DØNote 6266
Version 1.1
February 7, 2012

Measurement of the electron identification efficiency for the DØ RunIIb W mass analysis

Hengne Li, Jan Stark

LPSC, 53 Rue Des Martyrs
38026 Grenoble Cedex, France

Contents

1 Introduction **3**

2 Efficiency measurement from $Z \to ee$ collider data **3**

2.1 The “tag-and-probe” bias 4

2.2 An alternative approach 6

2.3 Those electrons fail a cut have a different energy scale 12

2.4 The resulting H-Matrix efficiency 14

3 Efficiency ratio between data and full MC **16**

3.1 Efficiency dependence on $p_T(e)$ 16

3.2 Efficiency dependence on SET and InstLumi 17

3.3 The efficiency correction 24

A Selection criteria **30**

A.1 Major regions excluded containing dead cells 30

A.2 Selection criteria for H-Matrix efficiency study 31

A.3 Selection criteria for loose-track-match efficiency study 32

A.4 Selection criteria for tight-track-match efficiency study 33

A.5 Pure background selection criteria for H-Matrix efficiency study 34

A.6 Pure background selection criteria for loose-track-match efficiency study . 34

A.7 Pure background selection criteria for tight-track-match efficiency study . 35

1 Introduction

The DØ W mass analysis uses a dedicated fast MC to describe the collider data. In the fast MC, the modeling of electron identification efficiency is developed in two steps. In the first step, we take the full `GEANT` MC as reference, tune the efficiency model in the fast MC so as to make it agree with the full MC. In the second step, we derive the relative efficiency between collider data and full MC, apply it as a correction on the efficiency model in the fast MC to describe the collider data.

The "two–step" modeling is demanded by the complexity of efficiency modeling in the Run IIb high instantaneous luminosity condition. In the first step, the dependences on electron p_T, scalar E_T (SET) and instantaneous luminosity (InstLumi), and electron η_{det} and η_{phys} and their correlations can be tuned based on MC truth, as detailed in Section 8 of Ref. [1]. These tuning methods can not be repeated in the collider data because we do not know the truth information and we do not have enough events for the detailed study and parameterization. In the second step, since we already managed to construct the fast MC that models the full MC well, the remaining task is to check if there is any difference of the efficiency dependence between data and full MC. In case there is any difference, we can correct the efficiency measured in full MC by the ratio between data and full MC.

This note presents the second step of our efficiency modeling for collider data analysis. In this study, we derive the efficiencies of track-matching and H-Matrix criteria as functions of electron p_T ($p_T(e)$), SET and Inst.Lumi, separately from collider data and full MC[1]. A new method to extract the electron identification efficiency from data $Z \to ee$ events is presented in Section 2 taking $p_T(e)$ dependence of H-Matrix efficiency as an example. Based on the same method, efficiencies of other selection criteria and their dependences are presented and discussed in Section 3, as well as the ratios between data and full MC. All the selection criteria used in this efficiency study are defined in Appendix A.

2 Efficiency measurement from $Z \to ee$ collider data

To measure the electron identification efficiency of a particular electron selection criterion from collider data, one needs to identify how many electrons pass this criterion and how many fail. However, the collider data are not pure signal samples, they are contaminated by background. Any background contamination biases the efficiency measurement. In principle, one does not know how many of these electrons that pass (fail) a selection criterion are actual electrons. It is especially true for electrons that fail a selection

[1]The study focusses on track-matching and H-Matrix selection criteria because they are the two criteria that sensitive to the underlying energy contamination.

criterion, because the selection criterion itself is designed to select the actual electrons. Therefore, the major difficulty in the efficiency measurement from collider data is the background suppression.

2.1 The "tag-and-probe" bias

Traditionally, the "tag-and-probe" method to derive the electron identification efficiency from $Z \to ee$ events of collider data is extensively used to suppress the background. The idea of "tag-and-probe" is that, we "tag" one electron using relatively tight criteria to identify the $Z \to ee$ events (thus, to suppress the background), and "probe" the other electron from the same Z decay to study the efficiency. However, the "tag-and-probe" method may introduce certain bias in the measured efficiency due to the fact that the two electrons from the same Z decay are kinematically strongly correlated.

Figure 1 shows a study of the "tag-and-probe" bias using full MC $Z \to ee$ events based on MC truth. In this study, the efficiency as a function of true $p_T(e)$ is extracted (red histogram), taking the number of MC true electrons as the denominator and the number of electrons passing all the electron selection criteria used in the W mass analysis (see Section 4.6 of Ref. [1]) as the numerator. Now, we do the same exercise to extract the efficiency but require the other electron from the same Z decay to pass all the electron selection criteria (blue histogram). The latter (blue histogram), equivalents to the efficiency derived using the "tag-and-probe" method. We can see from Figure 1 that the blue histogram is different from the red one, clearly a bias due to the "tag-and-probe" method.

The "tag-and-probe" bias shown in Figure 1 is that for electrons with true $p_T(e)$ below the jacobian edge (at 45 GeV) the "tag-and-probe" efficiency (blue) appears to be smaller than the actual efficiency (red), while above the jacobian edge the two efficiencies agree well. To understand this bias, one could start from the kinematics of a $Z \to ee$ event. Imagine an electron from a Z decay gains a high p_T above the jacobian edge. The most probable scenario is that the electron is flying along the Z boson boost direction. Therefore, the other electron must have a small p_T below the jacobian edge, and fly opposite to the Z boost direction, where there is more hadronic recoil contamination. At the same time, we also understand that high p_T electrons can be easier identified because they are less affected by energy contamination from underlying energy (hard recoil, spectator parton interactions, and additional $p\bar{p}$ interactions). Therefore, a "tag" electron is more likely to have a high p_T flying along the Z boost, while the "probe" electron is more likely to have a low p_T flying opposite to the Z boost thus a region with heavier recoil contamination resulting in a lower efficiency.

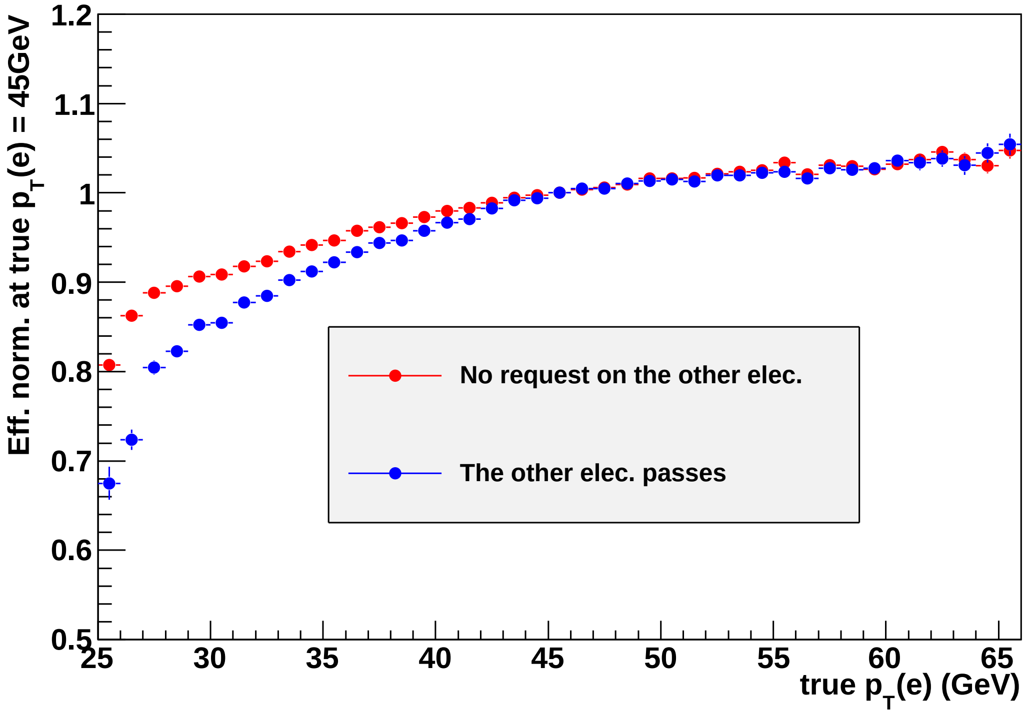

Figure 1: Electron identification efficiency extracted from full MC $Z \to ee$ events using truth information, without (red) and with (blue) requesting a “tag” electron.

2.2 An alternative approach

Due to the reason discussed above, an alternative approach to extract the efficiency from collider data without any requirement of a "tag" electron is developed in the Run IIb W mass analysis. In this approach, we make use of the shape of the invariant mass ($M(ee)$) distribution of the $Z \to ee$ events to subtract the background. We fit a function composed of a signal function and a background function to it to obtain an estimate of the background contribution. We present the method in this section taking the H-Matrix efficiency as a function of $p_T(e)$ as an example. The efficiencies of other selection criteria as functions of other observables following the same method are presented in Section 3. The definition of the selection criteria are left to Appendix A.

To extract the H-Matrix efficiency as a function of $p_T(e)$, for each of $p_T(e)$ bin we fill a histogram of $M(ee)$ with one entry per electron that passes the H-Matrix criterion and another histogram for the electrons that fail. An important point should be noticed is that, each entry of the two histograms of a particular $p_T(e)$ bin represents an electron (not an event). The electron under study is required to be in a given $p_T(e)$ bin and in CC region, while the other electron in the same $Z \to ee$ event can have any $p_T(e)$ above 25 GeV, and can be either in CC or EC. The usage of the other electron is limited to the calculation of the $M(ee)$ together with the electron under study. The precise selection criteria for the H-Matrix efficiency study are given in Appendix A.2.

Figure 2 and Figure 3 are the $M(ee)$ distributions for electrons that pass and fail the H-Matrix criterion, respectively, taking the $p_T(e)$ bin between 39 and 41 GeV as an example. Certain background contamination can be observed in the two distributions.

The next step is to fit a function composed of a signal function and a background function to the data $M(ee)$ distribution to subtract the background. The composed function shown in Figures 2 and 3 is given as

$$F(x; N, f) = N \cdot [f \cdot S(x) + (1 - f) \cdot B(x)] \tag{1}$$

where, N is the total number of entries in the data $M(ee)$ histogram, f ($f \in (0, 1)$) is the fraction of pure signal among N, $S(x)$ and $B(x)$ are the normalized signal and background functions, respectively.

The starting point to construct the signal function ($S(x)$) is our fast MC (`RESBOS` + `PHOTOS` + `wz_epmcs`) $M(ee)$ spectrum. It is obtained following the same electron selection criteria as the collider data $M(ee)$ distribution described above. However, in this study the other electron can be also in EC, but our fast MC is not yet tuned to describe the EC electrons to the level of precision as the CC electrons. At the same time, our fast MC is designed to describe correctly only those electrons that succeed the electron selection criteria, not those fail (see discussions in Section 2.3). Therefore, we need a method to introduce a certain flexibility to the fast MC $M(ee)$ spectrum so that it can describe the signal from the $Z \to ee$ data correctly.

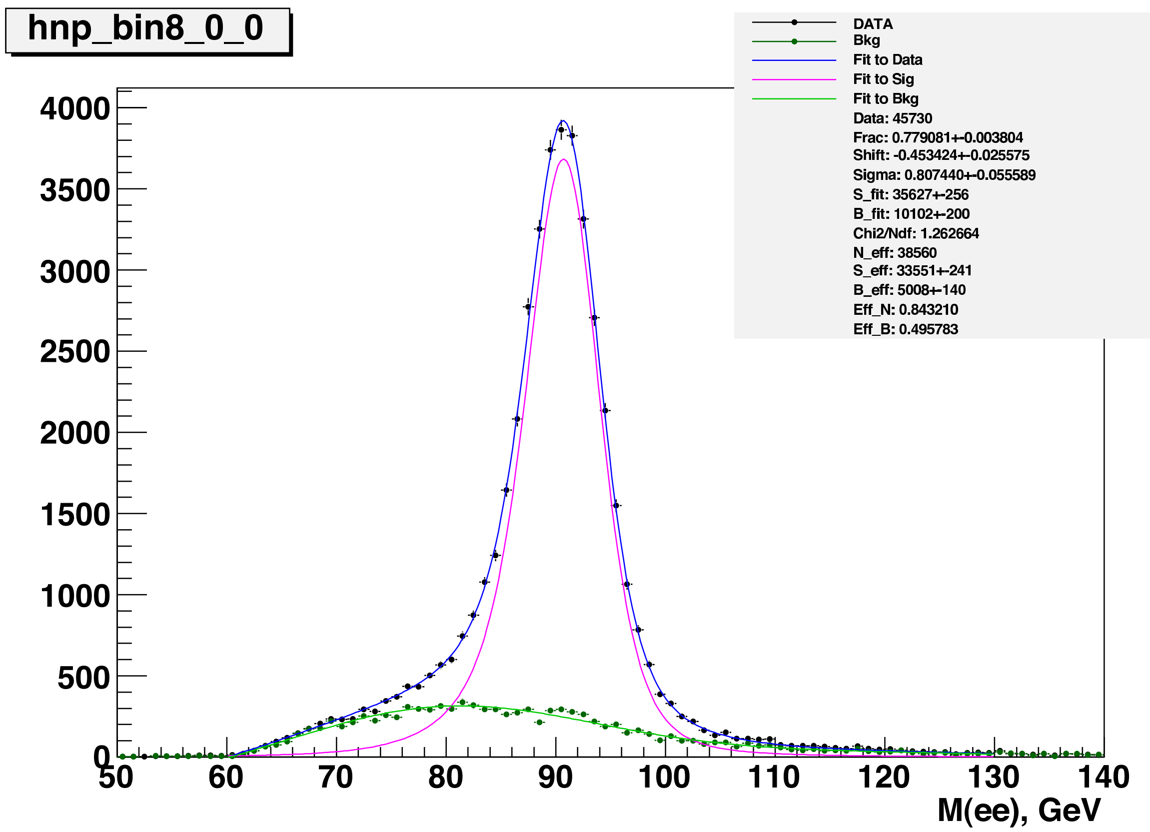

Figure 2: Fit to the $M(ee)$ distribution to subtract the background for electrons with $p_T(e)$ between 39 and 41 GeV that *pass* the H-Matrix criterion.

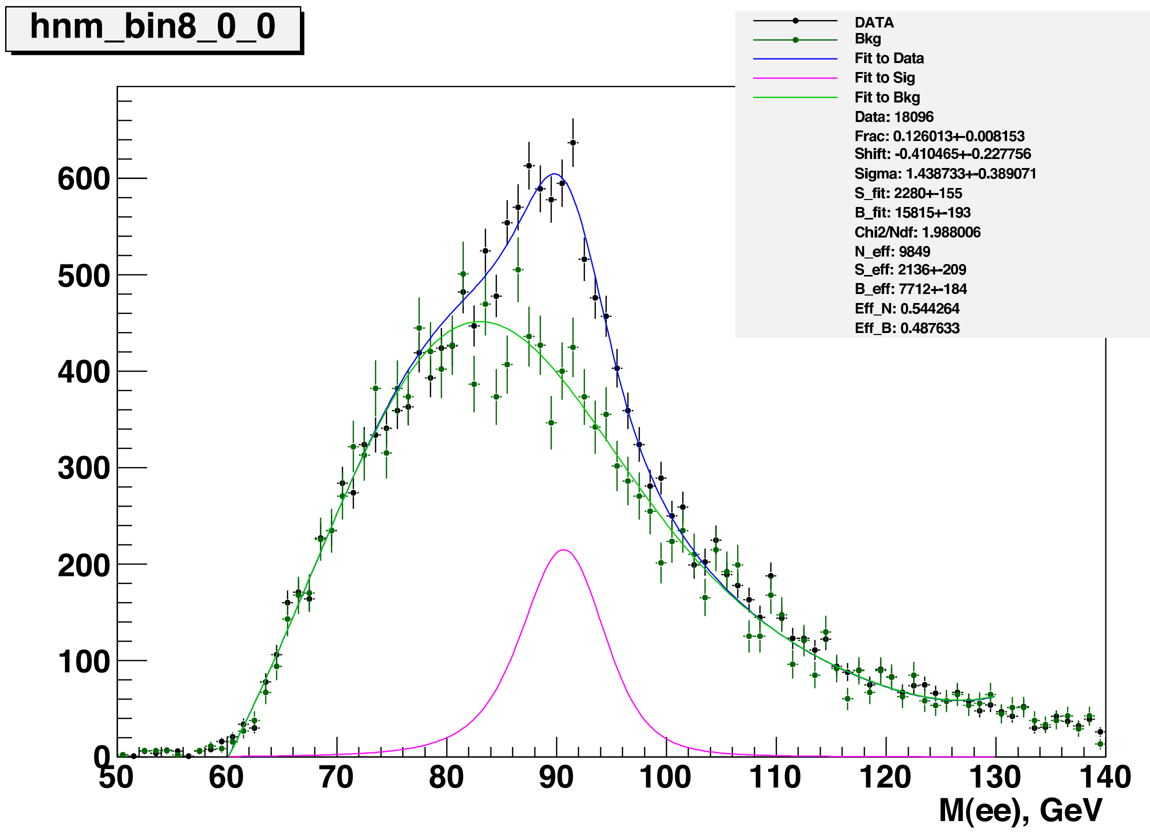

Figure 3: Fit to the $M(ee)$ distribution to subtract the background for electrons with $p_T(e)$ between 39 and 41 GeV that *do not pass* the H-Matrix criterion.

For this purpose, we need to introduce at least two free parameters to the fast MC M(ee) spectrum: a $M(ee)$ peak shift and an additional smearing factor. The Kernel Density Estimation method satisfies our need. However, the existing packages of Kernel Estimation (e.g. `RooFit`) are all based on un-binned data samples in a non-parametric way, which is inconvenient for the problem at hand. A modified version of the method is developed in Ref. [2], which allows the Kernel Estimation function to be initialized using a histogram.

In this way, one can see the two free parameters (the $M(ee)$ peak shift and the additional smearing factor) are naturally introduced in the modified Kernel Estimation parameterization below:

$$S(x; s, \sigma) = \frac{1}{N} \sum_{i=1}^{m} \frac{n_i}{\sqrt{2\pi (h_i + \sigma)^2}} e^{-\frac{[(x-s)-t_i]^2}{2(h_i+\sigma)^2}} \tag{2}$$

where, N is the number of entries of the initial histogram, it is divided as in the first term on the right hand side so that the function $S(x)$ is normalized with its area equals *one*; s is the $M(ee)$ peak shift, σ is the additional smearing factor, they are the two free parameters introduced; m is the number of bins in the initial $M(ee)$ histogram generated by our fast MC; n_i is the bin content of the ith bin, t_i is the center of the ith bin; and h_i is the bandwidth of the ith bin, which is chosen to be

$$h_i = \frac{1}{2} \cdot (\frac{4}{3})^{\frac{1}{5}} \cdot \sqrt{\frac{\Delta x_i}{3\sqrt{2}}} \cdot \sqrt{\frac{N}{n_i}} \tag{3}$$

where the Δx_i is the bin width of the ith bin.

We can see that except for the two free parameters (s and σ) all the other parameters in signal function $S(x; s, \sigma)$ (Equation 2) are given by a histogram. An $M(ee)$ histogram from fast MC is used to initialize a signal function $S(x; s, \sigma)$. The free parameters s and σ are used to provide certain adjustment of the function. An example is shown in Figure 4 for $p_T(e)$ between 35 to 37 GeV.

To describe the background, we firstly get the background $M(ee)$ histogram for each $p_T(e)$ bin from a sample of candidate di-electron events from collider data (keeping in mind that this data sample before strong electron identification is dominated by background) by inverting other electron selection criteria (see Appendix A.5 for the exact definition). In order to be convinced that the pure background spectrum does not contain any $Z \to ee$ signal, we increase the inverse level of the electron selection criteria, until the shape (not the normalization) of the background spectrum does not change anymore.

We are only interested in the shape (the probability density function (PDF), not the normalization) of the background spectrum in each $p_T(e)$ bin. The normalization (number of background) is determined later by the fit of the composed function (Equation 1) to

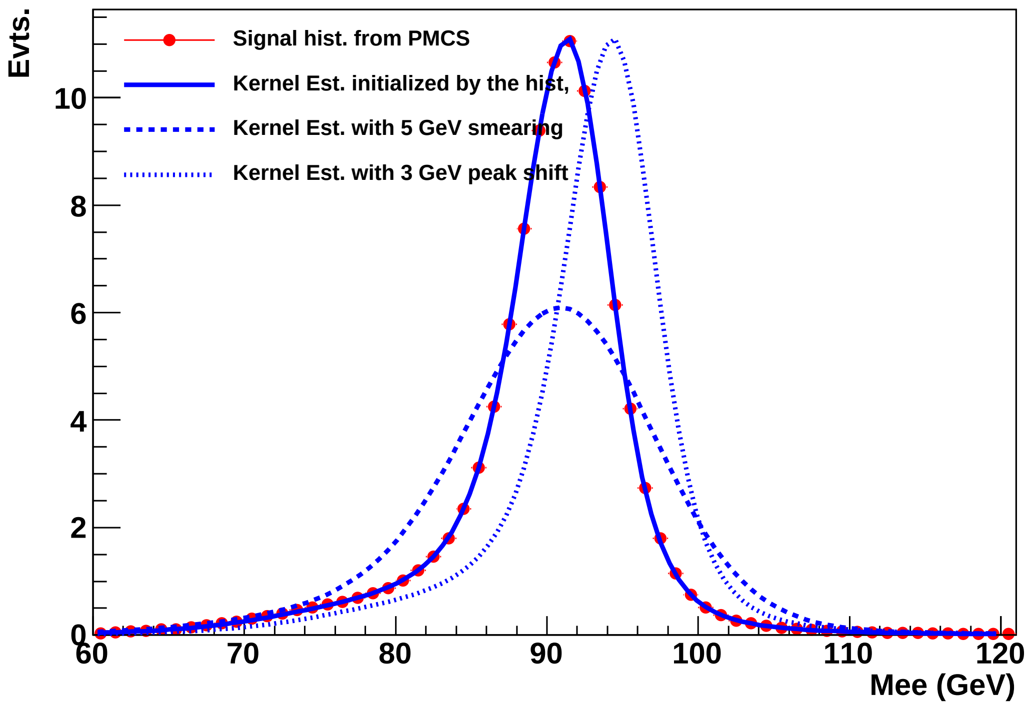

Figure 4: Example of using the modified Kernel Estimation function to describe the fast MC $M(ee)$ distribution with introduced flexibility. The dotted red histogram is generated by fast MC for $p_T(e)$ from 35 to 37 GeV, the solid blue curve is the signal function (Formula 2) just initialized by the dotted red histogram using. The two other blue curves are same as the solid blue function but with 5 GeV additional Gaussian smearing ($\sigma = 5$ GeV), and with 3 GeV shift of the mass peak ($s = 3$ GeV), respectively.

the data $M(ee)$ histogram. We describe the background shape for each electron $p_T(e)$ bin using an *ad hoc* function:

$$B(x) = p_0 \cdot e^{-\frac{(x-p_1)^2}{2p_2^2}} + \sum_{i=3}^{13} p_i \cdot x^{i-3} \tag{4}$$

where p_0 to p_{13} are the parameters. This *ad hoc* background function is fit to the pre-selected pure background histogram to determined the parameters. Thereafter, we scale the background function so that its area is normalized to be *one* and fix all the parameters.

The fit to extract the signal as shown in Figure 2 and 3 is performed in the following steps:

- Initialize the signal function $S(x)$ (Equation 2) using the fast MC $M(ee)$ histogram;
- Fit the background function $B(x)$ (Equation 4) to the background $M(ee)$ histogram from the $Z \to ee$ data selected by inverting the signal selection criteria;
- Fix all the parameters determined in the first two steps except f, s and σ, and fix the normalization N (the number of entries in the data $M(ee)$ histogram) in the composed function (Equation 1). Fit the composed function to the data $M(ee)$ distribution to determine parameters s, σ and f.

The efficiency and its error are defined as

$$\epsilon = \frac{n^+}{n^+ + n^-} , \tag{5}$$

$$\sigma_\epsilon^2 = \frac{(n^-)^2 \cdot \sigma_{n^+}^2 + (n^+)^2 \cdot \sigma_{n^-}^2}{(n^+ + n^-)^4} . \tag{6}$$

Where, n^+ and n^- are the number of electrons that pass and fail the H-Matrix criterion, respectively. They are determined from the fits to the $M(ee)$ histograms separately for electrons pass and fail the selection criterion.

Note that, we do all the fits in the $M(ee)$ window from 60 to 130 GeV, but we calculate the efficiency in the $M(ee)$ window from 80 to 100 GeV to increase the signal over background ratio. Therefore, n^+ or n^- is calculated as the number of data inside the efficiency calculation window minus the number of background inside this window,

$$n = N_{\text{eff}} - k_{\text{eff}} \cdot (1 - f) \cdot N_{\text{fit}} . \tag{7}$$

Where, n refers to either n^+ or n^-; N_{eff} and N_{fit} are the numbers of entries in the data $M(ee)$ histogram inside the efficiency calculation window (80–100 GeV) and the fitting window (60–130 GeV), respectively; f is the signal fraction determined by the fit as in Equation 1; k_{eff} is the fraction of background falls in the efficiency calculation window with respect to the fitting window, which is determined by the pre-selected pure background histogram used to initialize the background function (Equation 4). k_{eff} together with its error is calculated as

$$k_{\text{eff}} = \frac{N_{\text{eff}}^{\text{bkg}}}{N_{\text{fit}}^{\text{bkg}}} \ , \tag{8}$$

$$\sigma_{k_{\text{eff}}}^2 = \frac{(N_{\text{fit}}^{\text{bkg}})^2 \cdot N_{\text{eff}}^{\text{bkg}} + (N_{\text{eff}}^{\text{bkg}})^2 \cdot N_{\text{fit}}^{\text{bkg}}}{(N_{\text{fit}}^{\text{bkg}})^4} = \frac{k_{\text{eff}}}{N_{\text{fit}}^{\text{bkg}}} \cdot (1 + k_{\text{eff}}) \ , \tag{9}$$

where $N_{\text{eff}}^{\text{bkg}}$ or $N_{\text{fit}}^{\text{bkg}}$ is the number of background in the pure background histogram

Therefore, the error of n^+ or n^- is given by:

$$\sigma_n^2 = N_{\text{eff}} + k_{\text{eff}}^2 \cdot (1-f)^2 \cdot N_{\text{fit}} + k_{\text{eff}}^2 \cdot N_{\text{fit}}^2 \cdot \sigma_f^2 + (1-f)^2 \cdot N_{\text{fit}}^2 \cdot \sigma_{k_{\text{eff}}}^2 . \tag{10}$$

2.3 Those electrons fail a cut have a different energy scale

The fitting procedure introduced above is done for each (reconstructed) electron $p_T(e)$ bin, and the efficiency is calculated from n^+ and n^- (Equation 5) separately for each $p_T(e)$ bin. We notice that the energy responses of electrons that pass and fail a selection criterion do not agree with each other. This is a dangerous effect that can bias the measured efficiency as a function of reconstructed $p_T(e)$. For the same true electron p_T, a different energy response of electrons that fail a selection criterion may migrate those electrons (fail the criterion) from one reconstructed $p_T(e)$ bin to another. Therefore, the measured n^- of a given reconstructed $p_T(e)$ bin is biased, thus, the efficiency is biased. The existence of the different energy response between electrons passing and failing the selection criteria is possible, because all the electron energy scale studies and calibrations at DØ are using electrons that well selected. No dedicated study of the energy scale of electrons that fail the selection criteria has been performed so far.

One appearance of the different energy scale of electrons that fail a selection criterion is that it creates a separated bump in the $M(ee)$ distribution below Z peak. Using the full MC $Z \to ee$ events with truth information we can clearly see this effect. Figure 5 shows the full MC $M(ee)$ distribution of electrons with reconstructed $25 < p_T(e) < 30$ GeV that fail the H-Matrix criterion (red histogram). Recall that the full MC $Z \to ee$ events are pure signal, but we can see clearly there is a separated bump below the one at the Z boson mass. Now, we draw the same $M(ee)$ distribution, but in addition to $25 < p_T(e) < 30$ GeV we require the true $p_T^{\text{true}}(e) > 30$ GeV (blue histogram). We can see the bump below the Z peak is caused by the electrons with a very small energy response.

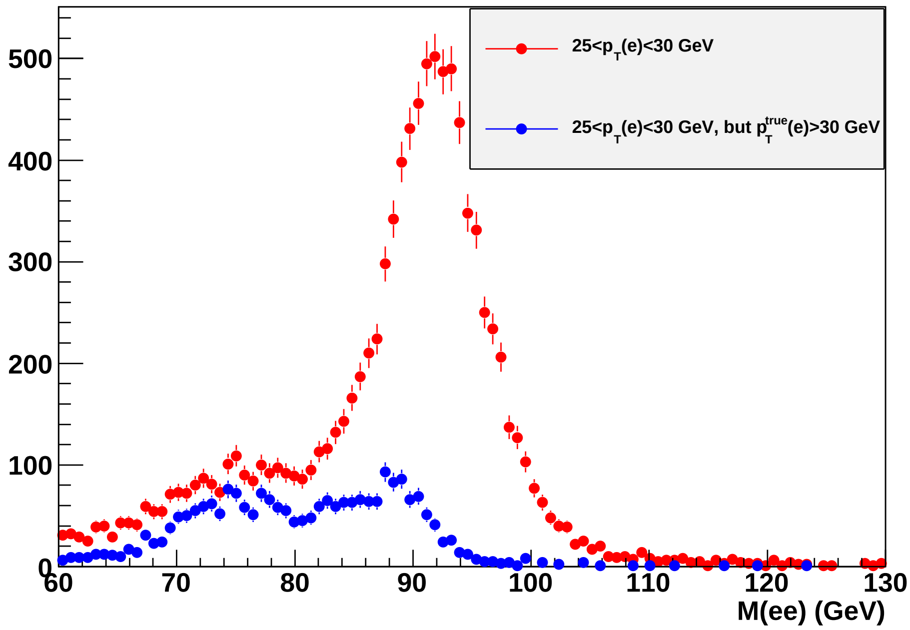

Figure 5: Full MC $M(ee)$ distribution of electrons fail the H-Matrix criterion with reconstructed $25 < p_T(e) < 30$ GeV (red). And the one with additional request on the true $p_T^{\mathrm{true}}(e) > 30$ GeV (blue).

We find that this very small energy response is coming from electrons that fall in calorimeter regions with dead cells. The energy response, together with the cluster shape, of these electrons is just not complete. For electrons that fall into these regions with dead cells it is therefore almost impossible to survive the H-Matrix and track-matching criteria.

One may note that this effect does not only bias the efficiencies depending on $p_T(e)$ due to the migration effect between $p_T(e)$ bins, it also causes the fit to the $M(ee)$ distribution to fail to converge. In order to remove the bias due to dead cells in the measurement of the efficiencies, we exclude electrons that fall in the major calorimeter regions with dead cells. The complete exclusion map of the calorimeter dead cells for both full MC and data is precisely defined in Appendix A.1.

Another appearance of the different energy scale of electrons that fail a selection criterion is not as apparent as the first one, but causes a small difference of the $M(ee)$ peak with respect to that of the electrons that pass the selection criterion. Recall that in signal function (Equation 2) we have a free parameter s to describe the relative shift of the measured $M(ee)$ peak to that of the fast MC. So, the $M(ee)$ peak shift between electrons failing and passing a selection criterion can be measured from our fit to the data $M(ee)$ histograms for each $p_T(e)$ bin.

For the H-Matrix efficiency, the measured relative $M(ee)$ peak shift ($s_{\rm rel}$) between electrons failing and passing the H-Matrix criterion as a function of $p_T(e)$ is shown in the plot on the top side of Figure 6. We can see $s_{\rm rel}$ is reaching a size of about 1 GeV. Then we convert the relative peak shift to an energy scale factor (α) using the following formula for each $p_T(e)$ bin to correct the energy of electrons that fail H-Matrix criterion to be the same as that of electrons pass:

$$\alpha = \frac{M_{\rm pass}}{M_{\rm pass} + s_{\rm rel}} \ , \tag{11}$$

where, $M_{\rm pass}$ is the $M(ee)$ peak position of electrons pass the H-Matrix criterion. The resulting energy scale (α) as a function of $p_T(e)$ is shown in the plot on the bottom side of Figure 6. We use a *ad hoc* curve to smoothen the energy scale function.

2.4 The resulting H-Matrix efficiency

After the exclusion of the calorimeter regions containing dead cells and the energy scale correction for electrons that fail the H-Matrix criterion, the H-Matrix efficiency as a function of $p_T(e)$ extracted from the $Z \to ee$ data (red) is shown in Figure 7. Also shown in this plot is the H-Matrix efficiency derived from full MC $Z \to ee$ events (blue), and the ratio between the efficiencies from data and full MC (data divided by full MC) (black).

The efficiency from full MC is directly extracted from the electron $p_T(e)$ distributions that pass and fail the H-Matrix criterion using Equation 5. We apply an $M(ee)$ cut ($80 <$

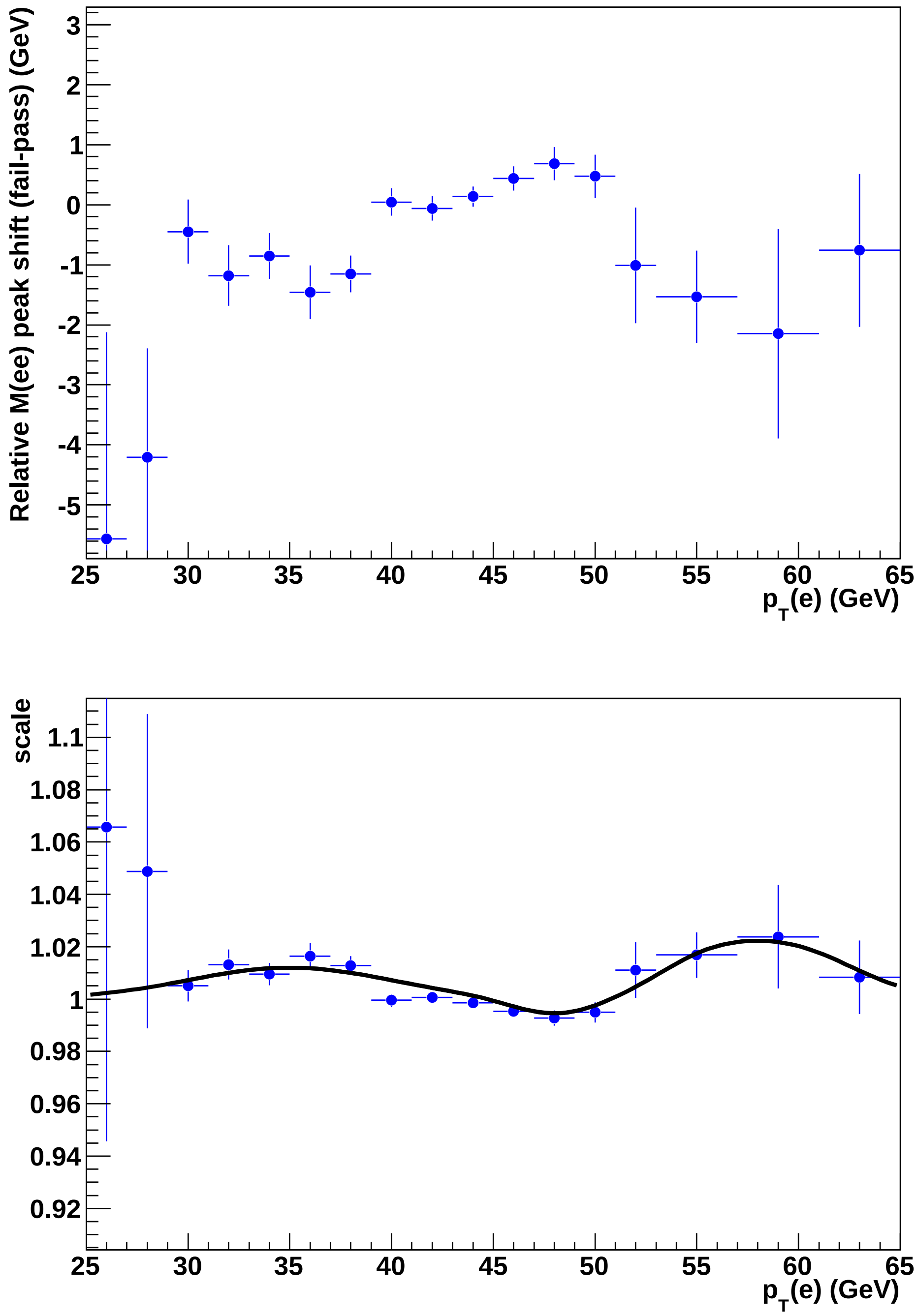

Figure 6: Top: Relative $M(ee)$ peak shift between electrons fail and pass the H-Matrix criterion ($s_{\rm rel}$) as a function of $p_T(e)$. Bottom: The energy scale factor (α) to correct the energy of electrons fail the H-Matrix criterion to be that of electrons pass, as a function of $p_T(e)$. This scale factor is converted from $s_{\rm rel}$ using Equation eq:escale.

$M(ee) < 100$ GeV) equivalent to the efficiency calculation $M(ee)$ window that is used in efficiency measurement from collider data. Those colorimeter regions containing dead cells that are simulated in the full MC (see Appendix A.1) are also excluded. The energy scale correction for electrons that fail the H-Matrix criterion is determined separately for full MC and applied in the efficiency measurement.

From the ratio of efficiencies between data and full MC as shown in Figure 7, we do not see any H-Matrix efficiency dependence on $p_T(e)$ need to be corrected for the W mass measurement using collider data.

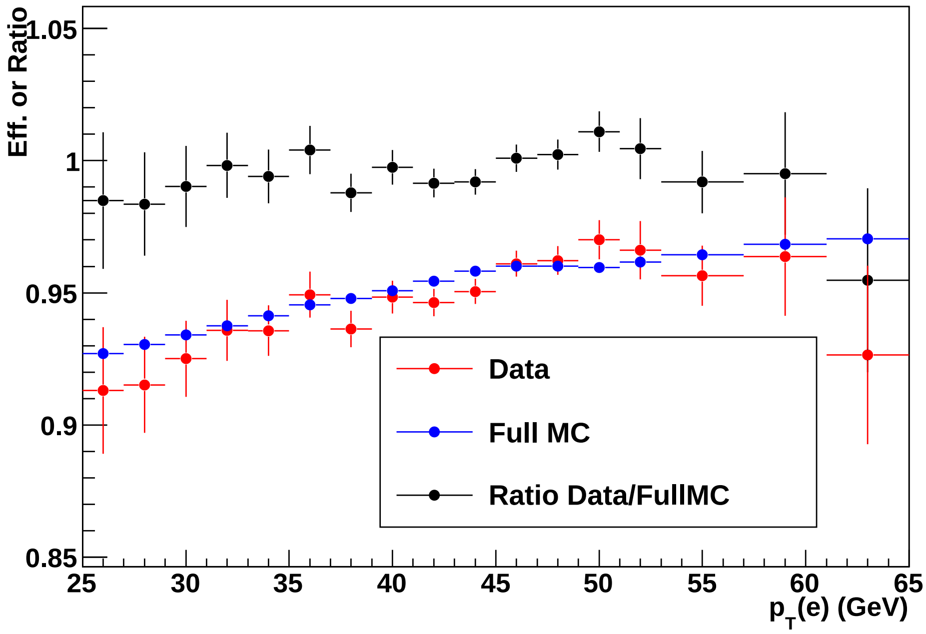

Figure 7: H-Matrix efficiency derived from $Z \to ee$ data and full MC as a function of electron $p_T(e)$, and the ratio between them (data over full MC).

3 Efficiency ratio between data and full MC

3.1 Efficiency dependence on $p_T(e)$

The H-Matrix efficiency as a function of $p_T(e)$ is shown in Section 2 as an example to present our method the extract the efficiency from collider data. Following the same method, in this section, we derive the loose-track-match efficiency and tight-track-match efficiency as a function of electron $p_T(e)$. The selection criteria for these efficiency studies are given in Appendix A.

The extracted loose-track-match efficiency and tight-track-match efficiency as a function of $p_T(e)$ are shown in Figure 8 and 9, respectively. Also on these two plots are the corresponding efficiencies derived from full MC $Z \to ee$ samples, and the ratio between data and full MC (data/full MC). Figure 10 shows the efficiency ratio between data and full MC (data/full MC) as a function of $p_T(e)$, separately for H-Matrix, loose-track-match, tight-track-match and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria. A flat line fit to the overall efficiency ratio as a function of $p_T(e)$ gives a $\chi^2/\mathrm{Ndf} = 24.9/21$. The $p_T(e)$ dependence of the overall efficiency agrees well between data and full MC.

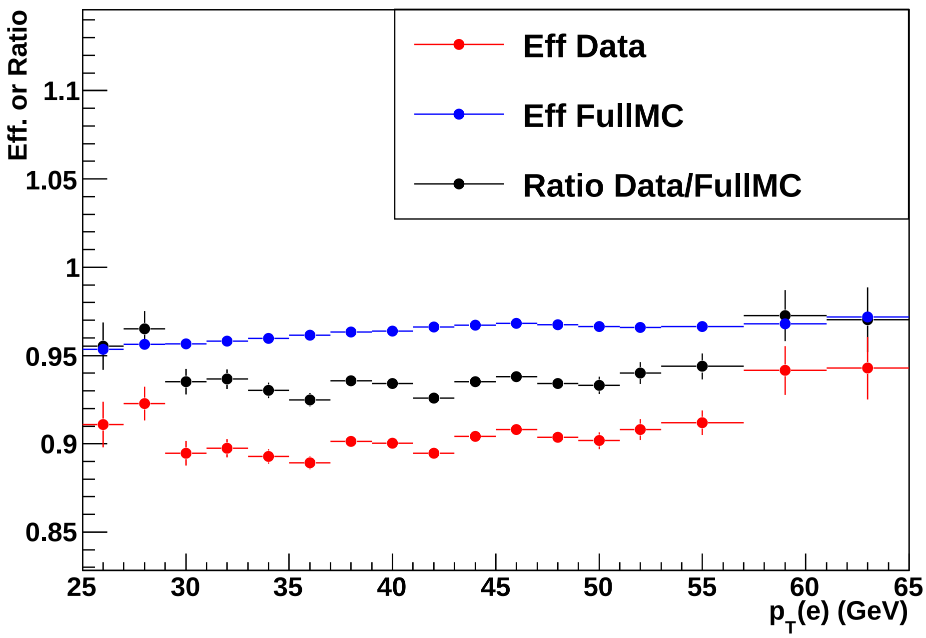

Figure 8: Loose-track-match efficiency derived from $Z \to ee$ data and full MC as a function of $p_T(e)$, and the ratio between them (data/full MC).

3.2 Efficiency dependence on SET and InstLumi

In this section, we study the efficiency dependence on SET and InstLumi. The selection criteria are defined in Section A, which are the same as in the study of $p_T(e)$ dependence. The only difference in the method to extract the efficiency is that we do not need to correct the energy response of electrons failing a selection criterion.

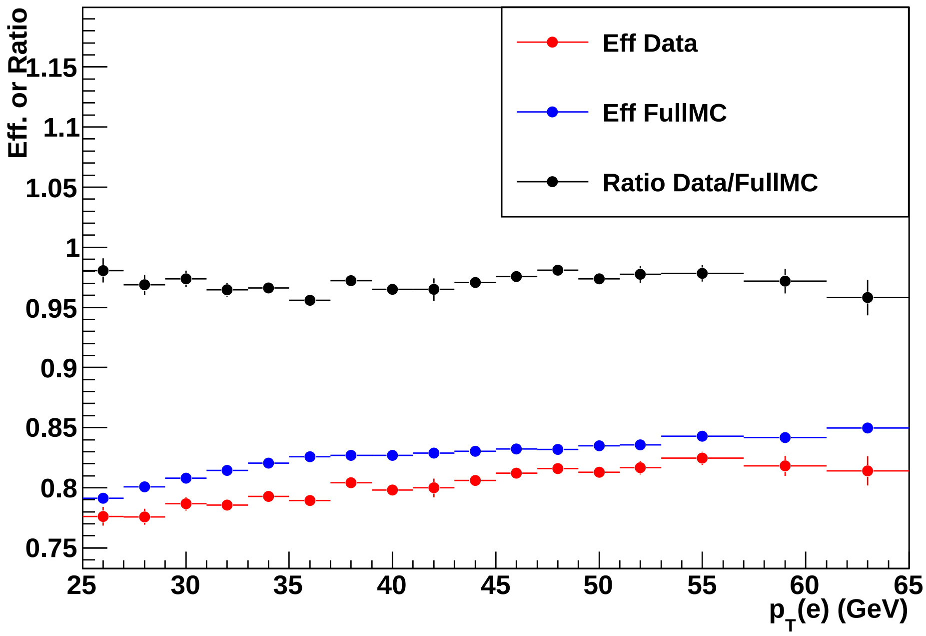

Figure 9: Tight-track-match efficiency derived from $Z \to ee$ data and full MC as a function of $p_T(e)$, and the ratio between them (data/full MC).

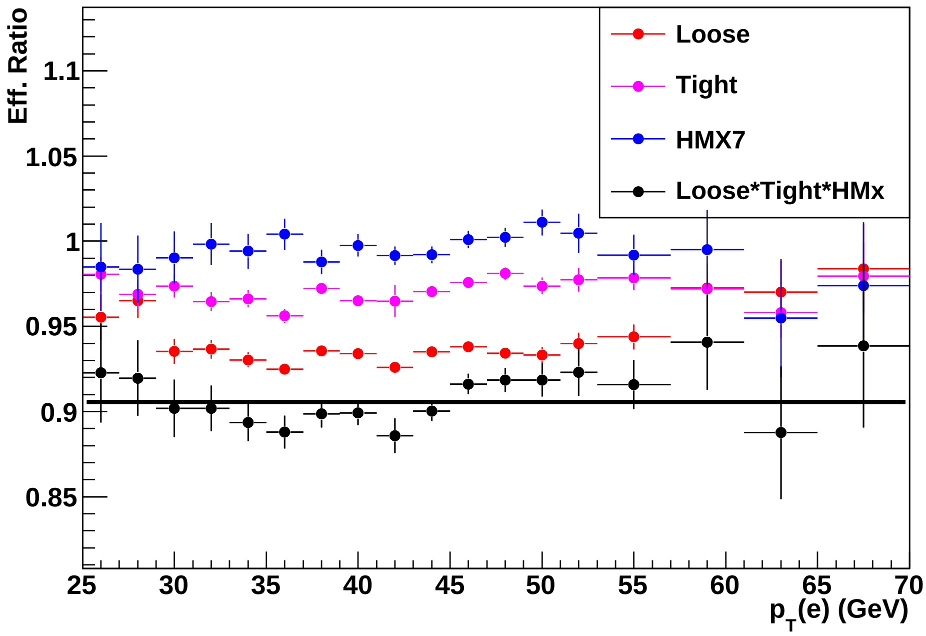

Figure 10: Efficiency ratio between data and full MC (data/full MC) as a function of $p_T(e)$ for H-Matrix, loose-track-match, tight-track-match, and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria.

The extracted efficiencies as a function of SET are shown in Figure 11, 12, and 13 for H-Matrix, loose-track-match and tight-track-match criteria, respectively. Also shown in these plots are the corresponding efficiency ratio between data and full MC. Figure 14 shows the efficiency ratio between data and full MC (data/full MC) as a function of SET, separately for H-Matrix, loose-track-match, tight-track-match and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria. The overall efficiency ratio is found to exhibit a dependence on SET, most notably at very large SET, of the order of $\sim 10\%$ for SET between 0 GeV and 200 GeV.

The extracted efficiencies as a function of InstLumi are shown in Figure 15, 16, and 17 for H-Matrix, loose-track-match and tight-track-match criteria, respectively. Also shown in these plots are the corresponding efficiency ratio between data and full MC. Figure 18 shows the efficiency ratio between data and full MC (data/full MC) as a function of InstLumi, separately for H-Matrix, loose-track-match, tight-track-match and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria. The overall efficiency ratio is also found to exhibit a dependence on InstLumi, most notably at very large InstLumi, of the order of $\sim 10\%$ for InstLumi between 0 to 6 ($36 \times 10^{30} cm^{-2} s^{-1}$).

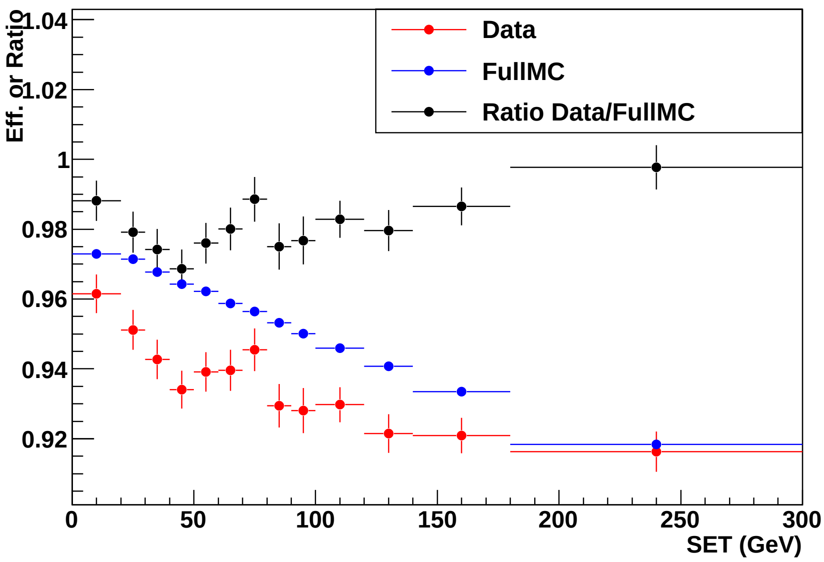

Figure 11: H-Matrix efficiency derived from $Z \to ee$ data and full MC as a function of SET, and the ratio between them (data/full MC).

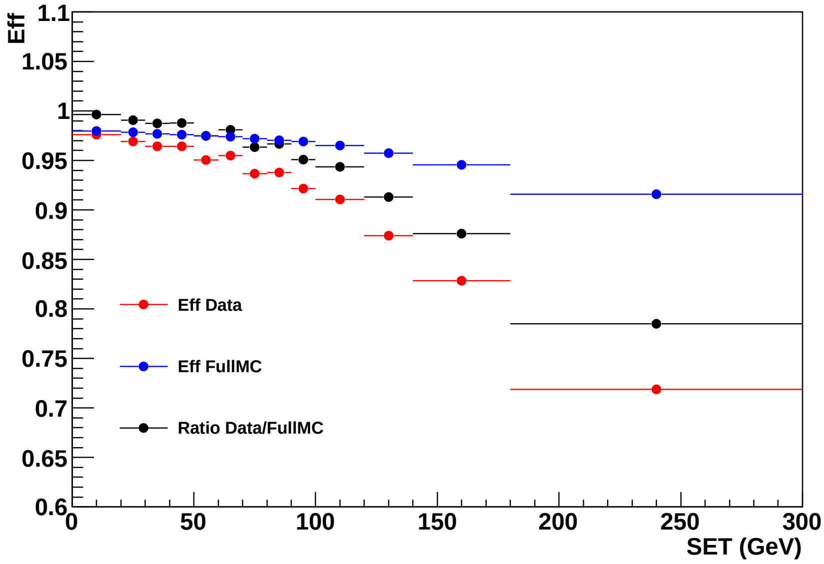

Figure 12: Loose-track-match efficiency derived from $Z \to ee$ data and full MC as a function of SET, and the ratio between them (data/full MC).

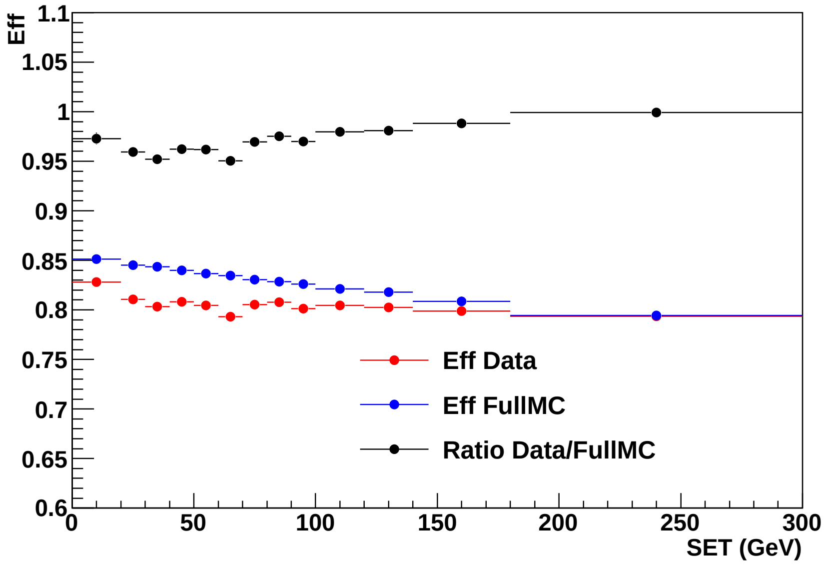

Figure 13: Tight-track-match efficiency derived from $Z \to ee$ data and full MC as a function of SET, and the ratio between them (data/full MC).

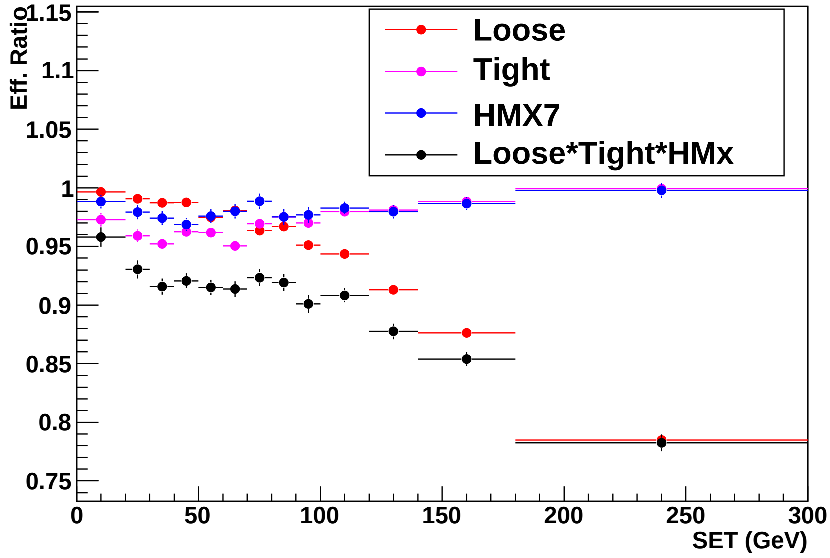

Figure 14: Efficiency ratio between data and full MC (data/full MC) as a function of SET for H-Matrix, loose-track-match, tight-track-match, and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria.

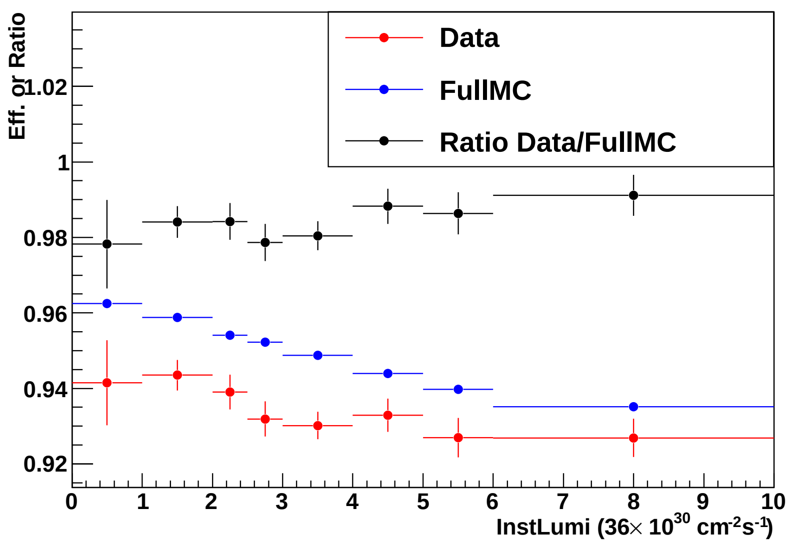

Figure 15: H-Matrix efficiency derived from $Z \to ee$ data and full MC as a function of InstLumi (in the unit of $36 \times 10^{30} cm^{-2} s^{-1}$), and the ratio between them (data/full MC).

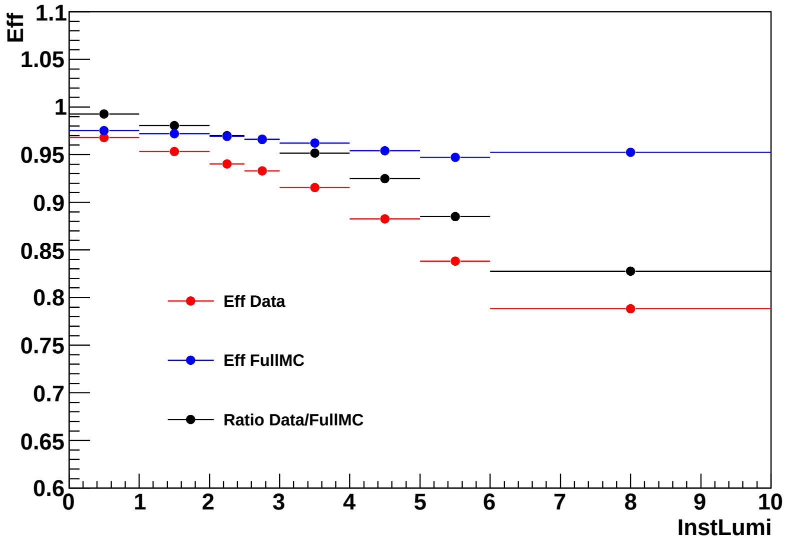

Figure 16: Loose-track-match efficiency derived from $Z \to ee$ data and full MC as a function of InstLumi (in the unit of $36 \times 10^{30} cm^{-2} s^{-1}$), and the ratio between them (data/full MC).

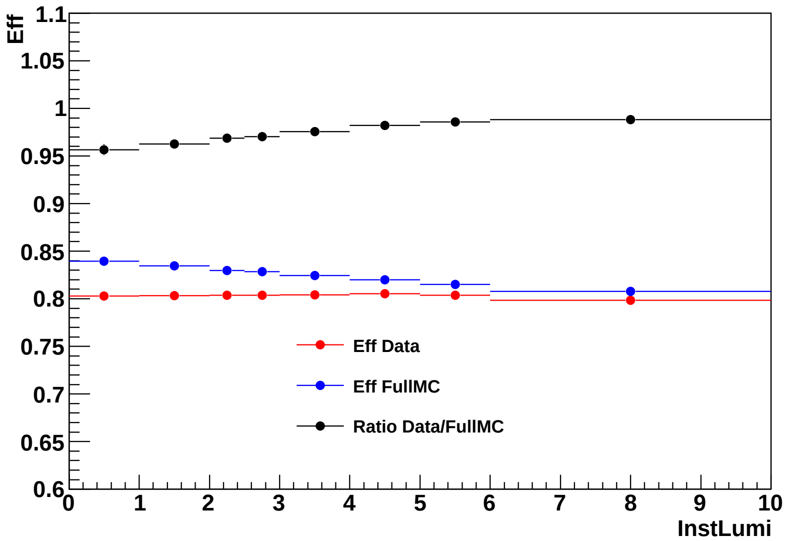

Figure 17: Tight-track-match efficiency derived from $Z \to ee$ data and full MC as a function of InstLumi (in the unit of $36 \times 10^{30} cm^{-2} s^{-1}$), and the ratio between them (data/full MC).

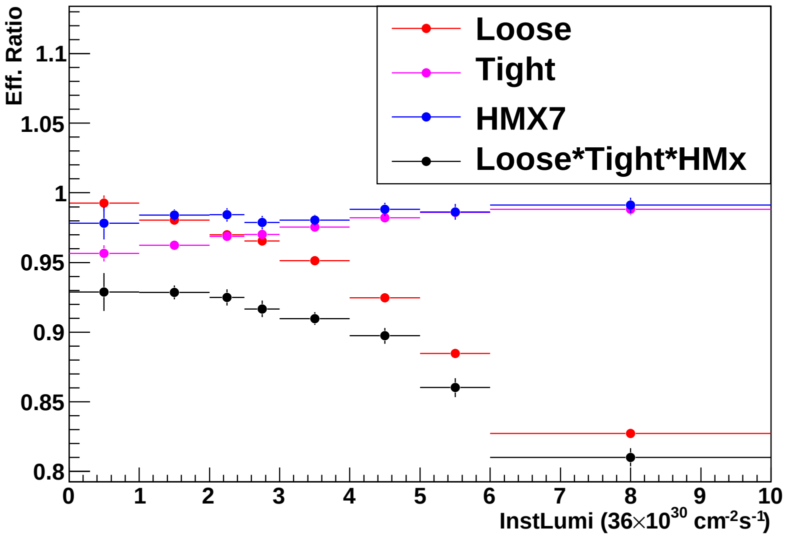

Figure 18: Efficiency ratio between data and full MC (data/full MC) as a function of InstLumi (in the unit of $36 \times 10^{30} cm^{-2} s^{-1}$) for H-Matrix, loose-track-match, tight-track-match, and the overall (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) criteria.

3.3 The efficiency correction

Based on the study above, we could conclude that the major mis-match of the efficiency ratio between data and full MC is dependent on SET and InstLumi. However, the SET and InstLumi are highly correlated. A study of the correlation between the dependences on SET and InstLumi is needed. Figure 19 shows the overall efficiency ratio (data/full MC) as a 2-dimensional function of SET and InstLumi, with the ratio value marked on each bin. Figure 20 shows the same ratio but in 3-dimensional view (black histogram) with a fit (red lined surface) using a multiplication of two polynomial functions as given by Equation 12 followed by the fitted parameters. From these two figures, one can observe the correlation between SET and InstLumi dependences.

$$R(\mathrm{SET}, \mathrm{InstLumi}) = \left[\sum_{i=0}^{3} a_i (\mathrm{SET})^i\right] \cdot \left[\sum_{i=0}^{3} b_i (\mathrm{InstLumi})^i\right] \tag{12}$$

where,

$$\begin{aligned} a_0 \ to \ a_3 \quad = \quad & 1.0 \pm 8.7 \times 10^{-3}, \quad -5.1 \times 10^{-4} \pm 2.5 \times 10^{-4}, \\ & 3.6 \times 10^{-6} \pm 2.6 \times 10^{-6}, \quad -1.7 \times 10^{-8} \pm 0.7 \times 10^{-8}. \end{aligned}$$

$$\begin{aligned} b_0 \ to \ b_3 \quad = \quad & 1.0 \pm 8.7 \times 10^{-3}, \quad 1.2 \times 10^{-2} \pm 0.7 \times 10^{-2}, \\ & -3.6 \times 10^{-3} \pm 1.8 \times 10^{-3}, \quad 1.2 \times 10^{-4} \pm 1.4 \times 10^{-4}. \end{aligned}$$

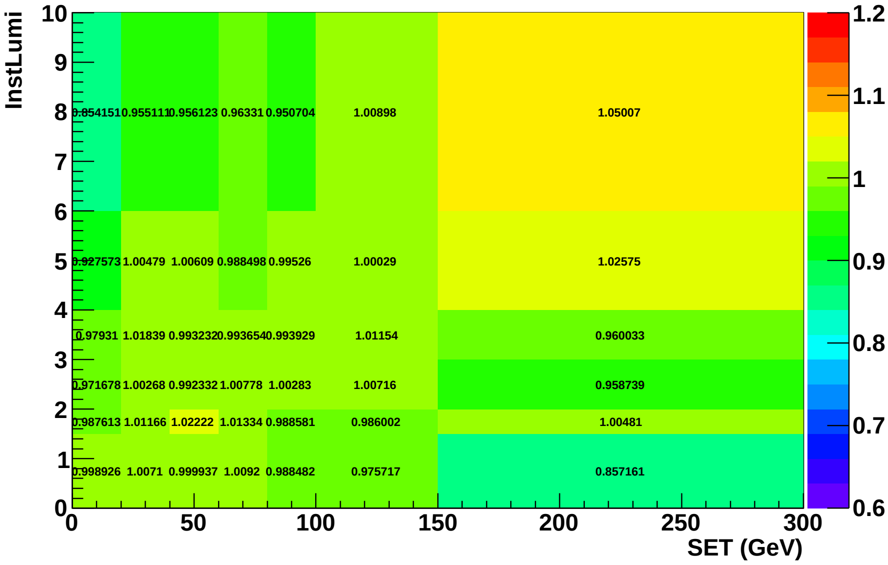

Figure 19: Overall efficiency ratio (H-Matrix $\times$ loose-track-match $\times$ tight-track-match) between data and full MC (data/full MC) as a function of SET and InstLumi. The ratio number is given on each bin.

We take the overall efficiency ratio between data and full MC (data/full MC) parameterized in Equation 12 as a function of SET and InstLumi as a correction of the efficiency model in the fast MC for the data analysis.

The plots comparing fast MC before and after the correction of the electron identification efficiency of the three observables of W$\rightarrow e\nu$ events we are using to extract the W boson mass are shown in Figure 21, 22 and 23, for W transverse mass, electron P_T and missing E_T, respectively. The number of events are scaled to 4.3 fb^{-1} RunIIb W$\rightarrow e\nu$ data. The χ distribution is shown in the bottom plot of each figures.

Observed from these comparison plots, the correction of efficiency does have a non-trivial impact on the shape of the $\not{E}_T$ distribution, and to a lesser extent also on the m_T distribution.

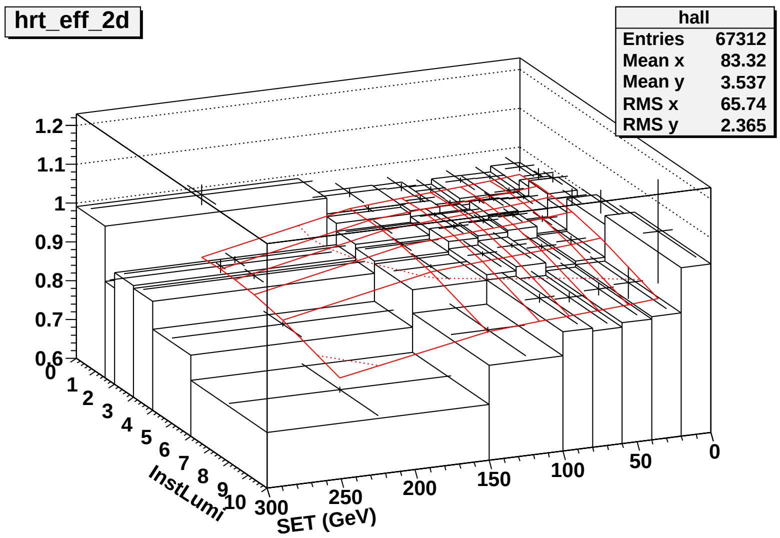

Figure 20: Overall efficiency ratio (H-Matrix × loose-track-match × tight-track-match) between data and full MC (data/full MC) as a function of SET and InstLumi in a 3-dimensional view (black histogram), and a fit to it (red lined surface).

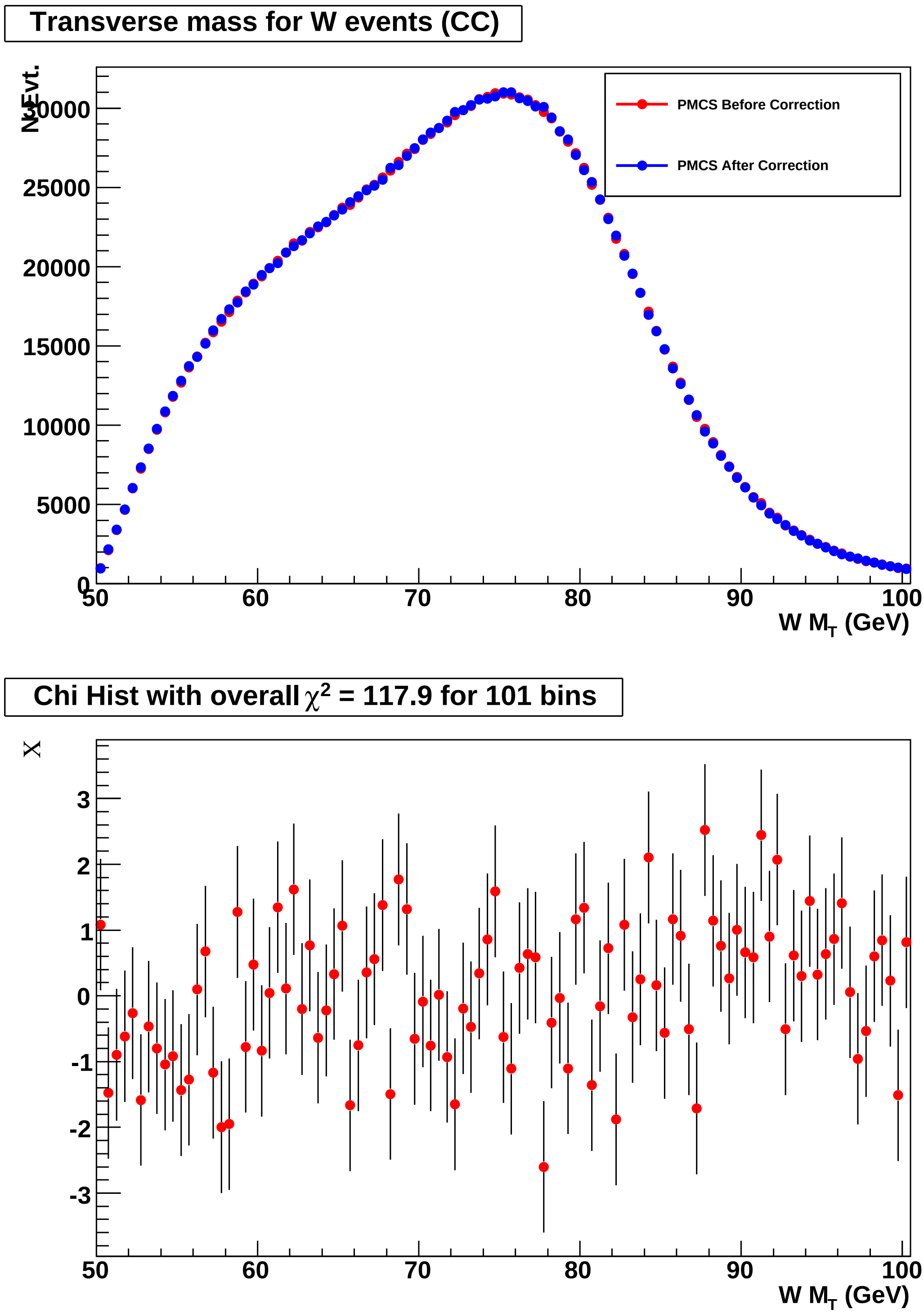

Figure 21: W transverse mass comparison of W$\rightarrow e\nu$ events between FastMC before and after the correction of efficiency. The number of events are scaled to 4.3 fb^{-1} RunIIb W$\rightarrow e\nu$ data. The bottom plot is the χ distribution of the comparison.

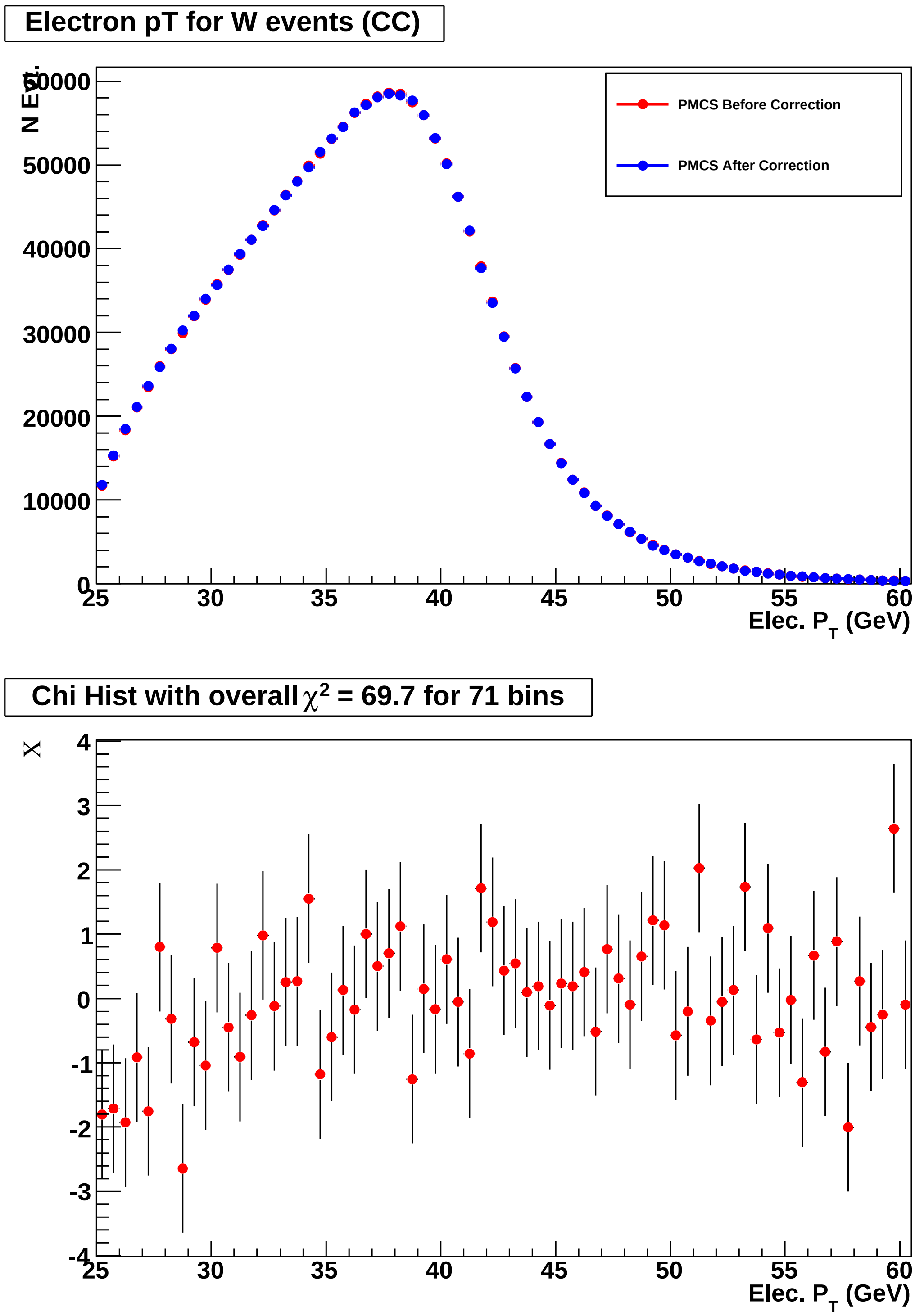

Figure 22: Electron P_T comparison of W→ $e\nu$ events between FastMC before and after the correction of efficiency. The number of events are scaled to 4.3 fb^{-1} RunIIb W→ $e\nu$ data. The bottom plot is the χ distribution of the comparison.

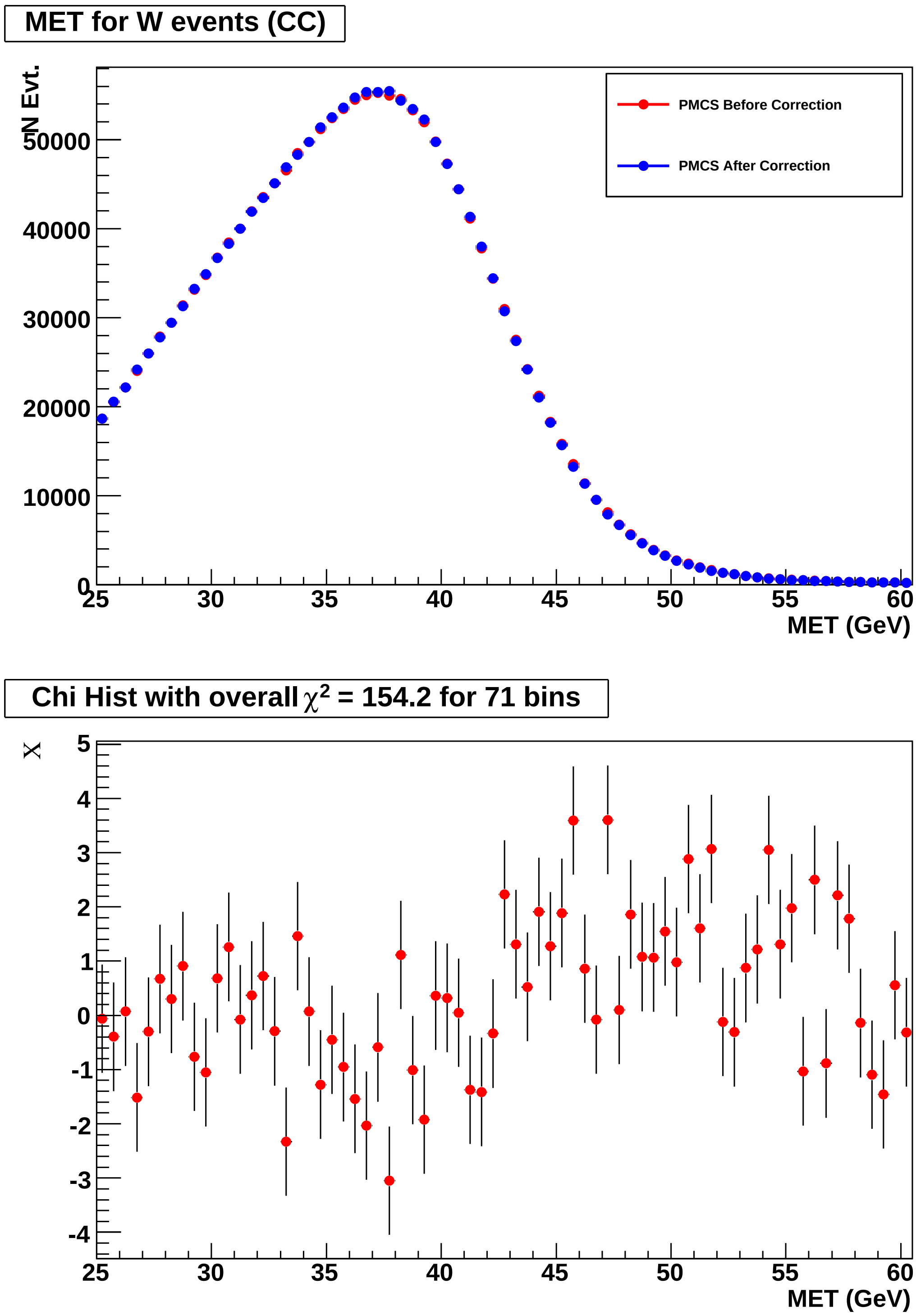

Figure 23: Missing E_T comparison of W$\rightarrow e\nu$ events between FastMC before and after the correction of efficiency. The number of events are scaled to 4.3 fb^{-1} RunIIb W$\rightarrow e\nu$ data. The bottom plot is the χ distribution of the comparison.

A Selection criteria

In this appendix we precisely define the selection criteria for the efficiency study. Some frequently referred selection observables and criteria are defined below:

- CC: $|\eta_{det}| < 1.05$.
- EC: $1.5 < |\eta_{det}| < 2.5$.
- ISO: Regular calorimeter isolation variable defined as the fraction of calorimeter energy in the isolation region ($0.2 < \Delta R < 0.4$) to the energy of the EM cluster, where ΔR is the distance between EM cluster barycenter and a calorimeter tower.
- EMFraction: Ratio of the energy deposited in the electromagnetic part of the calorimeter to the total energy including the hadronic calorimeter.
- HMx7 and HMx8: H-matrix represents lateral and longitudinal shower shapes of EM cluster. It is the χ^2 of the covariance matrix built for CC (HMx7) and EC (HMx8).
- TrkMatch: The track matching probability, is calculated based on the distance between the cluster barycenter and the associated track position at layer 3 of the EM Calorimeter.
- H-Matrix criterion: For electrons in CC, HMx7 < 12; for electrons in EC, HMx8 < 20 .
- loose-track-match criterion: A track exists in the vicinity of the cluster;
- tight-track-match criterion: TrkMatch > 0.01, associated track has at least one SMT hit and track $p_T > 10$ GeV.
- pass trigger: In the Run IIb analysis we employ triggers "E1_SHT25" and "E1_SHT27". The trigger object matching is performed for all levels of the trigger. Pass trigger is defined as the reconstructed electron EM cluster is matched with the trigger level EM object that fires the trigger.
- dead cells: Major regions inside the calorimeter containing dead cells are excluded from the measurement of the efficiencies. The exclusion regions are defined in Appendix A.1.

A.1 Major regions excluded containing dead cells

In the DØ calorimeter, there are cells that have been permanently disconnected from the readout. In the data, there can be more dead regions because of temporary failures or imperfections in parts of the readout electronics. Such faulty hardware is typically replaced very quickly, but some of the more subtle imperfections can affect larger chunks of data. In the full MC, only these permanently dead regions are simulated.

The major regions containing dead cells are flagged in $(\eta_{\rm det}, \phi_{\rm det})$ space. The following exclusion regions are only for CC calorimeter.

In the full MC samples, the following $(\eta_{\rm det}, \phi_{\rm det})$ regions are excluded:

(1) $-0.55 < \eta_{\rm det} < -0.35$ and $2.35 < \phi_{\rm det} < 2.5$,

(2) $-0.95 < \eta_{\rm det} < -0.75$ and $2.55 < \phi_{\rm det} < 2.7$,

(3) $0.35 < \eta_{\rm det} < 0.55$ and $2.6 < \phi_{\rm det} < 2.75$.

In the collider data samples, the following $(\eta_{\rm det}, \phi_{\rm det})$ regions are excluded:

(1) $-0.9 < \eta_{\rm det} < -0.8$ and $2.55 < \phi_{\rm det} < 2.75$,

(2) $-1.0 < \eta_{\rm det} < -0.9$ and $5.89 < \phi_{\rm det} < 6.28$,

(3) $-0.6 < \eta_{\rm det} < -0.5$ and $5.89 < \phi_{\rm det} < 6.09$.

(4) $-0.5 < \eta_{\rm det} < -0.4$ and $2.35 < \phi_{\rm det} < 2.56$.

(5) $0.2 < \eta_{\rm det} < 0.3$ and $2.75 < \phi_{\rm det} < 2.95$.

(6) $0.4 < \eta_{\rm det} < 0.5$ and $2.55 < \phi_{\rm det} < 2.75$.

A.2 Selection criteria for H-Matrix efficiency study

Base selection criteria applied to $Z \to ee$ events for H-Matrix efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy EMFraction > 0.9 .

- Both electrons satisfy ISO < 0.15 .
- Electrons can be either in CC or in EC. If in EC, it must satisfy HMx8<20 . There is no requirement for both electrons to be in CC at the same time.

Electrons pass or fail the H-Matrix criterion is defined as:

- **Pass:** *this electron* is in CC, HMx7<12 , and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.
- **Fail:** *this electron* is in CC, HMx7>12 , and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.

Note that we require *the other* electron to pass trigger when we study *this* CC electron. The idea is that we study the efficiency of the CC electrons, but we also accept the EC electrons to tag the events, and only use the trigger to tag. Therefore, we can remove the inefficiency dependences introduced by the trigger, but do not introduce bias in the "tag-and-probe" method.

A.3 Selection criteria for loose-track-match efficiency study

Base selection criteria applied to $Z \to ee$ events for loose-track-match efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy EMFraction > 0.9 .
- Both electrons satisfy ISO < 0.15 .
- Both electrons pass H-Matrix criterion.

Electron pass or fail the loose-track-match criterion is defined as:

- **Pass:** *this electron* is in CC, passes loose-track-match criterion, and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.

- **Fail:** *this electron* is in CC, fails loose-track-match criterion, and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.

Require *the other* electron to pass trigger when we study *this* CC electron, this follows the same idea as discussed in the end of Appendix A.2.

A.4 Selection criteria for tight-track-match efficiency study

Base selection criteria applied to $Z \to ee$ events for tight-track-match efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy EMFraction > 0.9 .
- Both electrons satisfy ISO < 0.15 .
- Both electrons pass H-Matrix criterion.
- Both electrons pass loose-track-match criterion.

Electron pass or fail the tight-track-match criterion is defined as:

- **Pass:** *this electron* is in CC, passes tight-track-match criterion, and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.
- **Fail:** *this electron* is in CC, fails tight-track-match criterion, and its reconstruction window does not contain dead cells; and *the other electron* pass trigger.

Require *the other* electron to pass trigger when we study *this* CC electron, this follows the same idea as discussed in the end of Appendix A.2.

A.5 Pure background selection criteria for H-Matrix efficiency study

Base selection criteria applied to $Z \to ee$ events to select pure background for H-Matrix efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy EMFraction > 0.9 .
- Both electrons satisfy ISO < 0.15 .
- Both electrons can be either in CC or EC. There is no requirement for both electrons to be in CC at the same time.
- Both electrons fail loose-track-match criterion.
- Both electrons fail tight-track-match criterion.

Electrons pass or fail H-Matrix criterion to determine the pure background shape is defined as:

- **Pass:** *this electron* is in CC, HMx7<12 , and its reconstruction window does not contain dead cells.
- **Fail:** *this electron* is in CC, HMx7>12 , and its reconstruction window does not contain dead cells.

A.6 Pure background selection criteria for loose-track-match efficiency study

Base selection criteria applied to $Z \to ee$ events to select pure background for loose-track-match efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy ISO < 0.15 .
- Both electrons fail H-Matrix criterion.
- No EMFraction criterion for both electrons.

Electrons pass or fail loose-track-match criterion to determine the pure background shape is defined as:

- **Pass:** *this electron* is in CC, pass loose-track-match criterion, and its reconstruction window does not contain dead cells.
- **Fail:** *this electron* is in CC, fail loose-track-match criterion, and its reconstruction window does not contain dead cells.

A.7 Pure background selection criteria for tight-track-match efficiency study

Base selection criteria applied to $Z \to ee$ events to select pure background for tight-track-match efficiency study:

- The event has at least two reconstructed electrons. If more than two, take the leading two according to $p_T(e)$.
- The invariant mass $M(ee)$ is between 60 and 130 GeV.
- Recoil p_T (u_T) satisfies $u_T < 30$ GeV.
- Both electrons do not fall in Phi Cracks (calorimeter inter-module gaps in ϕ).
- Both electrons satisfy $p_T(e) > 25$ GeV.
- Both electrons satisfy ISO < 0.15 .

- Both electrons fail H-Matrix criterion.
- Both electrons pass loose-track-match criterion.
- No EMFraction criterion for both electrons.

Electrons pass or fail tight-track-match criterion to determine the pure background shape is defined as:

- **Pass:** *this electron* is in CC, pass tight-track-match criterion, and its reconstruction window does not contain dead cells.
- **Fail:** *this electron* is in CC, fail tight-track-match criterion, and its reconstruction window does not contain dead cells.

References

[1] W mass group, W mass measurement in Full Geant Monte Carlo in RunIIb, DØ Note 6267 (2012)

[2] Hengne Li, Ph.D Thesis, LAL – Université Paris-Sud XI, LAL-09-118, October 2009