DØNote 2938

$W \rightarrow e\nu$ Data-based $W \rightarrow \tau\nu$ Monte Carlo

Greg Griffith, Chang Kee Jung, Jon Kotcher,
Hailin Li,Qi-zhong Li-Demarteau,
Serban Protopopescu, Djoko Wirjawan

May 1, 1996

Abstract

This note describes the procedure for generating the $W \rightarrow e\nu$ data-based $W \rightarrow \tau\nu$ Monte Carlo. To verify the procedure, the $W \rightarrow e\nu$ data-based $W \rightarrow e\nu$ Monte Carlo is generated and compared with the original $W \rightarrow e\nu$ data. It's found that data-based Monte Carlo models the data very well in the calorimeter. This gives us confidence in data-based $W \rightarrow \tau\nu$ Monte Carlo which is used to carry out $W \rightarrow \tau\nu$ analysis.

1 Motivation

The best way to identify the τ from W decays at DØ is through its hadronic decays which accounts for 64% of its total branching ratio, since it's very difficult to distinguish τ's leptonic decays from leptonic decays of Ws. Only less than 1% of τ's decays have more than three charged tracks. Therefore the event signature for

$$\bar{p}p \rightarrow W \rightarrow \tau\nu$$

where $\tau \rightarrow$ hadrons $+ \nu$ will be a large $\not\!\!E_T$ and a τ–jet. The τ–jet is characterized by its narrowness, its hadronicity and its low multiplicity of charged particles.

In order to identify τs from W decays, we need a good Monte carlo with a detailed detector simulation which models the data very well, since there's no data available as pure signals to study taus from W. In addition, several things are difficult to model with a Monte Carlo: noise, underlying event and jet multiplicities. Those things are crucial to tau identifications from W decays though.

Since $W \rightarrow e\nu$ is well defined and understood and also has the same event topology as $W \rightarrow \tau\nu$, the best way is to replace electrons in the $W \rightarrow e\nu$ data with taus simulated by Monte Carlo. This way only τ decays and decay products through the detector are simulated with a Monte Carlo. And also the noise, underlying event and multiple interactions which become more important at higher luminosities are taken care of automatically.

2 Generation of Data-based Monte Carlo(DBMC)

Generation of Data-based Monte Carlo(DBMC) consists of selection of good $W \rightarrow e\nu$ events, generation of monte Carlo τ events and merging of the above two.

2.1 Selection of a good $W \rightarrow e\nu$ sample

We used the following selection criteria:

- $|\eta_{\text{detector}}| \leq 1.0$.
- $E_T \geq 25$ GeV.
- $\not\!\!E_T \geq 25$ GeV.
- Quality cuts:
  1. $\chi^2 \leq 100$.
  2. EM fraction ≥ 0.90.
  3. isolation(0.4-core) ≤ 0.15.
  4. $\sigma_{trk} \leq 10$.
  5. excluding crack regions in ϕ (0.01 away from cracks)
- remove bad runs

This sample only contains about 3% background, according to the estimate by the electron group.

2.2 Generation of Monte Carlo τ events and merging with the original $W \rightarrow e\nu$ events

Let's consider just one good $W \rightarrow e\nu$ event, and follow it through the whole procedure. Let's call this good $W \rightarrow e\nu$ event the parent event.

Here are the procedures:

- **Decay through ISAJET**: take the momentum $\mathbf{P}_e^p$ of the leading electron and the event vertex $\mathbf{V}^p$ from the parent event, and input to ISAJET. In the ISAJET, the τ parton gets the same momentum $\mathbf{P}_e^p$ and the event gets the same event vertex $\mathbf{V}^p$. What ISAJET does is just to let τ decay.
- **Detector Simulation and Reconstruction**: the event with the τ's decay products from ISAJET goes through the detector simulation DØGEANT and the reconstruction DØRECO. This event is called the daughter event.

- **Calorimeter Hit Dropping:** the electron calorimeter hits in the parent event are stored in the corresponding CASH bank and their energies are just zeroed. By doing so, we also remove the noise associated with each cell in the electron cluster. In order to put back the noise, for each cell in the electron cluster, a cell is searched which is ± 4 ϕ_w away in ϕ from the electron cluster center and at the same η and layer as the cell in the electron cluster, where ϕ_w is the electron cluster width in ϕ from its centroid. We also check the cell ± 4 cluster width away should not belong to any reconstructed object and not align with the original electron cluster within 15° in ϕ. If the cell belongs to any object, then the energy of the cell consists of not only noise but the true energy from a particle, which is not what we want. If such a cell can not be found, the event is not kept and we go to the next parent $W \rightarrow e\nu$ event. In fact this occurs very rarely, less than 1%.

- **Tracking Hit Dropping:** the electron tracking hits which are compressed in the parent event are stored in the DHIT bank associated with the electron track for a CDC track, since we are only concerned with central τs. The hits on the electron track are removed from the DHIT bank.

- **Merging:** Once we have the daughter event which only contains τ's decay products and the parent $W \rightarrow e\nu$ event with the electron hits dropped, we can merge the daughter event with the parent event. For calorimeter hit merging, the energy of the cell with the same index in both events are summed up. For tracking hit merging, the hits on a CDC track in the daughter event are corrected for the difference in the detector geometry between Monte Carlo and data, and then put together with the rest of the hits in the parent event after the hits on the electron track are removed. Note the hits have to be reordered to make sure all hits in the same wire, sector and layer of CDC are put together and numbered. If any hit in the parent event is within the hit resolution (18 ns) of a Monte Carlo hit in the same wire, sector and layer, the hit in the parent event is dropped to avoid double counting. The FDC hits in the parent event, if there's any, is left untouched, since the τ in the daughter event and the electron in the parent event are all in the central region.

The above procedure is repeated for each good $W \rightarrow e\nu$ event. Then the merged events are run through DØRECO and trigger simulator.

3 Verification of Data-based Monte Carlo

To verify the whole procedure above, and also double check our DØGEANT, we generated data-based $W \rightarrow e\nu$ Monte Carlo and compared them with the original data used to generate them, using the same procedure as described above.

Figures 1 to 9 all have a cut $E_T > 25$ GeV for the electrons in DBMC. Figure 1 and Fig. 2 show the E_T and E comparisons respectively. The mean values in E_T and E are both shifted by about 1 GeV. To understand this, we first checked the differences between the reconstructed energy and parton energy in a Monte Carlo, shown in Fig. 10. We require a match between the reconstructed electron cluster and the electron parton in both η and ϕ.

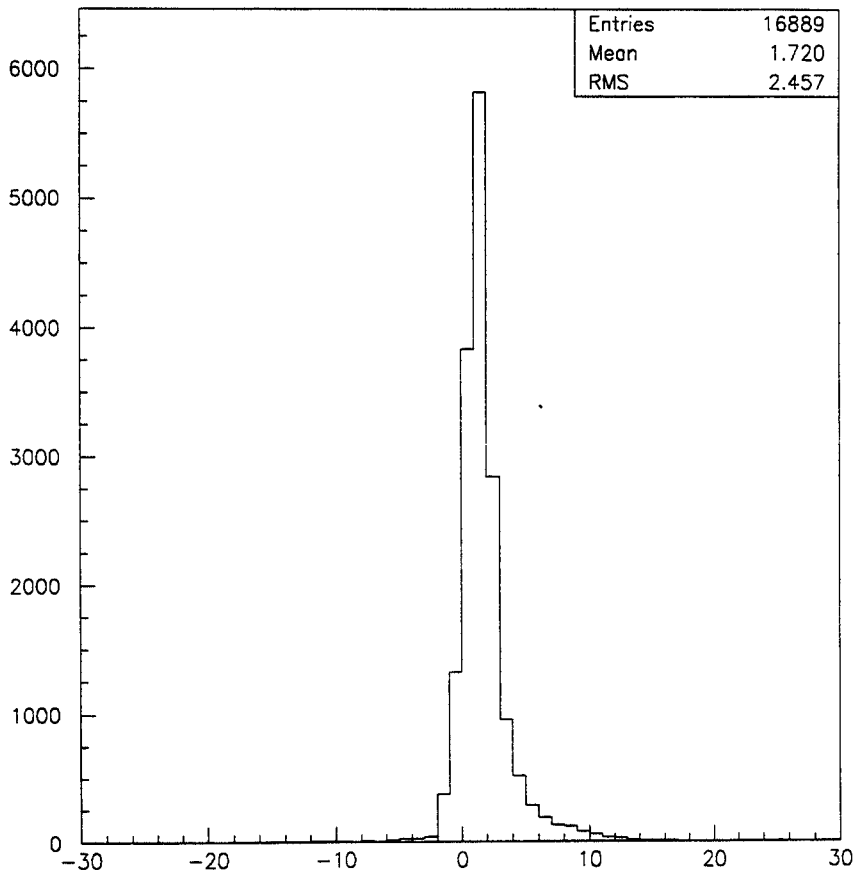

Figure 10: Difference in E_T between the reconstructed electron cluster and the electron parton.

We also applied the same cut $E_T > 25\text{GeV}$ for the reconstructed electron cluster. We see a similar shift in the Monte Carlo. Since the electrons in the DBMC are simulated with Monte Carlo, the shift is due to Monte Carlo itself, not due to the above procedure.

Figures 3 and 4 compare the η and ϕ of the leading electron in DBMC with those in DATA. An excellent agreement is evident. Fig. 5 compares the isolation in the 0.4 cone in DBMC with that in DATA. We also see a very good agreement.

The isolation in 0.7 cone is worth looking at carefully. Note one important cut in the tau identification is the jet size parameter:

$$RMS = \sqrt{(\delta\eta)^2 + (\delta\phi)^2}$$

where

$$\delta\eta = \sqrt{\sum_i \frac{E_T^i(\eta_i - \bar{\eta})^2}{E_T^i}}$$

$$\delta\phi = \sqrt{\sum_i \frac{E_T^i(\phi_i - \bar{\phi})^2}{E_T^i}}$$

i denotes tower indices in a jet.

For a tau cluster, RMS is very small; while for a QCD jet, RMS will be large. Cone 0.7 is chosen to maximize the difference in RMS between a narrow jet and a fat jet. Note also RMS is very sensitive to the noise, as can be seen from its definition. The isolation in the cone 0.7 is similar to RMS, as least occupying the same region in η-ϕ space. So the agreement in the isolation 0.7 between DBMC and DATA can give us much confidence in that RMS in DBMC $W \rightarrow \tau\nu$ simulates that of $W \rightarrow \tau\nu$ data well. We did see a very good agreement in the isolation 0.7 between DBMC and DATA, as shown in Fig. 6. In fact, it's the very suspicious RMS shape in the existing Monte Carlo that prompts us to start the DBMC project.

Fig. 7 shows the distributions of the EM fraction of the electron cluster in DBMC and DATA. The differences is probably due to the fact that the electron deposits its energy in the EM calorimeter in DØGEANT more than it does in DATA. Note this may not affect $W \rightarrow \tau\nu$ much, since τ decays hadronically.

Fig. 8 compares the $\not\!\!E_T$ distributions. The $\not\!\!E_T$ for the $W \rightarrow e\nu$ data is corrected with both EM and jet corrections, while $\not\!\!E_T$ for DBMC is corrected with jet corrections only. Since the electron E_T in DBMC is smaller by about 3% than that in DATA, we corrected the $\not\!\!E_T$ after we scaled up the electron E_T in DBMC. And we found an excellent agreement as shown in Fig. 11 and Fig. 12. This is expected, since only the electron is simulated with Monte Carlo, and the rest of the event is untouched.

We have seen a very good agreement between DBMC and DATA in the calorimeter quantities. Fig. 9 compares the number of tracks within the electron road 0.1 by 0.1 in η and ϕ. DBMC has more tracks than the original DATA: one track events fewer than DATA; two or more track events more than DATA. Briefly, this is due to the fact that the electron tracking hits are simulated with DØGEANT and they are very ideal;but track finding with those hits uses real DATA parameters which constrain the track finding(RCP) and are relatively looser than the corresponding Monte Carlo parameters, thus building lots more tracks than it should. Note DBMC contains two types of hits: electron hits from Monte Carlo and the rest from real DATA. Fortunately, the discrepancy will not affect the $W \rightarrow \tau\nu$ analysis much, since we cut very loosely on the number of tracks in $W \rightarrow \tau\nu$.

We also compared the trigger simulation results. Since the calorimeter quantities in DBMC agree fairly well with the DATA, we would expect trigger simulation results agree well, too. We did double-check this.

We ran the trigger simulator on DBMC events with the same trigger conditions as used in the data. Because tau trigger uses $\not\!\!E_T$ which is not easy to model well in a Monte Carlo, we want to pick a trigger which has the $\not\!\!E_T$ requirements. we chose the following trigger:

- Level 1 requirements (EM_1_MAX): EM cluster energy > 12 GeV.
- Level 2 requirements (EM1_EIS06_MS):
  - EM cluster energy > 20 GeV.
  - isolation(0.6-0.2) < 0.15.

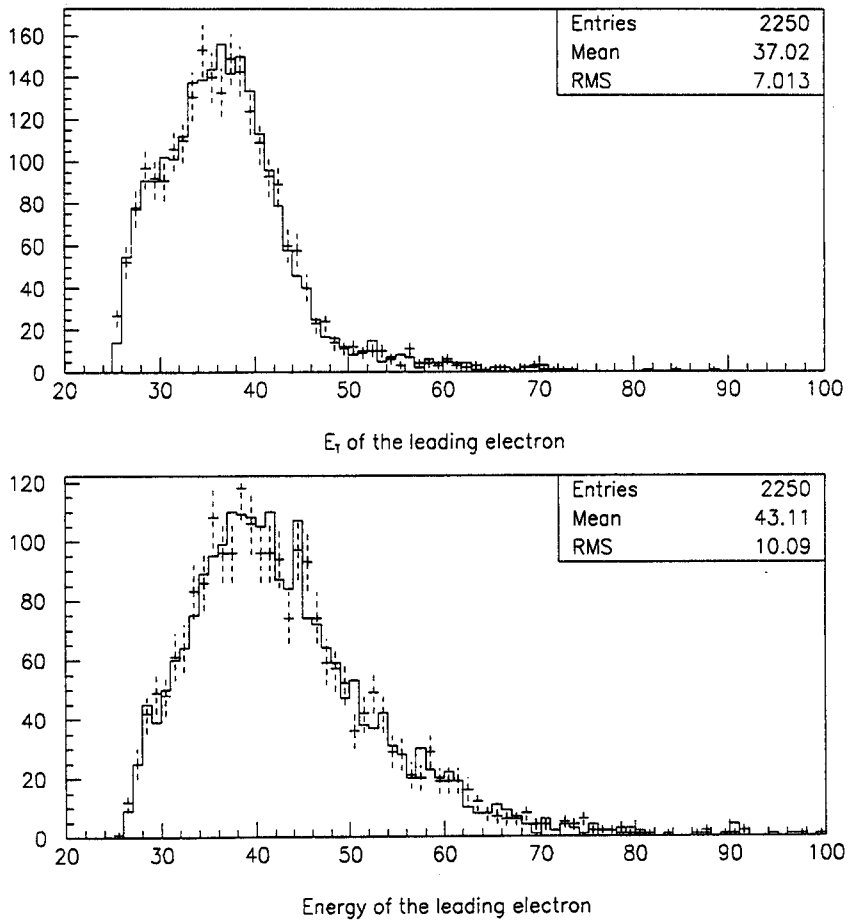

Figure 11: Comparison of the energy and E_T of the electron in DBMC with DATA after the 3% scale correction to the electron energy in DBMC. The solid histogram denotes DATA and the dashed line denotes DBMC.

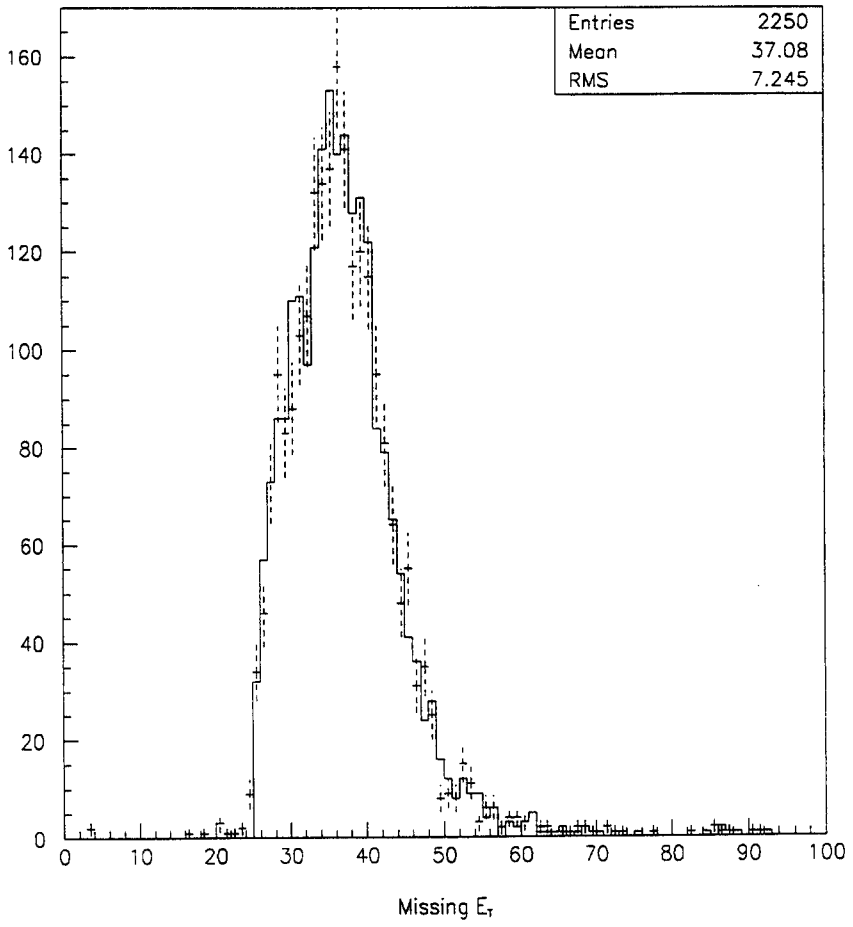

Figure 12: Comparison of $\not\!\!E_T$ in DBMC with DATA after the 3% scale correction to the electron E_T in DBMC. The solid histogram denotes DBMC and the dashed line denotes DATA.

- $\not\!\!E_T > 15$ GeV.

Out of 2706 events originally passing the above trigger, 2659 events in DBMC pass the same trigger. So the ratio

$$R = \frac{2659}{2706} = (98.26 \pm 0.25(\text{stat}))\%$$

shows DBMC models DATA very well in terms of trigger level quantities. Discrepancies are expected from comparisons of offline quantities, which already exists in our existing Monte Carlo.

4 Conclusions

We have shown DBMC models our data very well in the calorimeter. More work is needed to get tracking close to DATA. Because $W \to \tau\nu$ analysis mainly uses calorimeter quantities, tracking will not have a major hit on the analysis. Moreover, tracking related efficiencies can be derived from data directly.

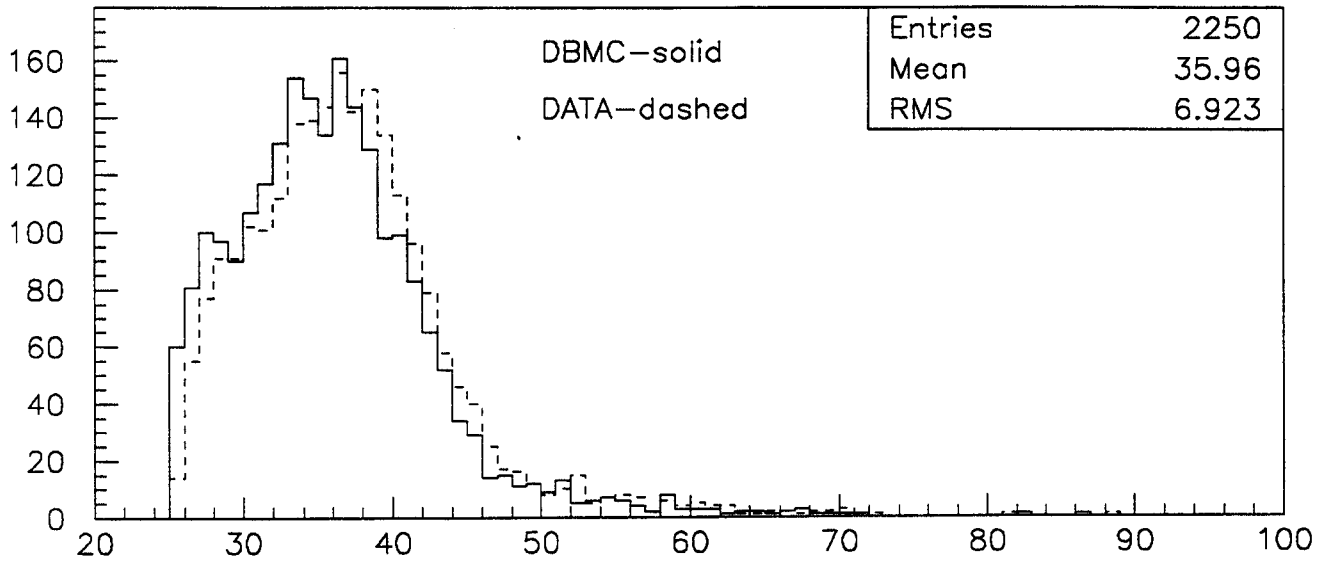

Figure 1(a) Et of leading electron

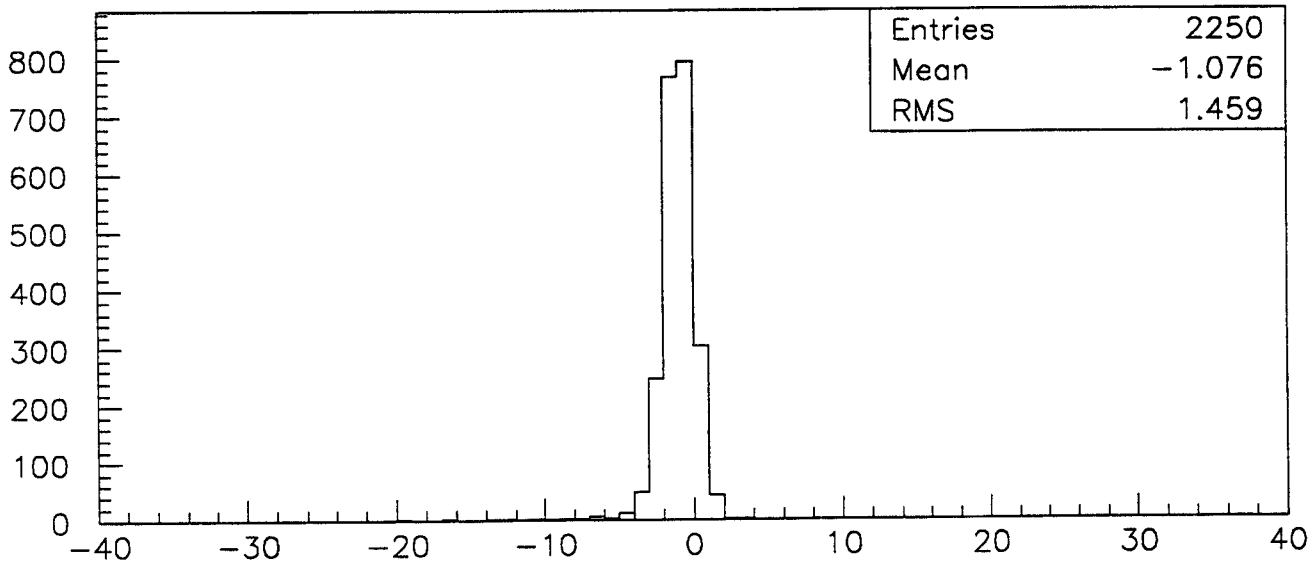

Figure 1(b) Et difference(DBMC−DATA)

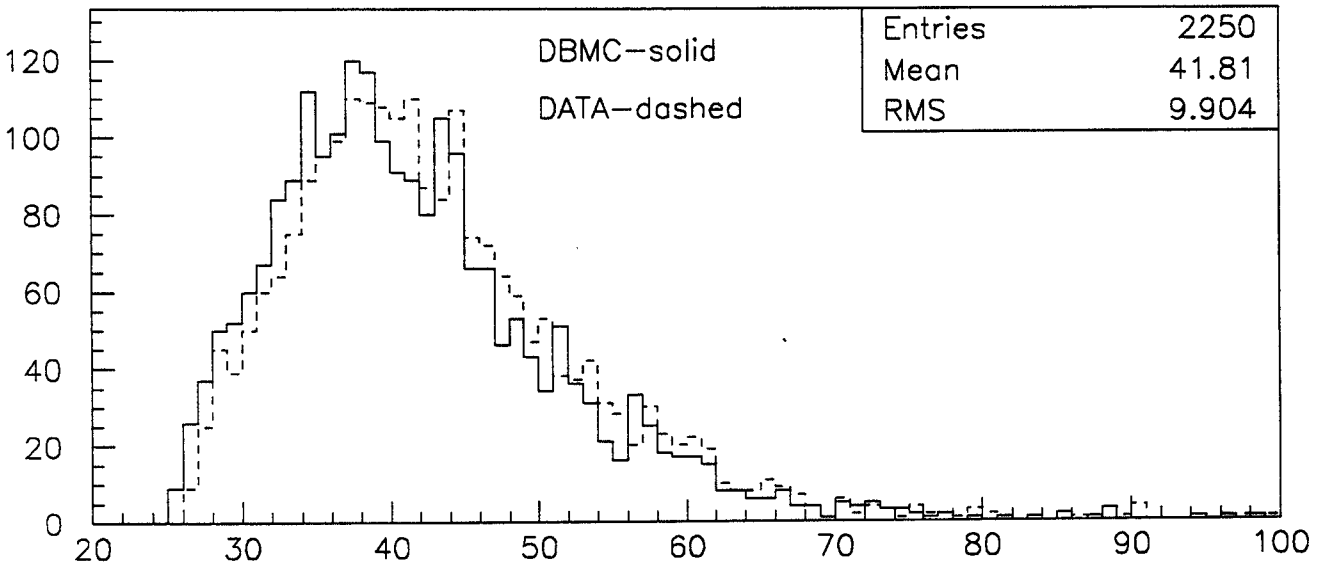

Figure 2(a) Energy of leading electron

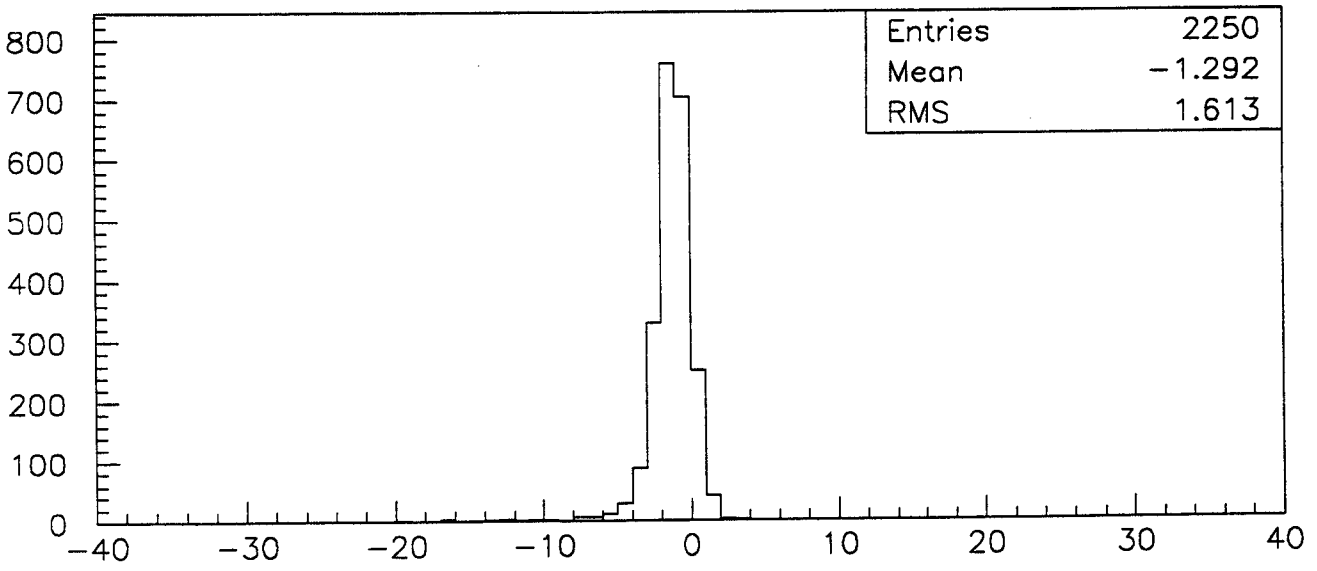

Figure 2(b) E difference(DBMC—DATA)

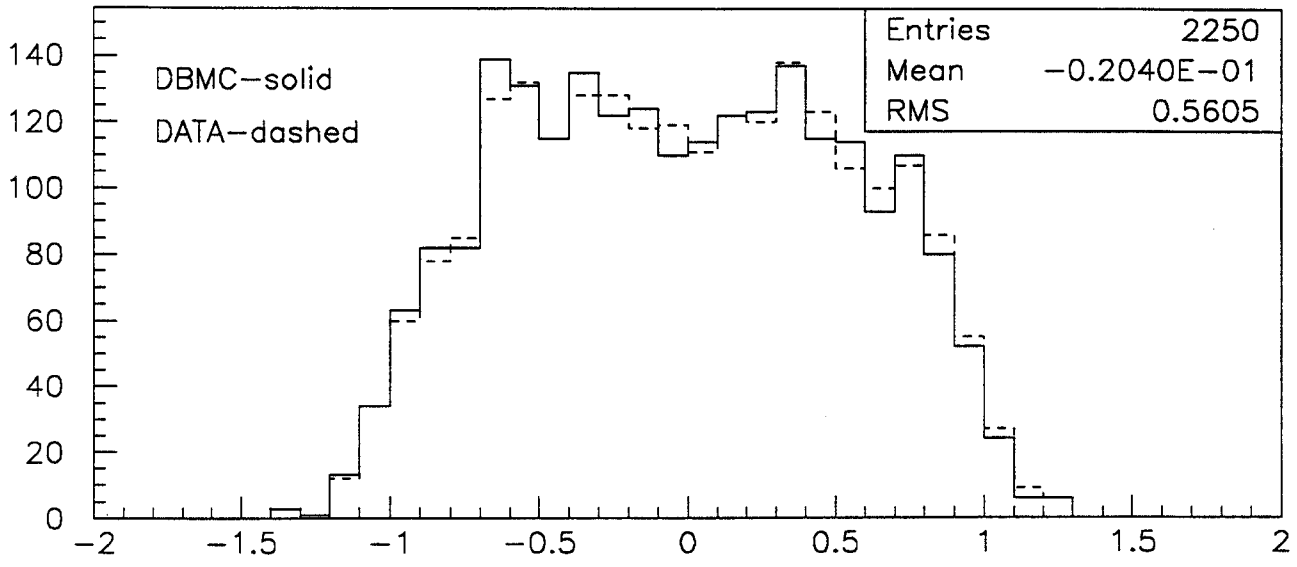

Figure 3(a) Eta of leading electron

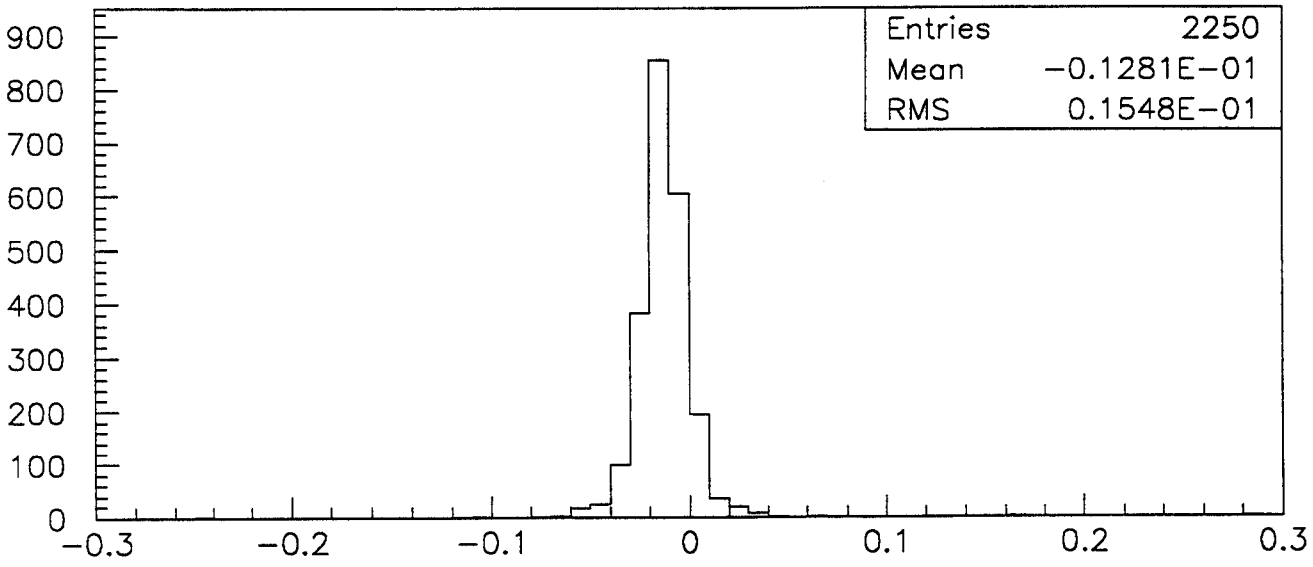

Figure 3(b) Eta differences (DBMC−DATA)

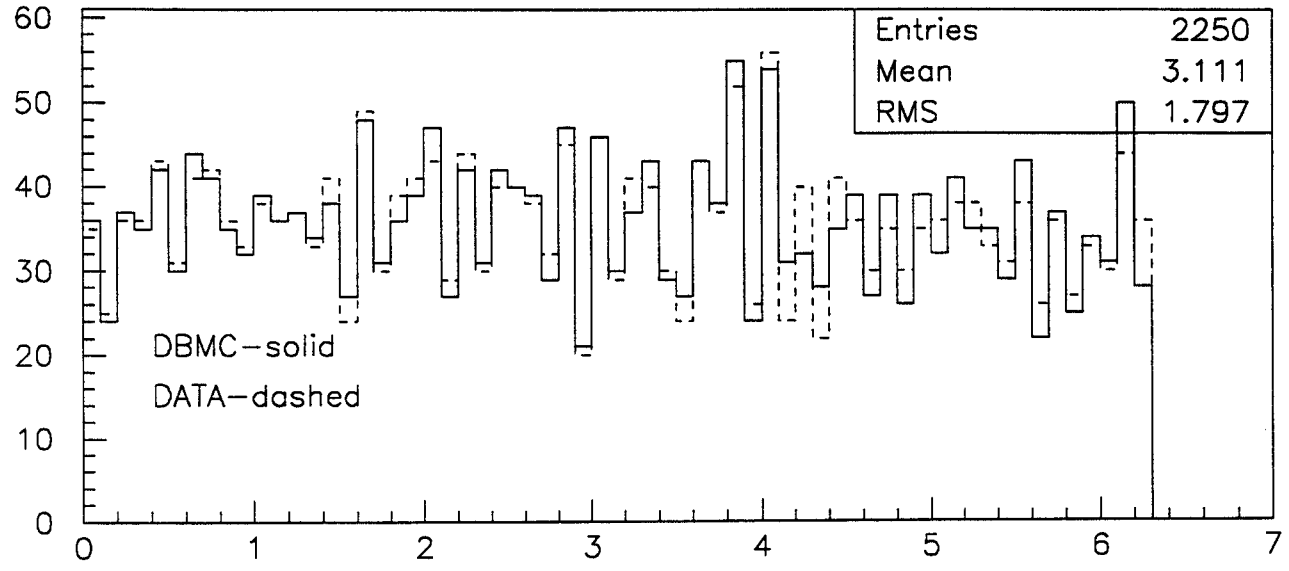

Figure 4(a) phi of leading electron

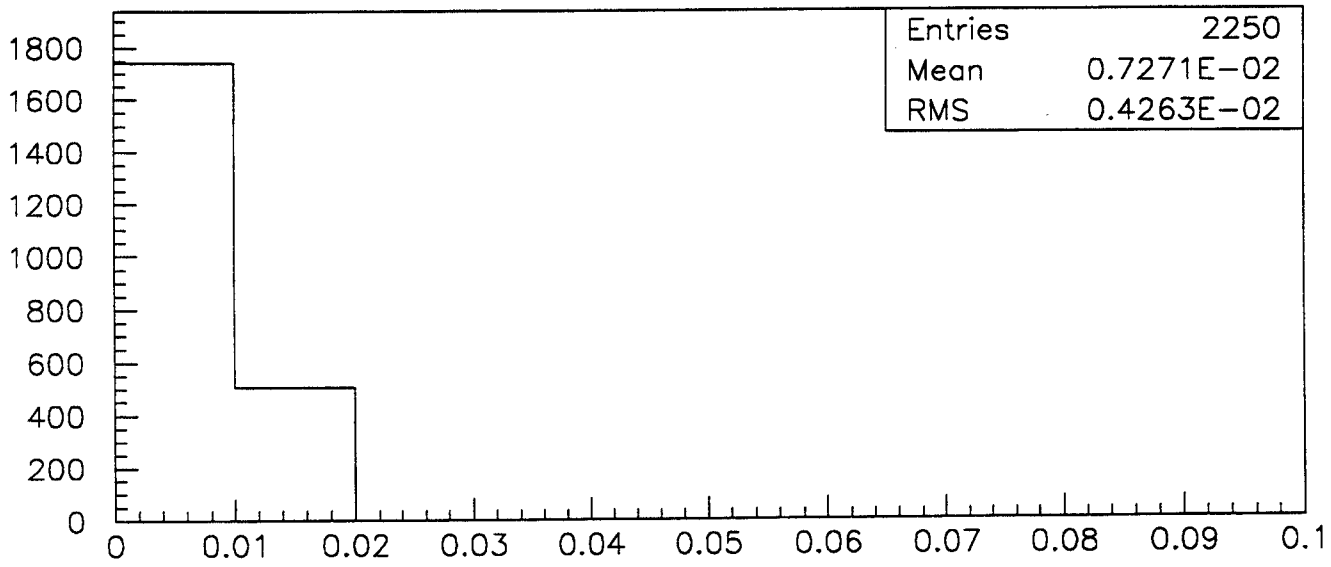

Figure 4(b) phi difference(DBMC−DATA)

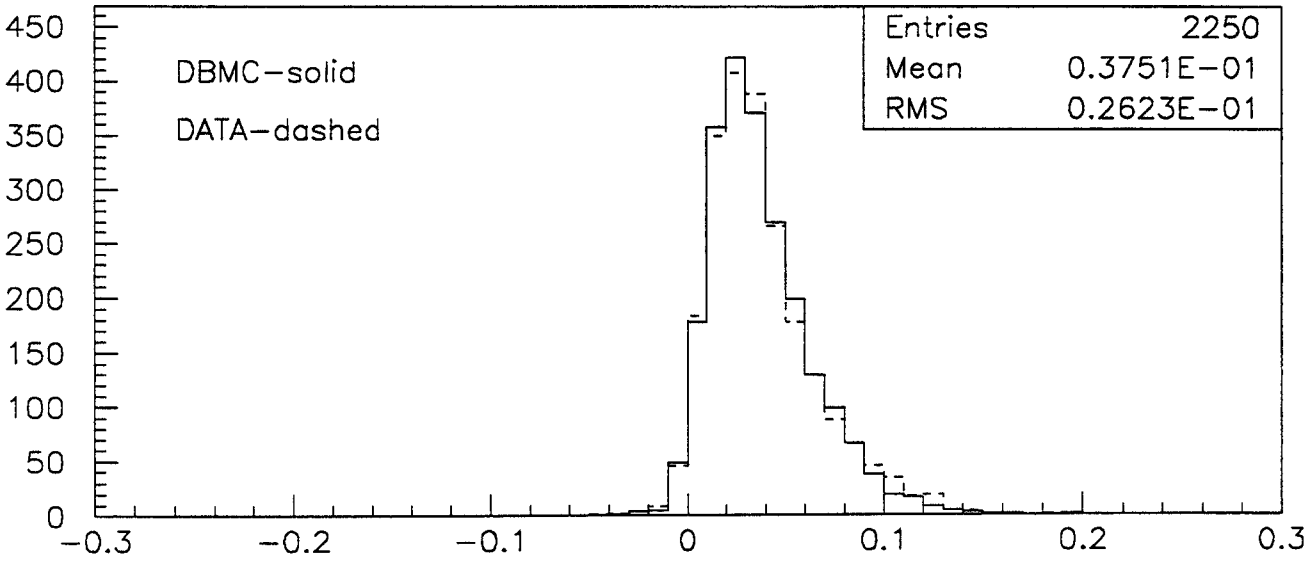

Figure 5(a) isolation in 0.4−0.2 of leading electron

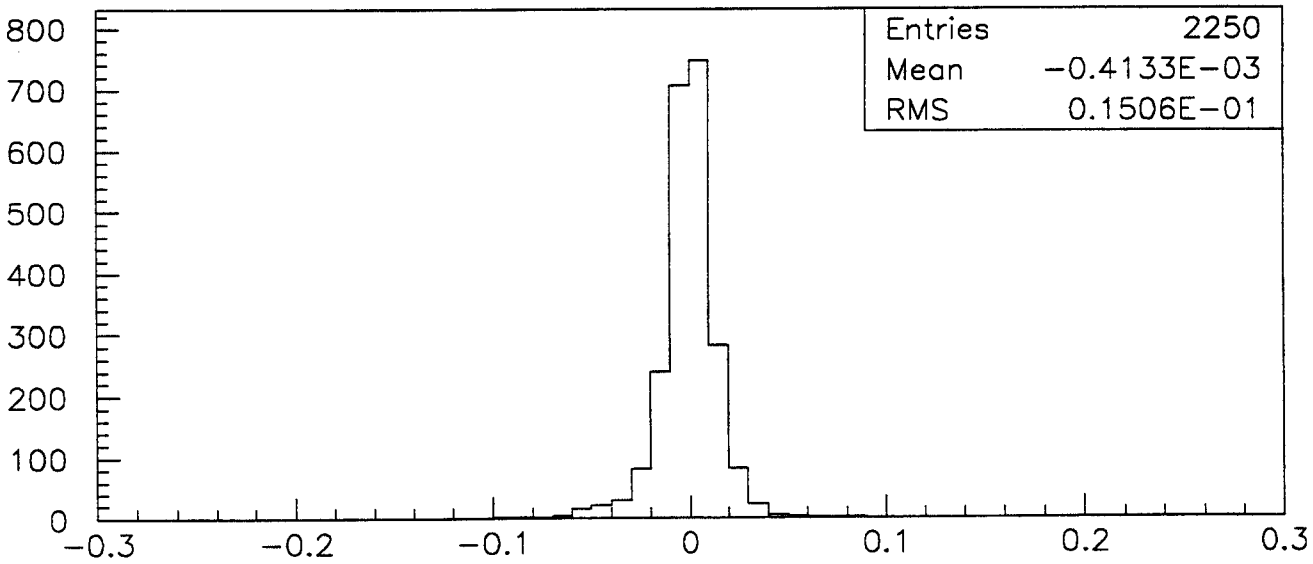

Figure 5(b) isolation difference in 0.4−0.2(DBMC−DATA)

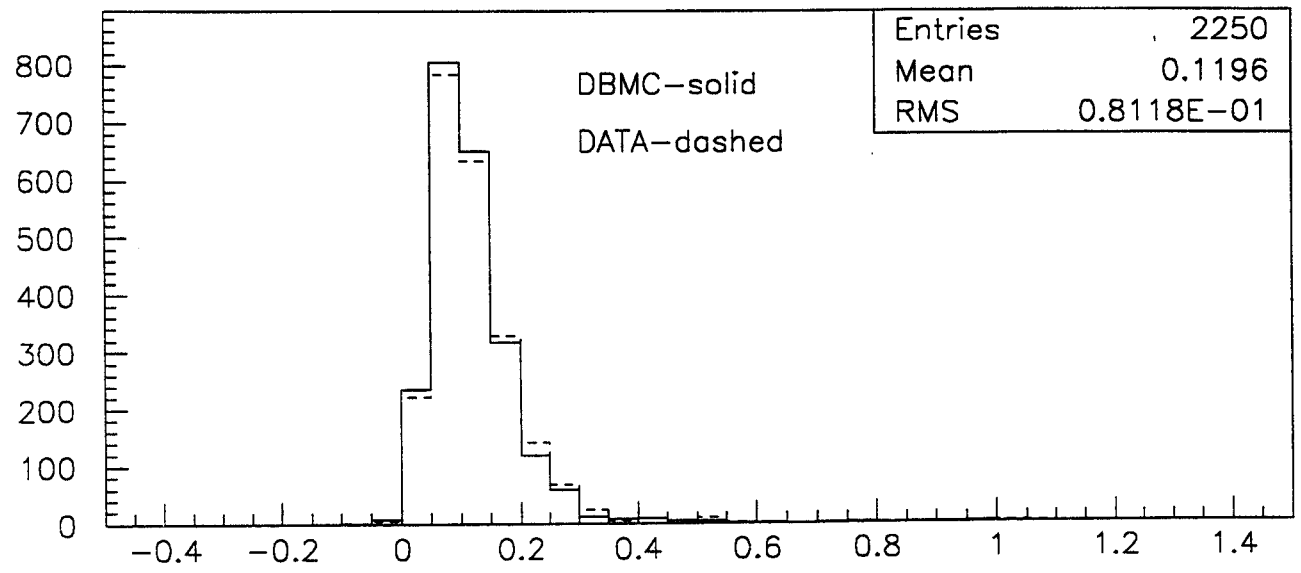

Figure 6(a) isolation in 0.7-0.2 of leading electron

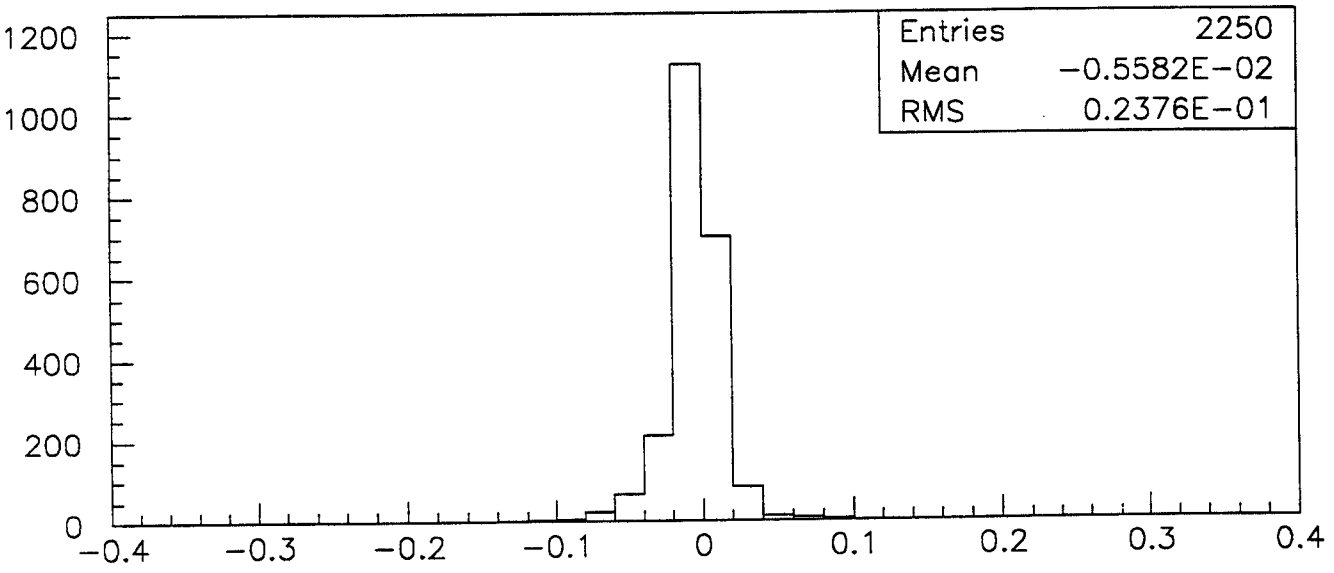

Figure 6(b) isolation difference in 0.7-0.2(DBMC-DATA)

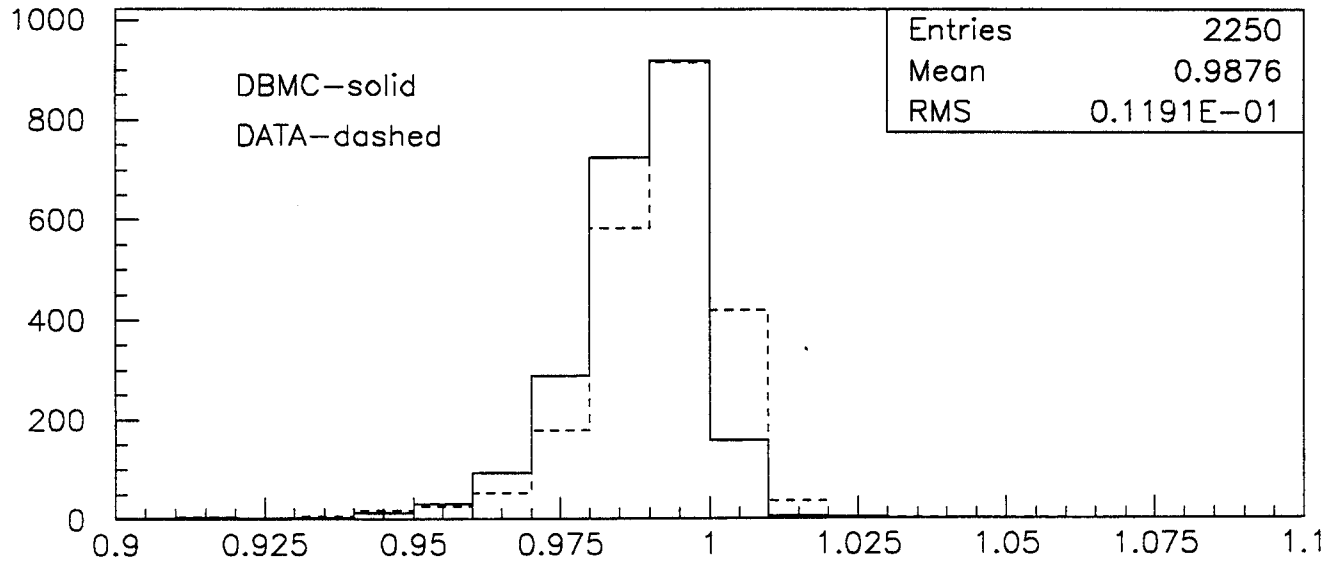

Figure 7(a) EM fraction of leading electron

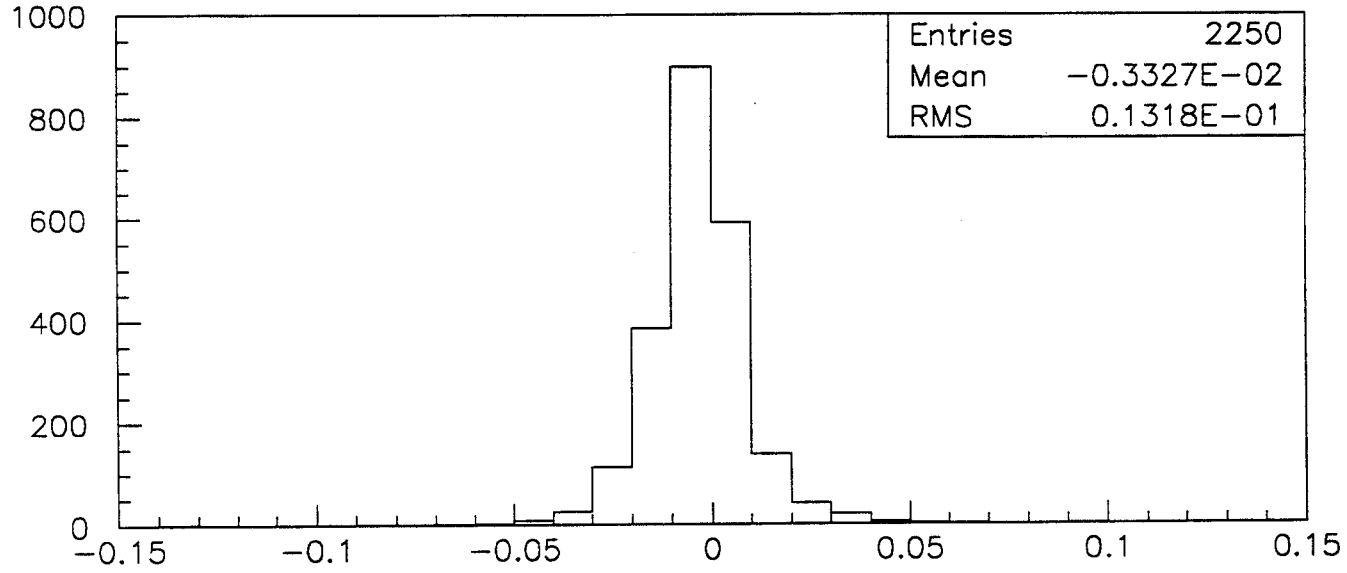

Figure 7(b) EMF difference(DBMC-DATA)

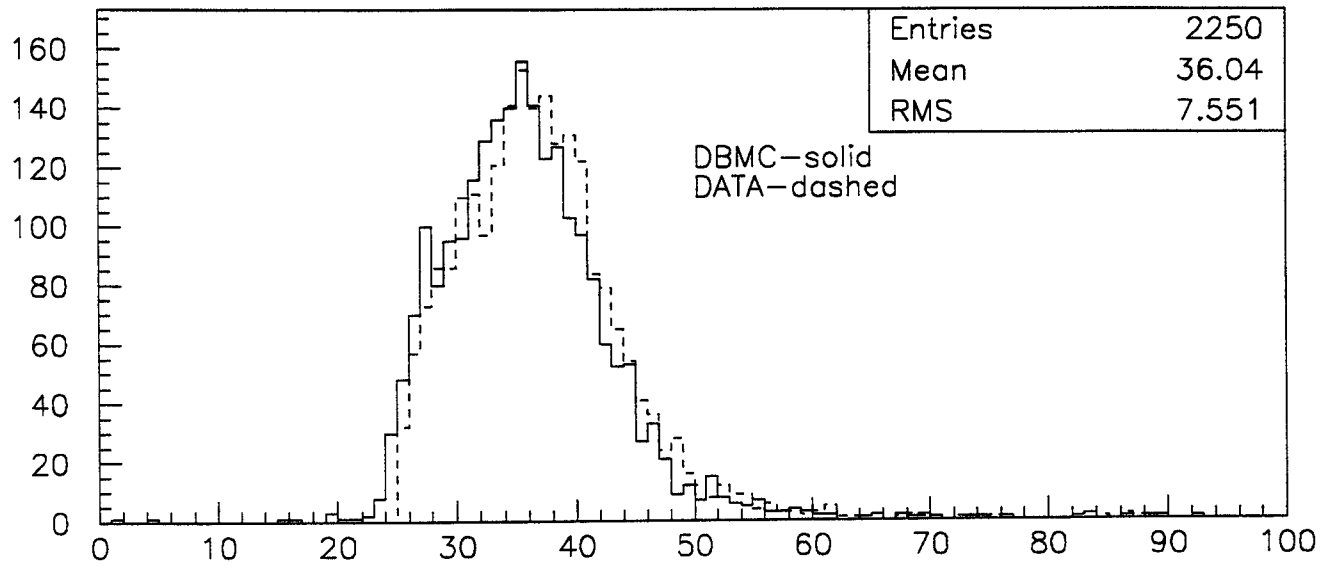

Figure 8(a) Missing Et

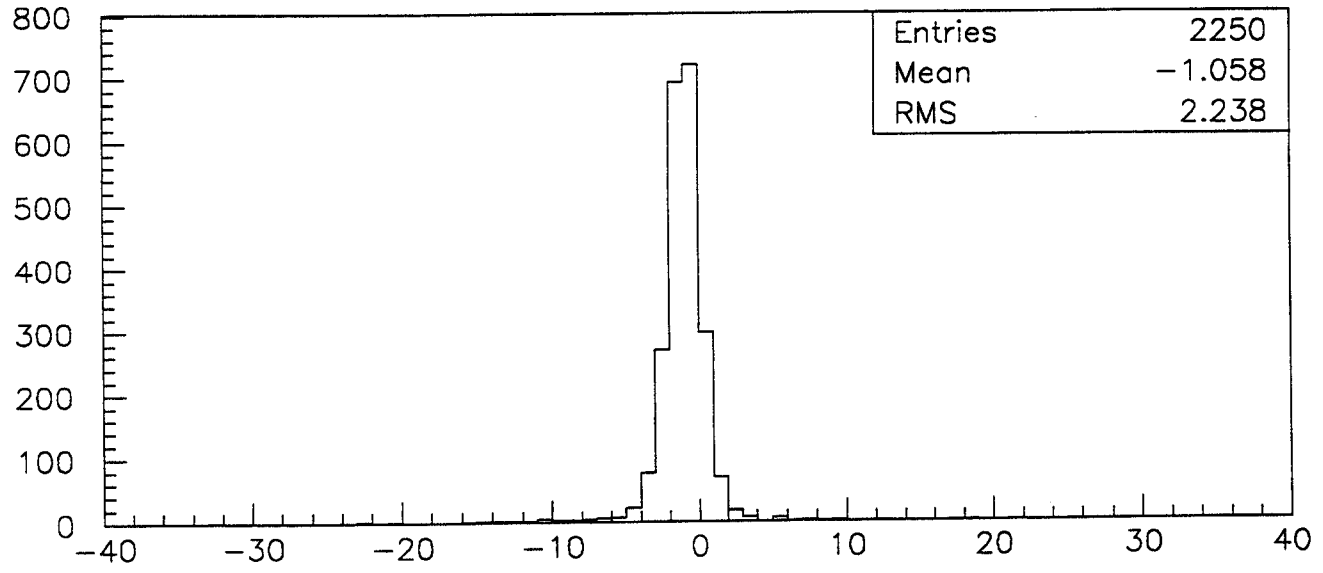

Figure 8(b) Missing Et difference(DBMC—DATA)

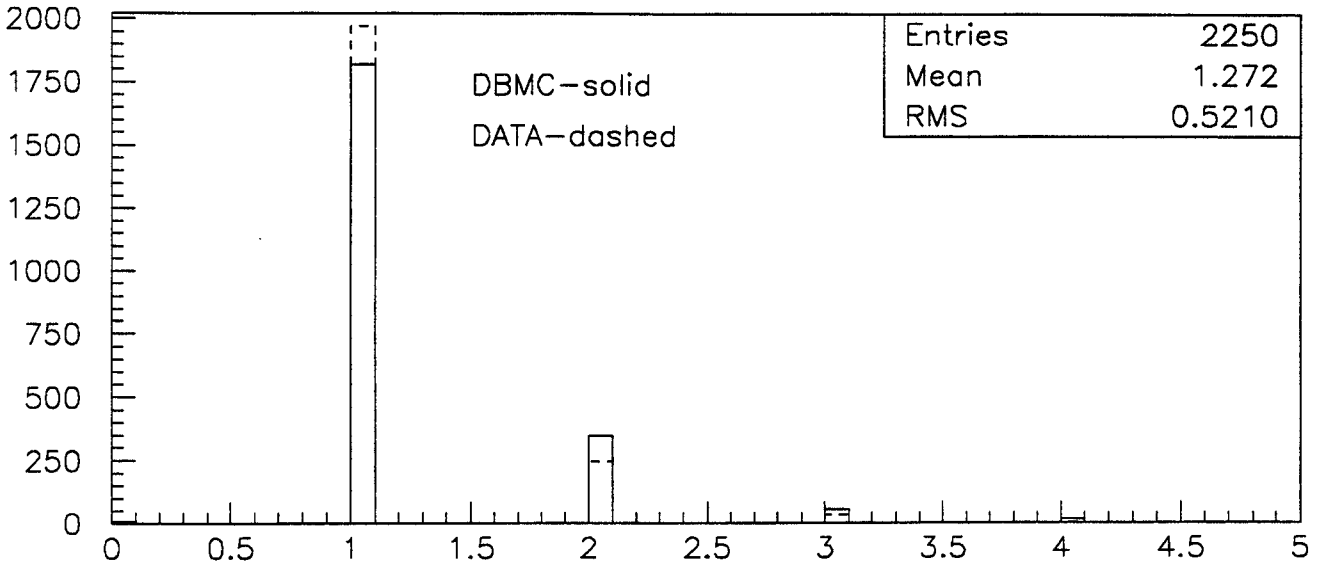

Figure 9(a) Number of Ztracks in road of leading electron

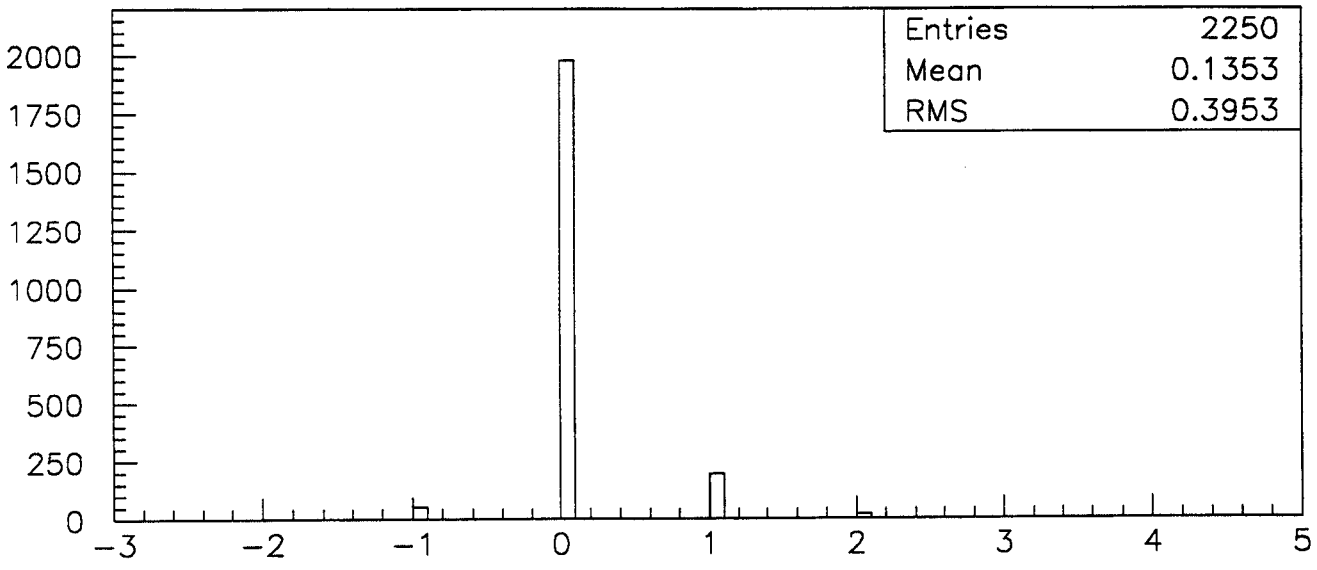

Figure 9(b) Difference in number of Ztracks in road(DBMC−DATA)