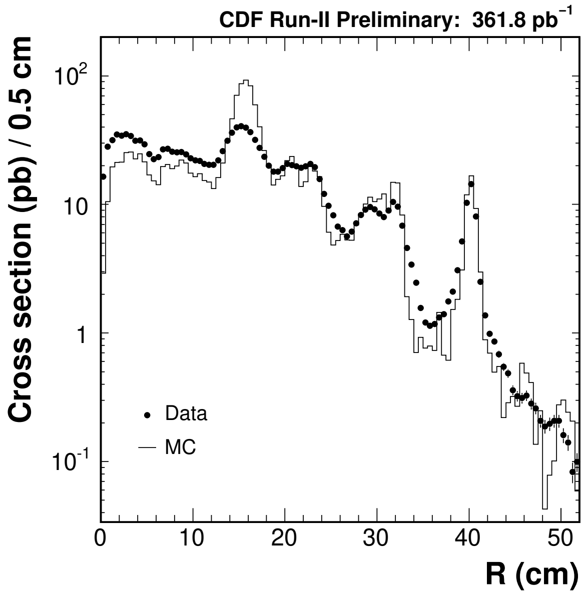


Figure 5: Distribution of the radius of conversion points from the beam line.

Mode103.Page28

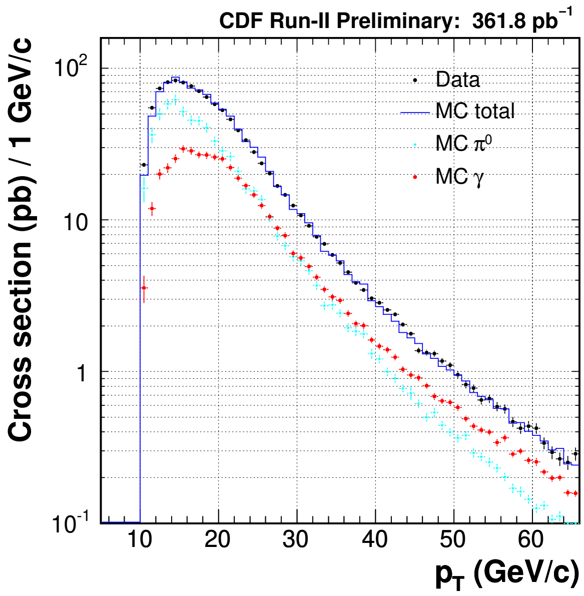


Figure 6: Conversion-electron p_T spectrum.

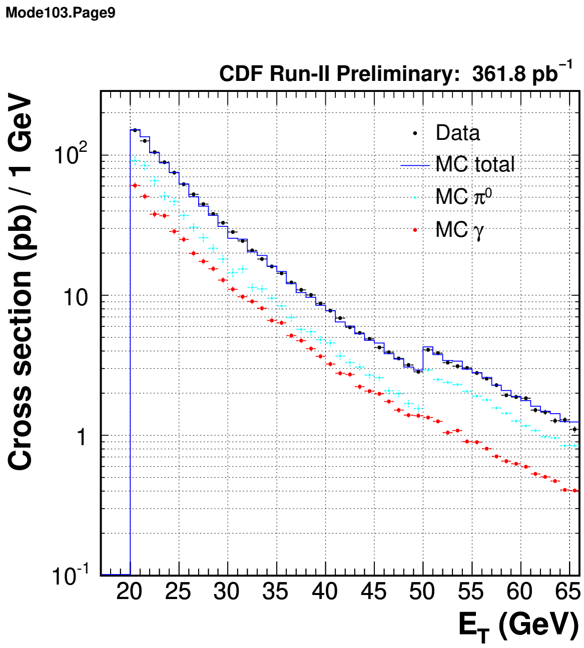


Figure 7: Conversion-electron E_T spectrum.

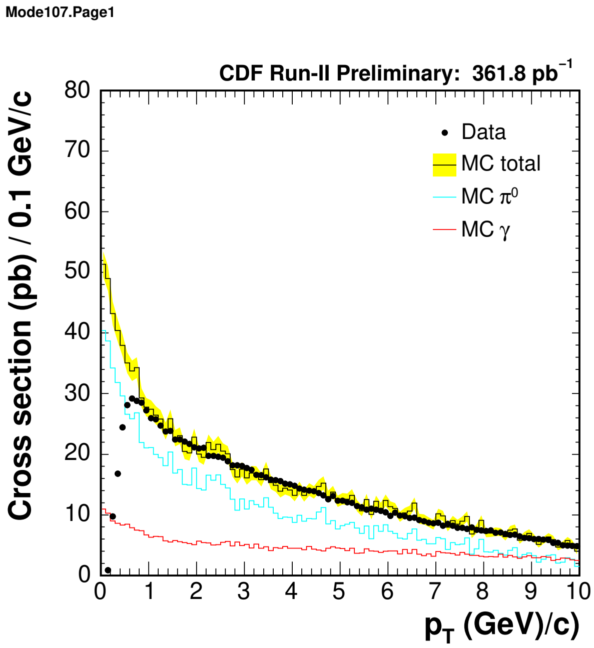


Figure 8: Partner-track p_T spectrum for conversions. The Monte Carlo spectrum is shown at the generator (OBSP) level and fitted to the data in the p_T region $2 < p_T < 5$ GeV/c.

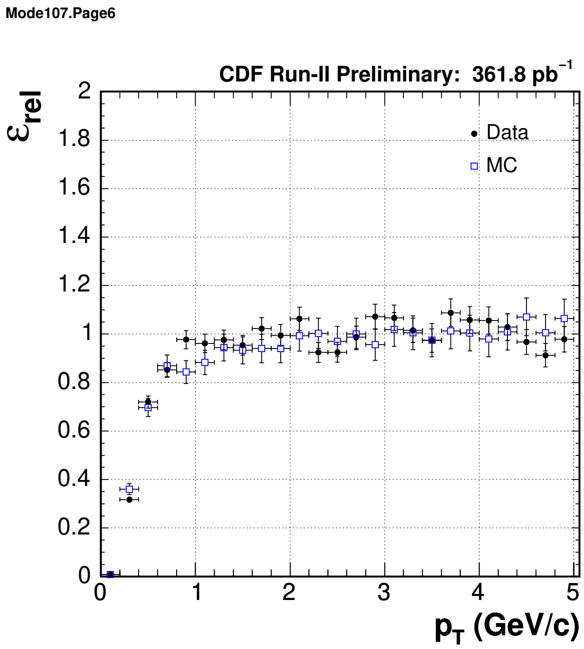


Figure 9: Relative conversion-detection efficiency as a function of partner-track p_T. Only statistical uncertainties are shown.

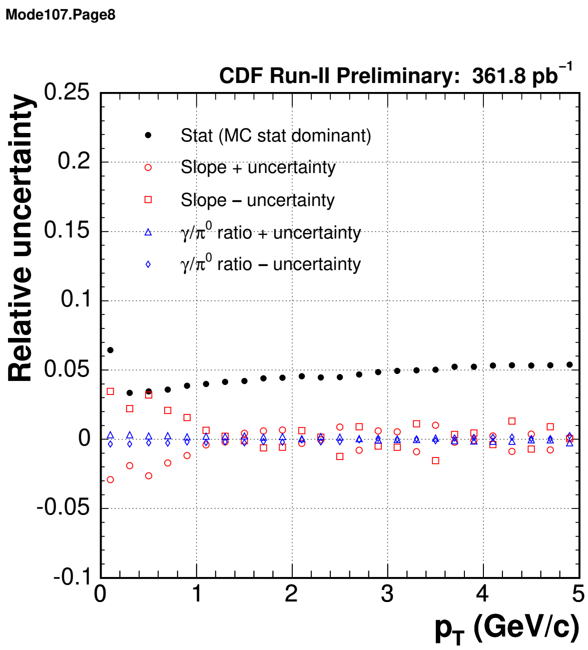


Figure 10: Relative uncertainty of the relative conversion-detection efficiency originating from different sources.

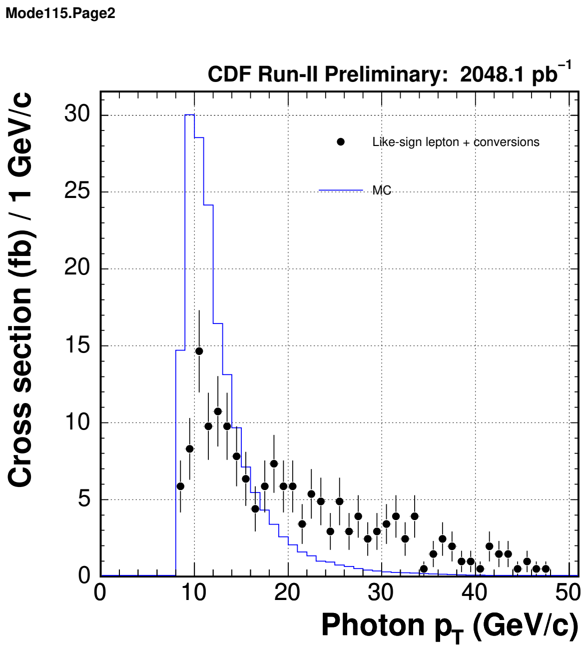


Figure 11: Comparison of photon-p_T spectra between like-sign lepton + conversion events and Monte Carlo tuned by conversions in inclusive electron data.

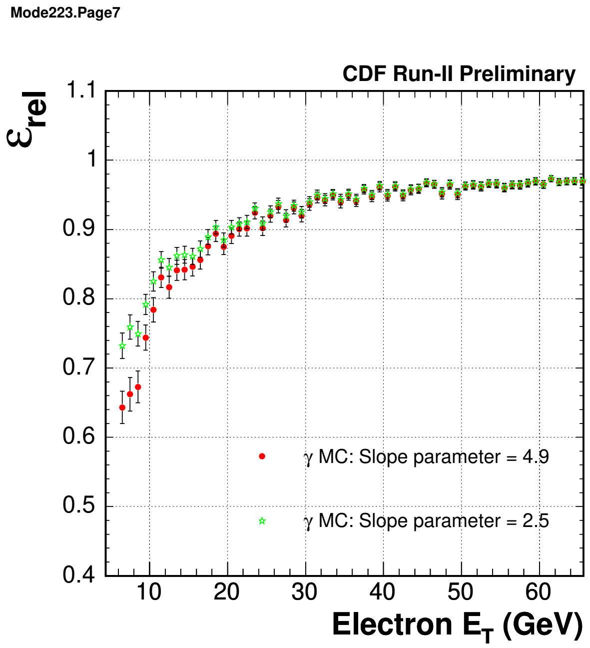


Figure 12: Relative conversion-detection efficiency as a function of seed-electron E_T for two cases of slope parameters in MC tuning.

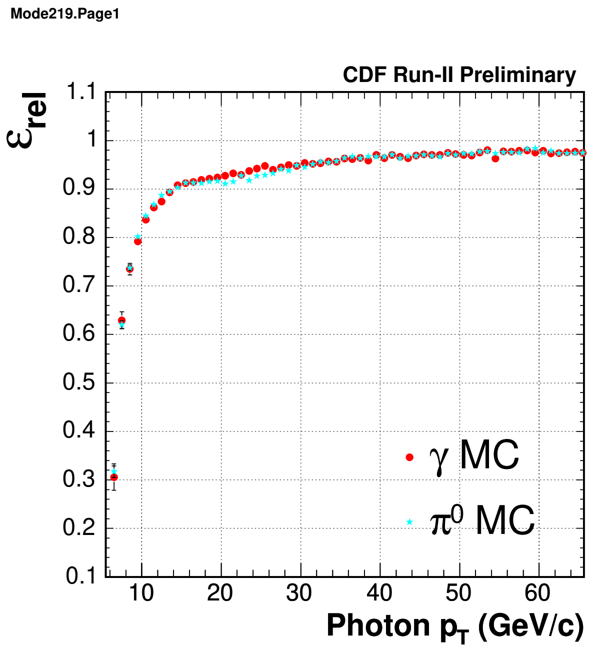


Figure 13: Relative conversion-detection efficiency as a function of photon p_T. The uncertainties are statistical only.

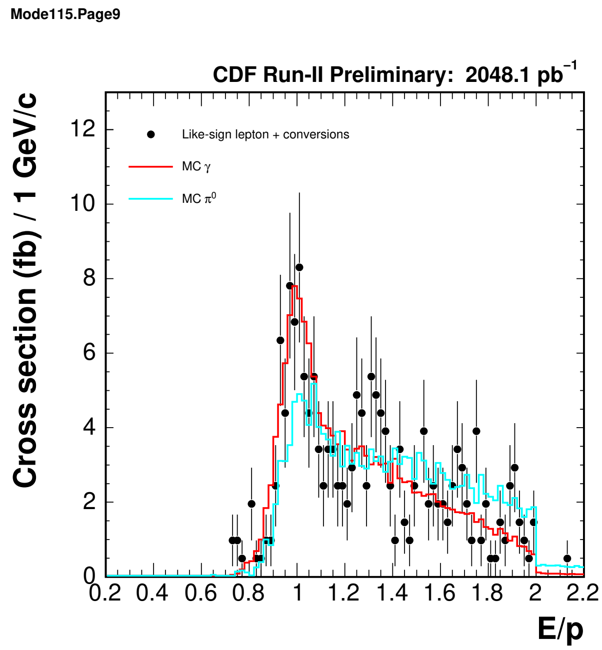


Figure 14: Comparison of E/p distributions for conversions in like-sign lepton+conversion events, the γ Monte Carlo, and the π^0 Monte Carlo. The p_T-slope parameters of the Monte Carlo data are tuned by the like-sign lepton+conversion events. The Monte Carlo distributions are normalized to the data. The uncertainties of the real data are statistical only.

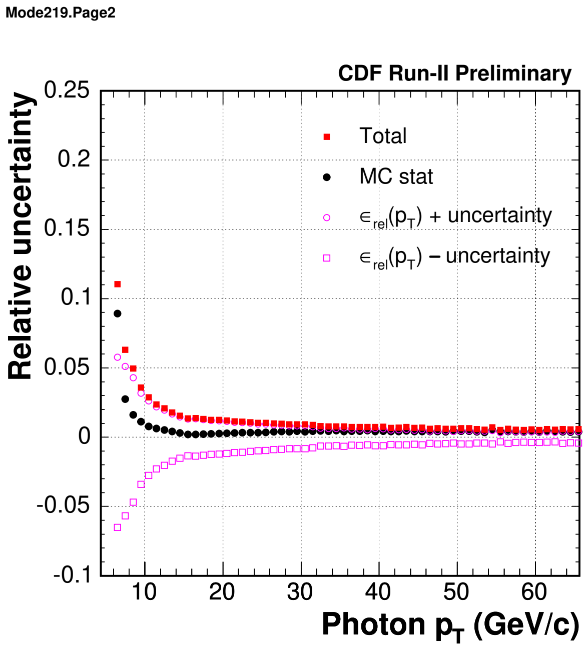


Figure 15: Relative uncertainty, shown for different sources, of the relative conversion-detection efficiency as a function of photon p_T.

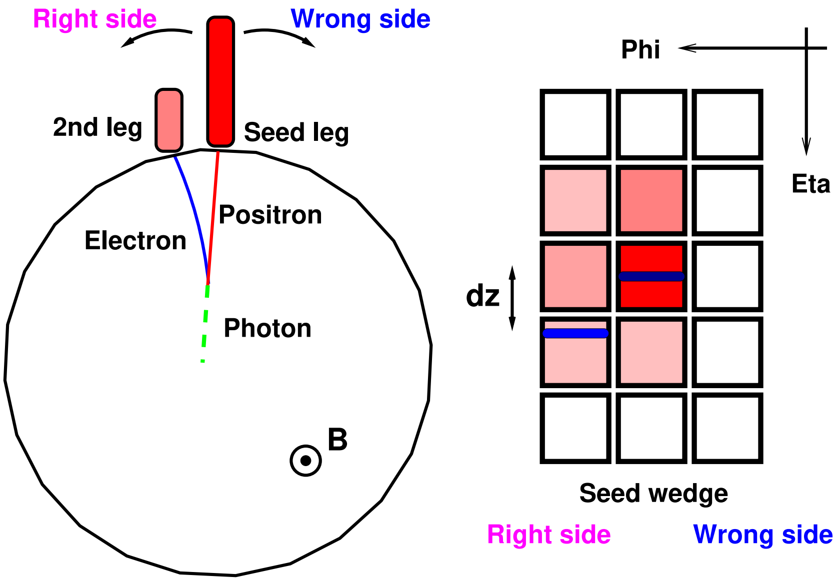


Figure 16: Illustration of the conversion tagging by means of CES strip information.

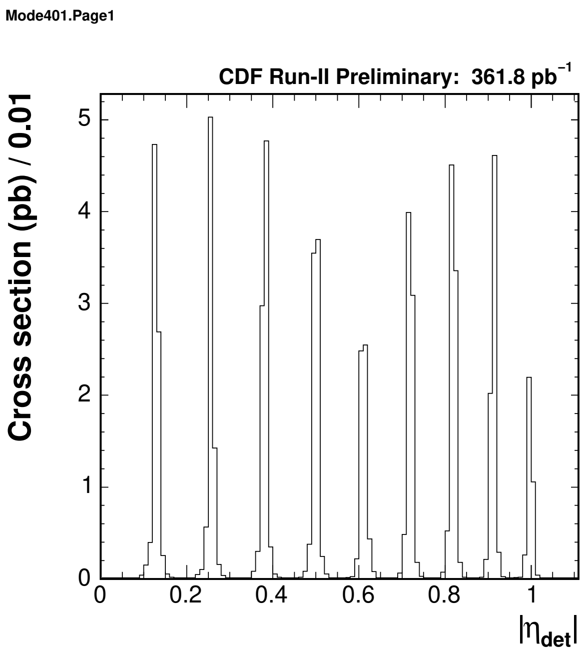


Figure 17: Distribution of absolute detector η for conversion electrons with the partner track hitting the tower with an η-index different from the electron.

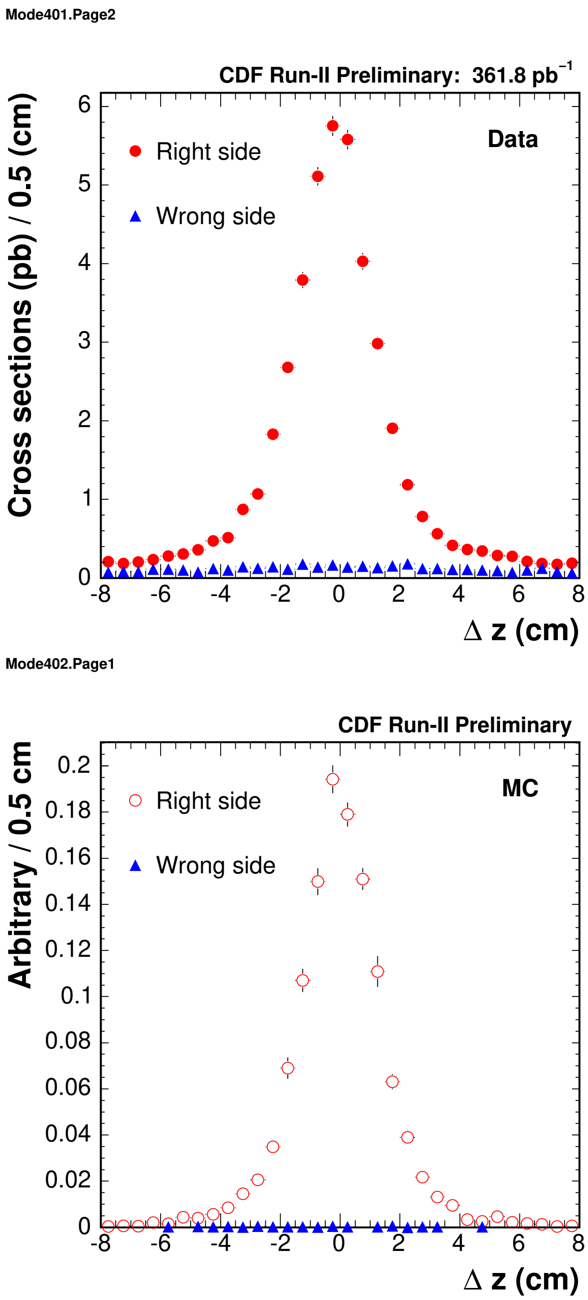


Figure 18: Distribution of z position difference on the CES plane between the seed electron and the highest-energy CES cluster in the nearest three towers of each ϕ side (right side and wrong side).

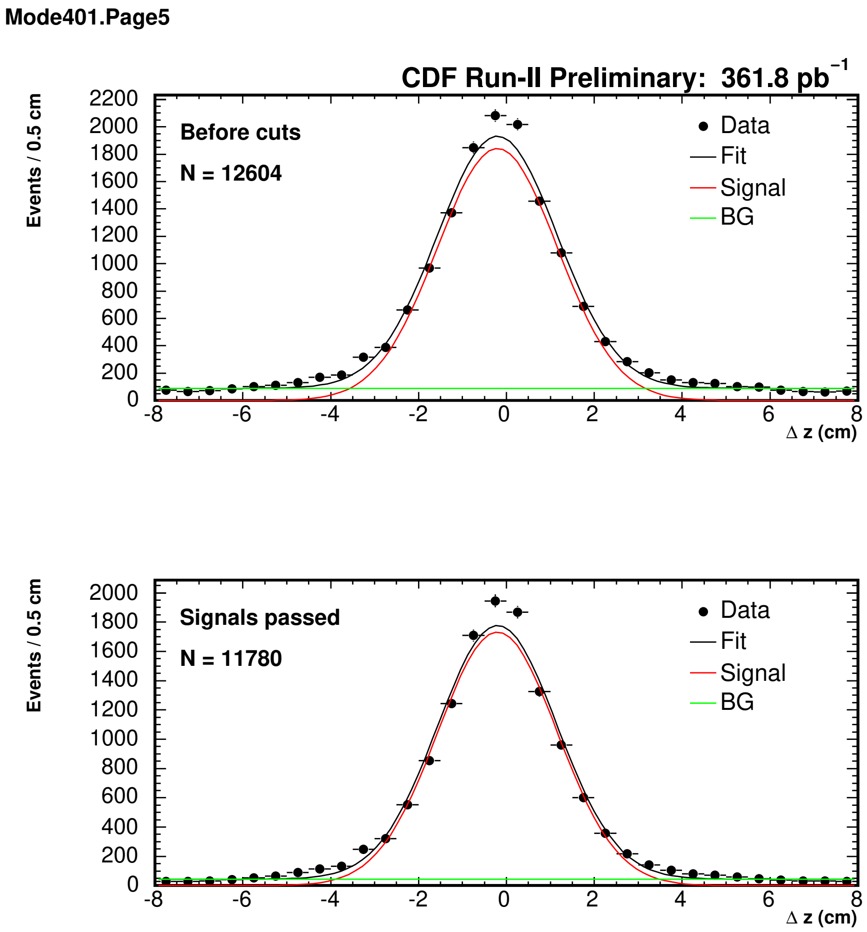


Figure 19: Distribution of z position difference before the conversion tagging cut (top) and after (bottom).

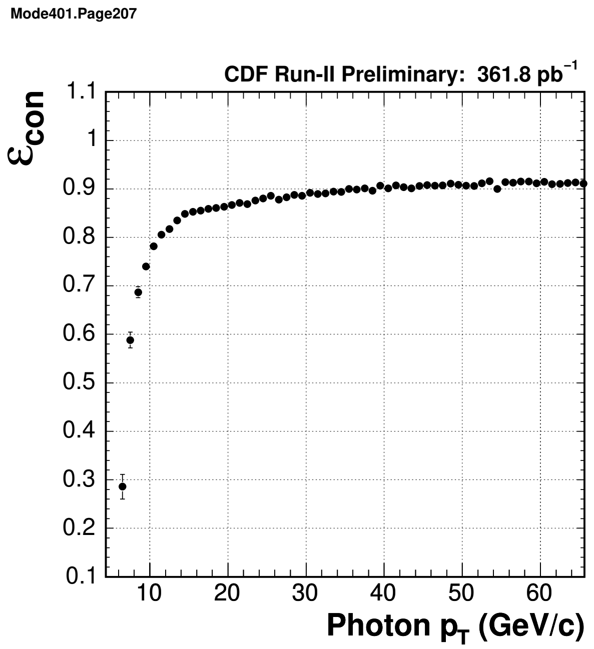


Figure 20: Conversion-detection efficiency as a function of photon p_T. The uncertainties are statistical only.

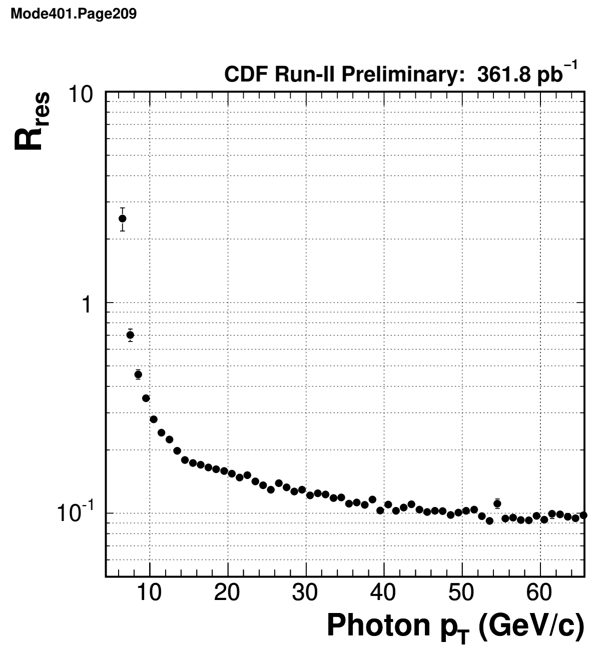


Figure 21: Residual ratio as a function of photon p_T. The uncertainties are statistical only.

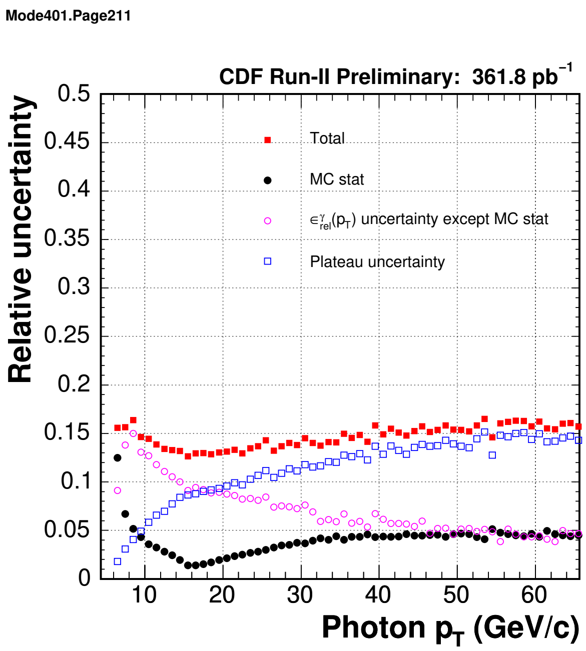
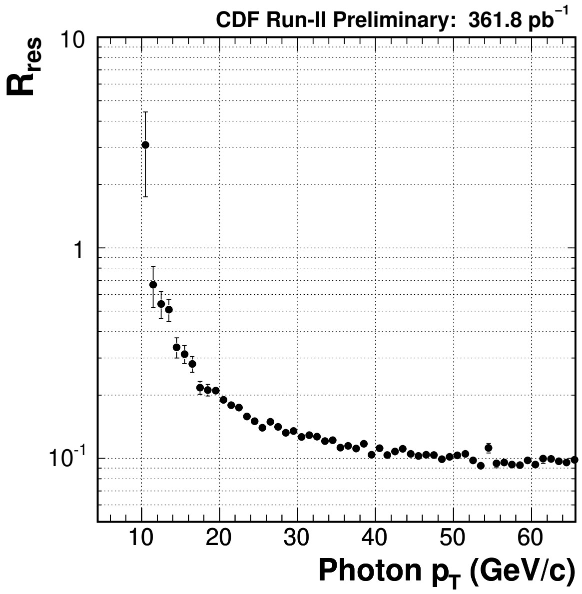


Figure 22: Relative uncertainty, shown for different sources, of the residual ratio as a function of photon p_T.

Mode401.Page210



Mode401.Page212

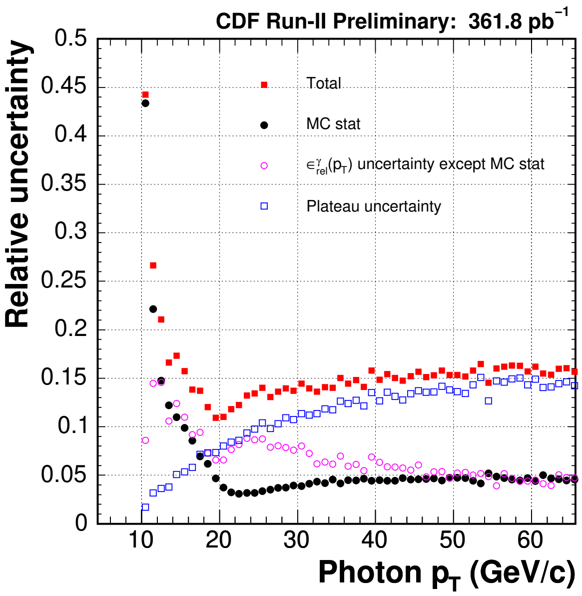


Figure 23: Residual ratio (top) and its relative uncertainties (bottom) as a function of photon p_T for the case of our high-p_T cut, $E_T > 20$ GeV and $p_T > 10$ GeV/c. The uncertainties in the top plot are statistical only.

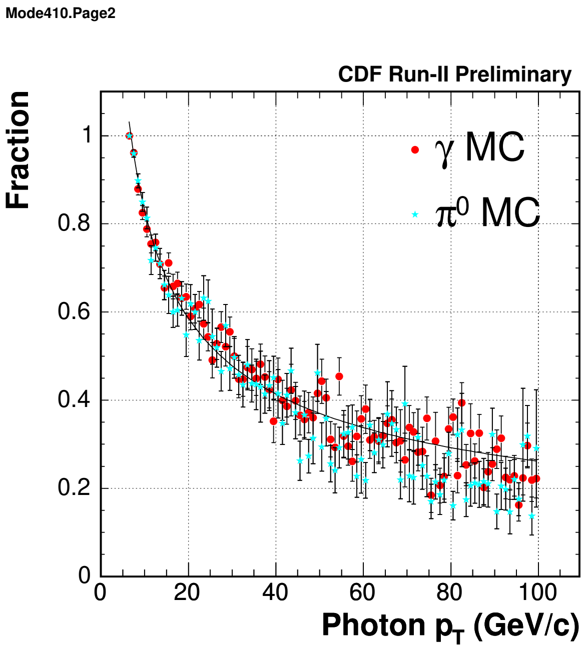


Figure 24: Fraction of residual conversions with partner-track p_T less than 1 GeV/c as a function of photon p_T for the case of low lepton-p_T cut.

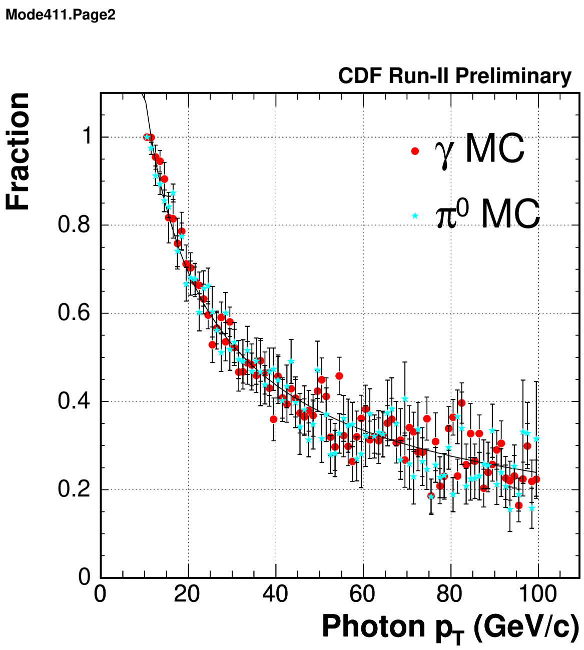


Figure 25: Fraction of residual conversions with partner-track p_T less than 1 GeV/c as a function of photon p_T for the case of high lepton-p_T cut.

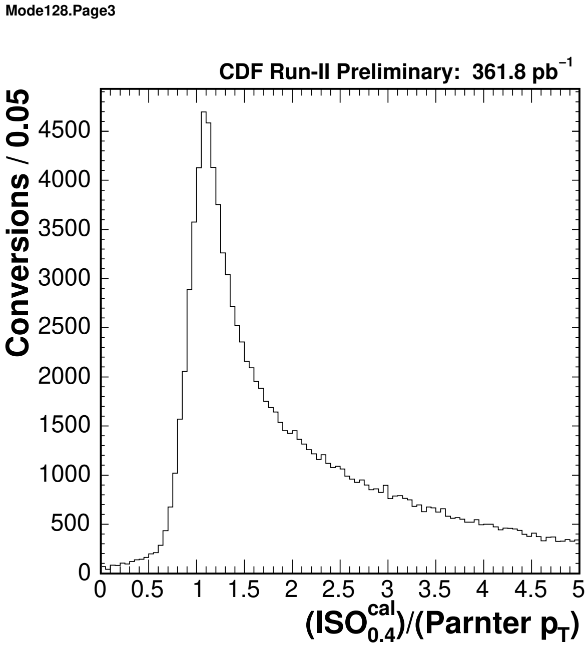


Figure 26: Ratio between the calorimeter isolation and the partner-track p_T for conversions.

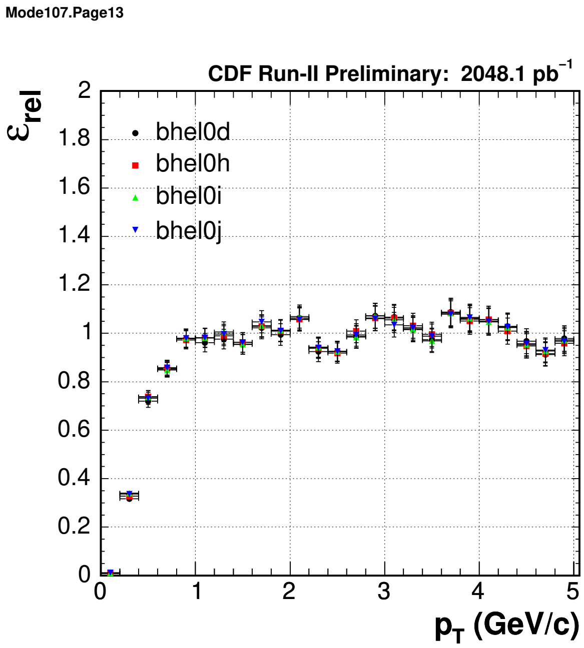


Figure 27: Comparison of relative conversion-detection efficiencies as a function of partner-track p_T between four datasets.

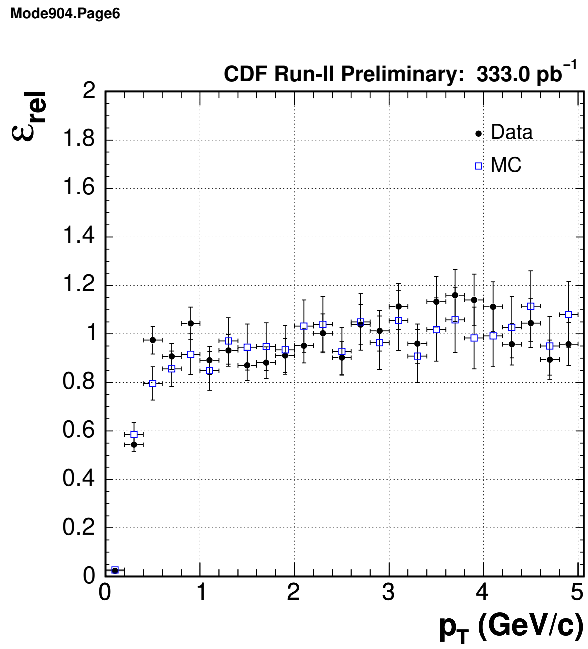


Figure 28: Relative conversion-detection efficiency as a function of partner-track p_T for the case that silicon hits are required for the seed-electron track. Only statistical uncertainties are shown.

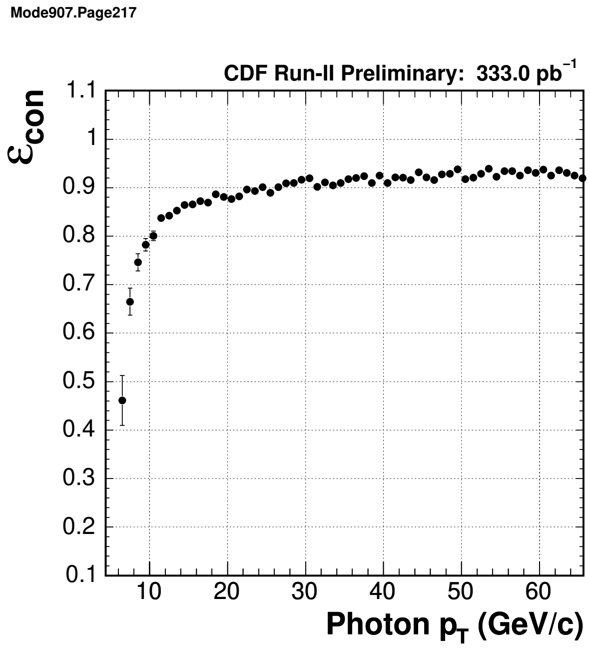


Figure 29: Conversion-detection efficiency as a function of photon p_T for the case that silicon hits are required for the seed-electron track. The uncertainties are statistical only.

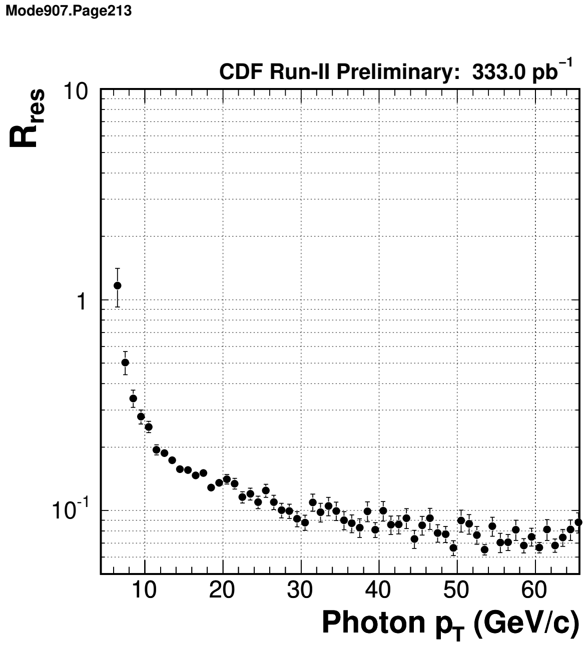


Figure 30: Residual ratio as a function of photon p_T for the case that silicon hits are required for the seed-electron track. The uncertainties are statistical only.

References

[1] Hirokazu Kobayashi, Yoshihiro Seiya, and Kazuhiro Yamamoto, "Background Estimation for High-p_T Like-Sign Dilepton Analysis", CDF7261 (Version 1.0).

[2] Toru Okusawa, Yoshihiro Seiya, Takayuki Wakisaka, and Kazuhiro Yamamoto, "Baseline Selection of High-p_T Dilepton Events to Search for the Wh Production", CDF7260 (Version 2.0). Our selection is essentially the same as the standard high-p_T cuts.

[3] B. Rossi, High-Energy Particles, Prentice-Hall.

[4] Hirokazu Kobayashi, Kazuhiro Yamamoto, and Yoshihiro Seiya, "Search for the WH Production Using High-p_T Isolated Like-Sign Dilepton Events in Run II", CDF7262 (Version 1.0).

[5] Toru Okusawa, Yoshihiro Seiya, Takayuki Wakisaka, and Kazuhiro Yamamoto, "Estimation of Monte-Carlo Scale Factors for the High-p_T Like-Sign Dilepton Analysis to Search for the Wh Production", CDF9238 (Version 2.0).

[6] Toru Okusawa, Yoshihiro Seiya, Takayuki Wakisaka, and Kazuhiro Yamamoto, "Background Estimation for the High-p_T Like-Sign Dilepton Analysis", CDF7261 (Version 2.0).