DØNOTE# 3071

15 Aug.96

Top via eμ using Neural Nets at DØ (1992-93 Data)

HARPREET SINGH and SUMAN BERI

Department of Physics, Panjab University,

Chandigarh, INDIA 160014

Abstract

We report on the results of a study on the channel $t\bar{t} \rightarrow e\mu$ using neural networks (NN). This study is different from previous studies of $e\mu$ channel in the sense that here we have taken into account the electron and muon id selection criterion as chosen in conventional analysis. These studies, conducted for run 1a correspond to an integrated luminosity of 13.9pb^{-1}. We observe one event in data with a background of $0.038 \pm 0.008(stat.)$ considering $Z \rightarrow \tau\tau \rightarrow e\mu$ and $WW \rightarrow e\mu$ as major background to this channel. We shall present a full status report of this analysis for run 1a, 1b and 1c data in an upcoming note.

1 Introduction

Top quark[1] events are being studied in the dilepton ($ee, e\mu, \mu\mu, e\nu$), lepton+jets and all-jets channels at DØ. In this note we discuss the use of feed forward neural networks for the study of $e\mu$ channel. The neural network approach is a multivariate technique, where networks map input feature space into one or more outputs. It is well established that this output is an approximation to a Bayes posterior probability,[2] and thus provides an optimal method to separate signal from background. The discriminant in this case is the output of the neural network, which for three layer feed forward neural net is given by

$$O(x) = g(\sum_k w_{jk} g(\sum_i w_{ki} x_i)$$

Where the x_i' s are the input vectors like $E_T^e, P_T^\mu, E_T^{jet1}, E_T^{jet2}, H_T, \not{E}_T$, etc. which in our case specify an event, w_{jk} and w_{ki} are the weights that are adjusted during the learning of the neural net and g represents a non linear transfer function like $\frac{1}{(1+e^{-2x})}$. The Bayes discriminant for neural nets is $R(x) = \frac{O(x)}{(1-O(x))}$. Neural networks have to be trained with the sufficient number of training patterns to obtain an optimal discrimination. We have done the studies of permuting different variables and effect of using different neural network training schedules. From this work we have arrived at the set of variables and the training schedule that gives us better discrimination.

In section 2 we describe the data sets used in this analysis, section 3 describes the neural network training schedule and section 4 presents the analysis and results with some conclusions.

2 Data Sample

In $p\bar{p}$ collisions the top quarks are produced mostly in pairs through $q\bar{q}$ annihilation. The standard model top decays to b quark through W boson. These are flavour changing charged currents.

$$q\bar{q} \rightarrow t\bar{t} \rightarrow Wb + W\bar{b}$$

Now, W boson further decay to leptons or quarks.

$$W \rightarrow \begin{cases} e/\mu/\tau + \nu & (1/3) \\ q\bar{q} \rightarrow 2jets & (2/3) \end{cases}$$

In $t\bar{t} \rightarrow e\mu$, one of the W goes to e and other goes to μ or vice versa. So in final state we have two high P_T leptons, two high P_T jets and missing $\not{E}_T$ from neutrinos.

Study of top quark is performed in $e\mu$ channel of collider data collected during run 1a (1992-93 data sample), which corresponds to an integrated luminosity of 13.9 pb^{-1}.

The major background to this channel is the production of $Z \rightarrow \tau\tau \rightarrow e\mu$ and $WW \rightarrow e\mu$. Together with collider data we have 6 different classes of events. We have simulated top quarks events for different top mass (160GeV, 170GeV, 180GeV, 190GeV) using Herwig5.8 Monte Carlo. We have also simulated background events for $Z \rightarrow \tau\tau \rightarrow e\mu$ having ${P^Z}_T \geq$ 25GeV and $WW \rightarrow e\mu$ using Pythia. All these events were energy scale corrected with the package CAFIX 5.0. We will use a small set of these MC events to train a neural net.

3 Training Schedule and Variables

The training of the neural network [3] is done by using a small set of Monte Carlo events. These sets are prepared by using the following cuts

${E_T}^e \geq 15\text{GeV}$

${P_T}^\mu \geq 15\text{GeV}$

Njets ≥ 1 ($\eta \leq 2.5$ and ${E_T}^{jet} \geq 15\text{GeV}$)

plus some loose particle identification cuts. We have used total 2000 training pattern vectors. To make a priori probability same for the signal and the background, we have taken 1000 pattern vector (events) from top 170GeV Monte Carlo as signal and 1000 from the background monte carlo, which consists of 500 $Z \rightarrow \tau\tau \rightarrow e\mu$ and 500 $WW \rightarrow e\mu$.

3.1 Training Schedule

We started training the net with 500 cycles and went up to 10000 cycles. There was no significant effect over this spectrum, but we dia observe that the efficiency of net fell with over training. We reached the conclusion that training with 1000 cycles was optimal. Further we have tried to split the net to over come the problem of a small training sample [4]. We even tried to permute different input variables to get better discrimination. The different variables which we tried are

1) E_T^e, Transverse energy of the leading electron
2) P_T^μ, Transverse momentum of the leading muon
3) E_T^{jet1}, Transverse energy of the leading jet
4) E_T^{jet2}, Transverse energy of next to leading jet
5) H_T, Where H_T is defined as $\sum_{alljets} E_T^{jet}$ with $|\eta^{jet}| \leq 2.5$ and $E_T^{jet} \geq 15GeV$
6) $MASSLL$, Electron Muon invariant mass (MEEM)
7) $\not{E}_T$, Muon corrected missing transverse energy
8) $\not{E}_T^c$, Missing transverse energy (Calorimeter only)
9) $\Delta\phi_{e\mu}$, ϕ difference between electron and muon

10) $\Delta\phi_{jet1,jet2}, \phi$ difference between jet1 and jet2

3.2 Variables

We tried different permutations of the input variables mentioned above. We used the set which gives us an optimal discrimination. This set consists of six variables.

1) E_T^e, Transverse energy of the leading electron

2) E_T^{jet2}, Transverse energy of next to leading jet

3) H_T Defined above

4) $\not{E}_T^c$, Missing transverse energy (Calorimeter only)

5) $MASSLL$, Electron Muon invariant mass (MEEM)

6) $\Delta\phi_{e\mu}, \phi$ difference between electron and muon

We trained the neural net with these six input variables (Fig2,3), five hidden nodes and one output node. During training the weights between different nodes are adjusted according to backpropagation [5] updating. The figure 1 shows the connectivity between different nodes. The thickness of the connecting lines represents the different strengths between different nodes.

4 Analysis and Results

4.1 Analysis

Besides the electron id and muon id cuts, which are the same as that of the conventional analysis. The other cuts are

- $E_T^e \geq 15$GeV, $|\eta| \leq 2.5$
- $P_T^\mu \geq 15$GeV
- Njets$\geq$ 2 with $E_T^{jets} \geq 15$GeV, $|\eta| \leq 2.5$

- Neural Network cut NN$\geq$ 0.85

The electrons are selected with an isolation of $\leq$ 0.1 and a five parameter likelihood[6] of $\leq$ 0.5. Here we are not using the H_T, $\not\!\!E_T^c$ and $E_T^{jets} \geq 20GeV$ cuts. Instead of these three cuts we are using Neural network cut of 0.85. We have chosen this cut for the maximum S/B at a given efficiency (times branching fraction). For this analysis we considered the events having electrons in cc or ec and muons in cf. we are not considering those events having muons in ef. Our trigger + e/μ id efficiencies for $\in$ (cc,cf) is 0.63 and for $\in$ (ec,cf) is 0.46.

4.2 Results

Based on 13.9pb^{-1} of data, the expected top and background yields are shown in table 1. After all cuts, 1 event remains (Fig.6), with an expected background of $0.038 \pm 0.008(stat.)$. The distributions for various top masses and two backgrounds after passing through the neural net are shown in Fig.4,5. Figures 7,8,9,10 show the one and two dimensional distributions of various variables before and after the neural cut for Herwig $t\bar{t}$ Monte Carlo and $Z \rightarrow \tau\tau \rightarrow e\mu$ Monte Carlo.

5 Conclusions

We have described an analysis method looking for $t\bar{t} \rightarrow e\mu$ using neural networks. We observe 1 event in data with a background of $0.038 \pm 0.008(stat.)$. The S/B $\sim$ 8 for a top mass of 170 GeV. The efficiency (times branching fraction) is $0.378 \pm 0.007(stat.)$ for top a mass of 170 GeV. We are applying this analysis to run 1b and 1c also. We will also estimate other backgrounds like

Samples	Expected Yield in $13.9pb^{-1}$
Top160	0.40± 0.008
Top170	0.31± 0.006
Top180	0.24± 0.004
Top190	0.20± 0.003
WW	0.010±0.003
$Z\tau\tau$	0.028±0.006
Total Bkg.	0.038±0.008
Data (1a)	1

Table 1: Expected top, background and data events passing all the cuts. Errors are statistical only

fakes etc. using this analysis.

6 Acknowledgements

We thank Jim Cochran (Chief), Pushpa Bhat, Harrison Prosper and J.M.Kohli for useful discussions.

References

1. $D\emptyset$ collaboration, S.Abachi et al. ,PRL (74,2632 1995), Observation of top quark.
2. Internal $D\emptyset$ note no 1606, Harrison Prosper, Some mathematical comments on feed forward neural nets.
3. C.Peterson and T.Rognvaldsson, An introduction to Artificial Neural Networks, Lectures given at the 1991 CERN school of computing. Lund Preprint LU TP 91-23.
4. Internal $D\emptyset$ note no 2230, Beri,Prosper,Bhat, Searching for high mass top$\rightarrow e\mu$ using neural nets.
5. R. Beale and T. Jackson NEURAL COMPUTING : AN INTRODUCTION (Adam Hilger, New York 1990).
6. Internal $D\emptyset$ note no ????, Heintz,Narain,Chopra,Adam, A likelihood test for electron ID.

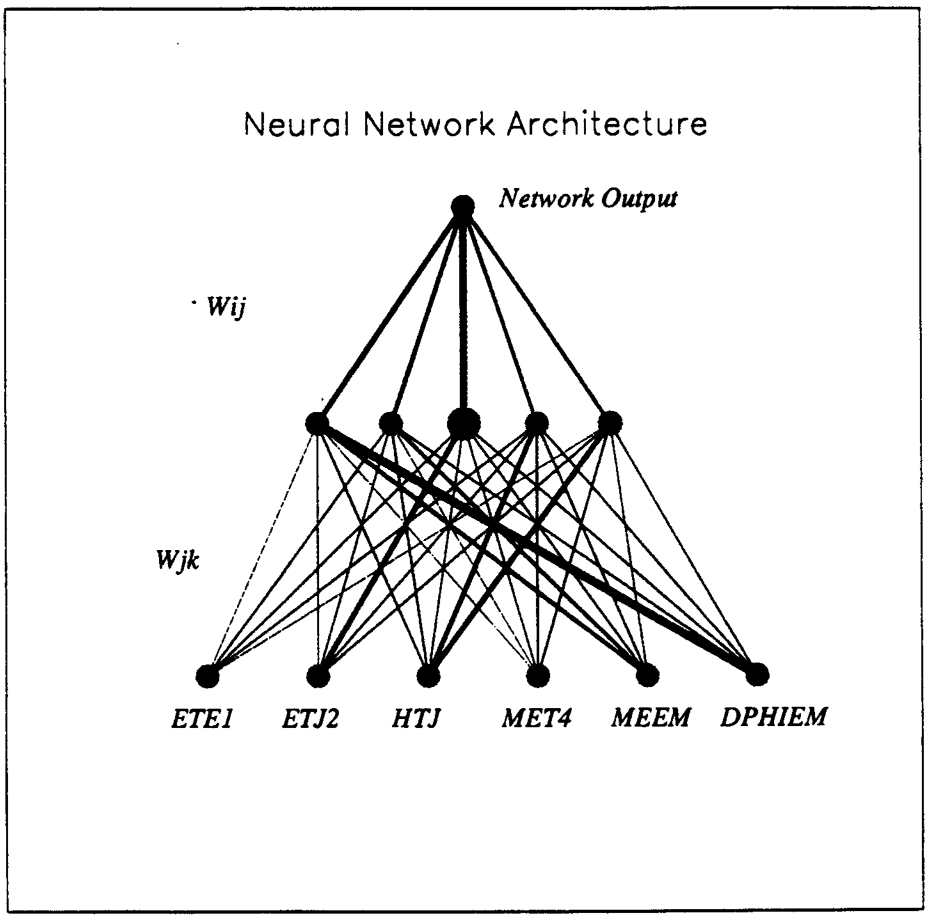

Figure 1: Neural Net Architecture showing different strengths between different nodes.

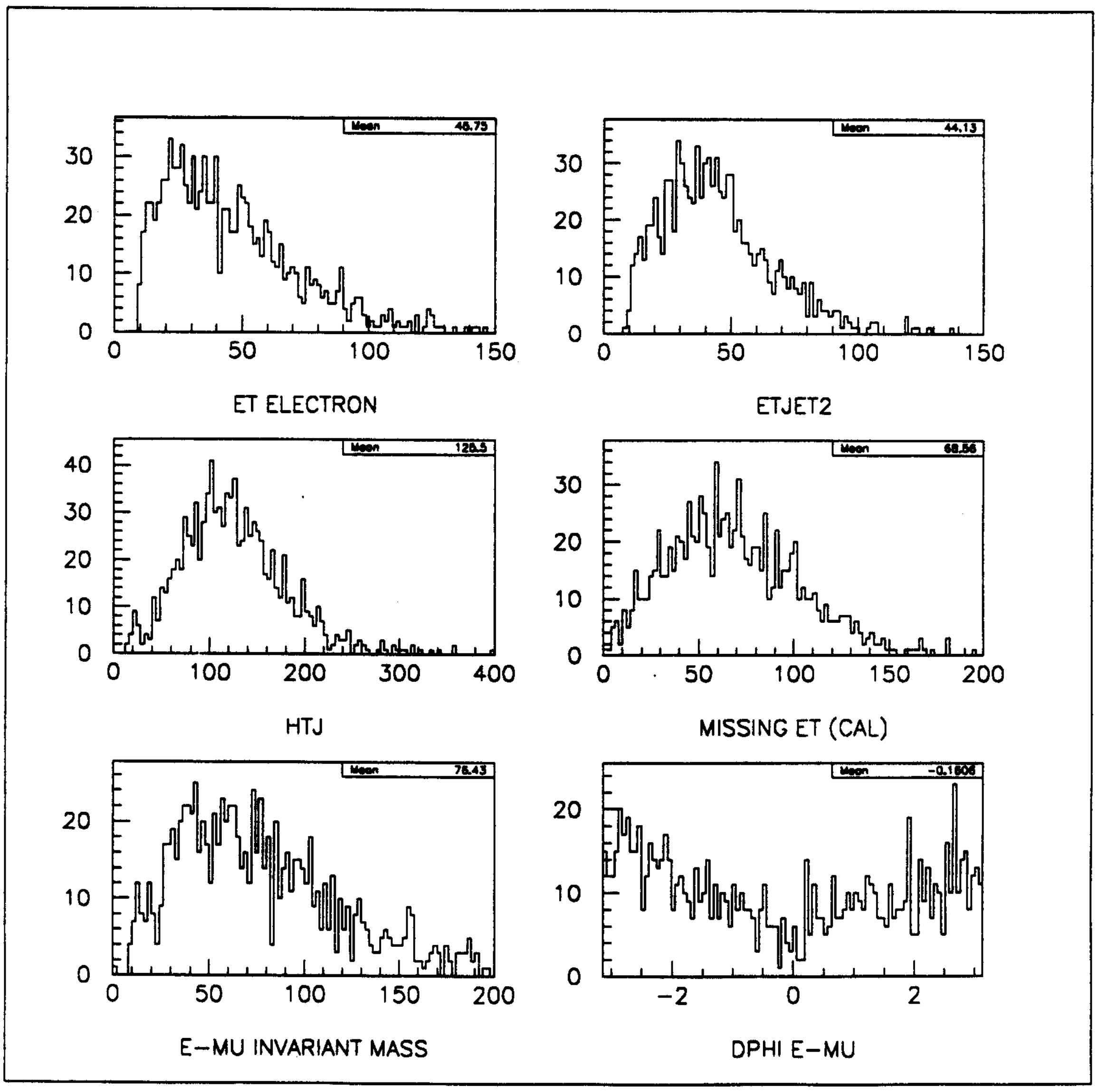

Figure 2: Distributions of six signal variables having 1000 patterns These are the inputs to neural net as signal.

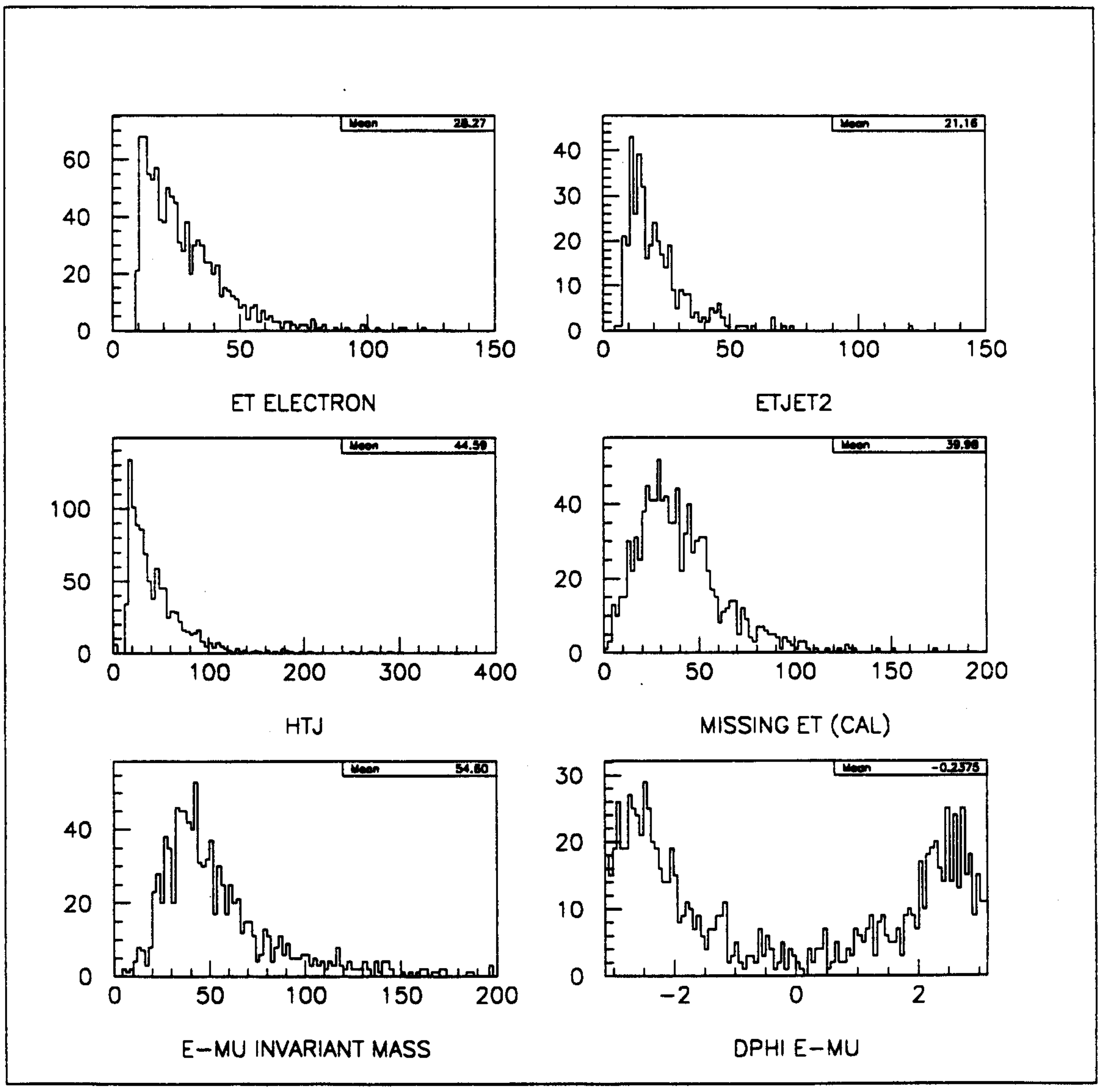

Figure 3: Distributions of six background variables having 1000 patterns. These are the inputs to neural net as background.

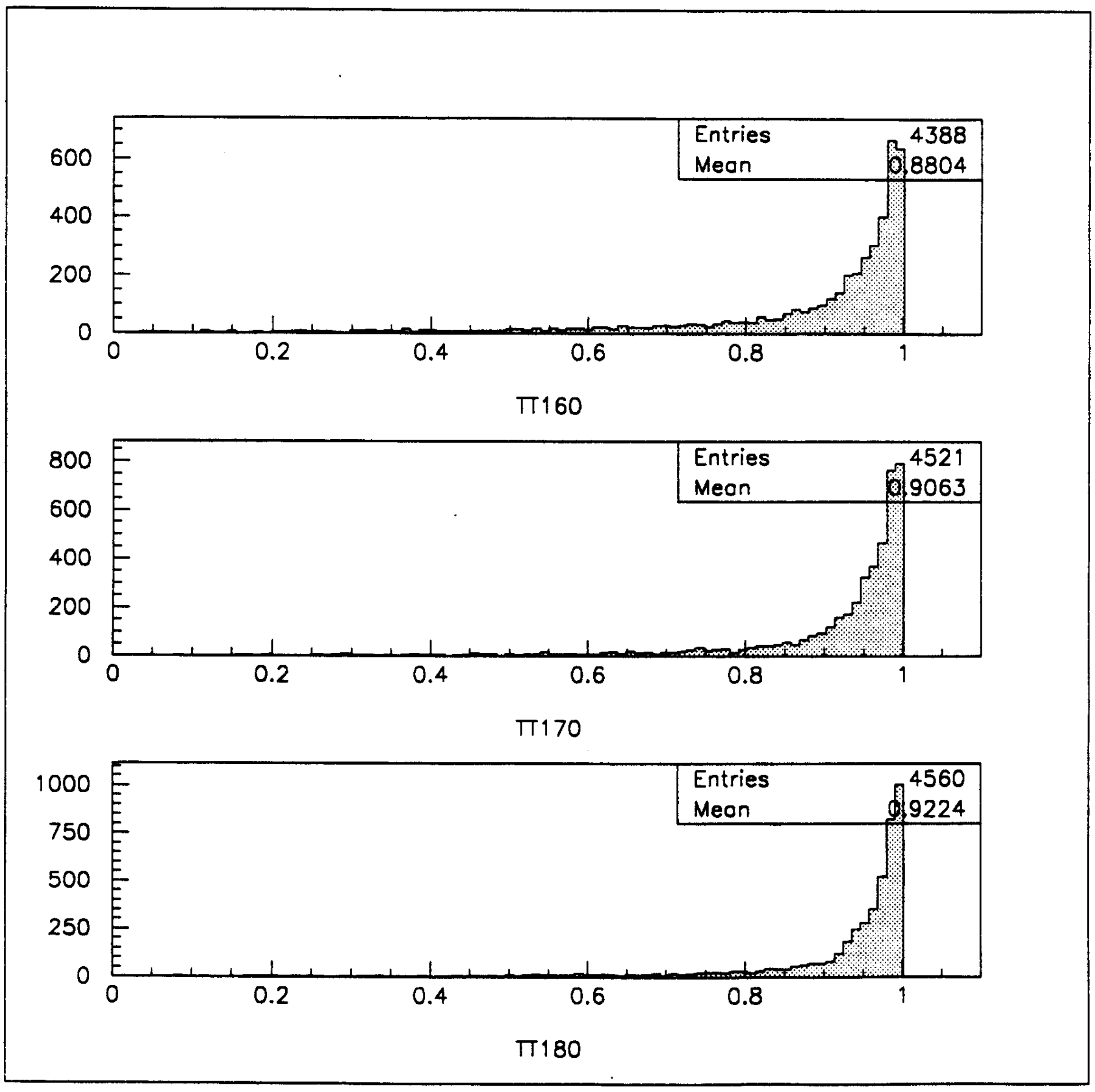

Figure 4: Distributions of neural network output for top160, top170 and top180

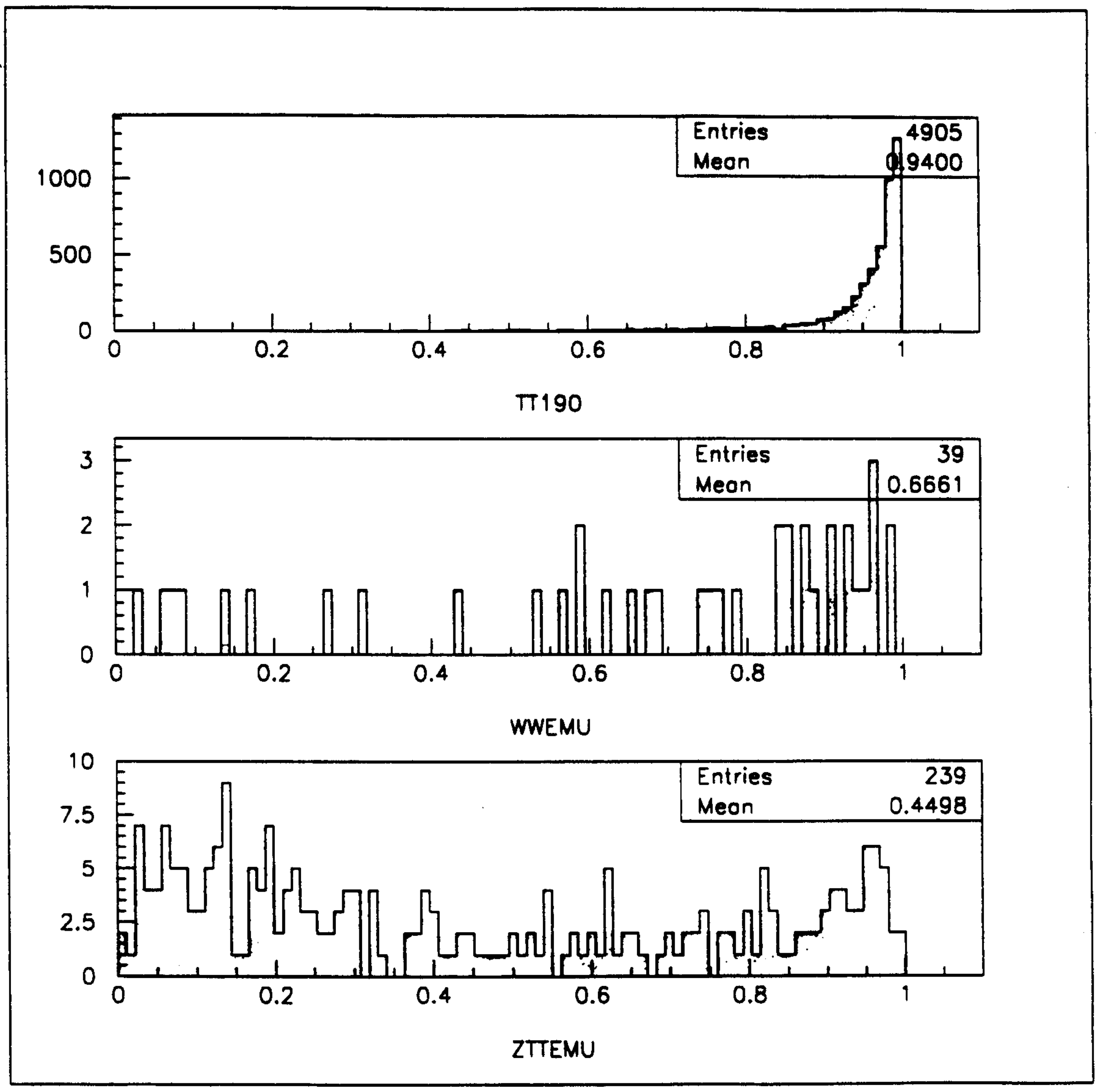

Figure 5: Distributions of neural network output for top190, $WW \to e\mu$ and $Z \to \tau\tau \to e\mu$

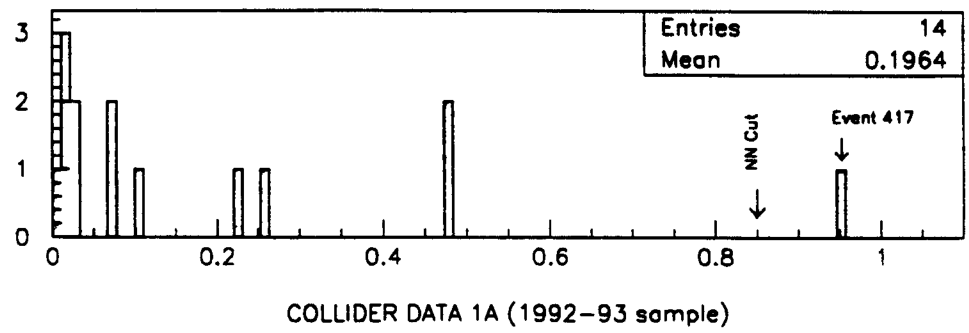

Figure 6: Distributions of neural network output for collider data, Only one event survives the final neural cut.

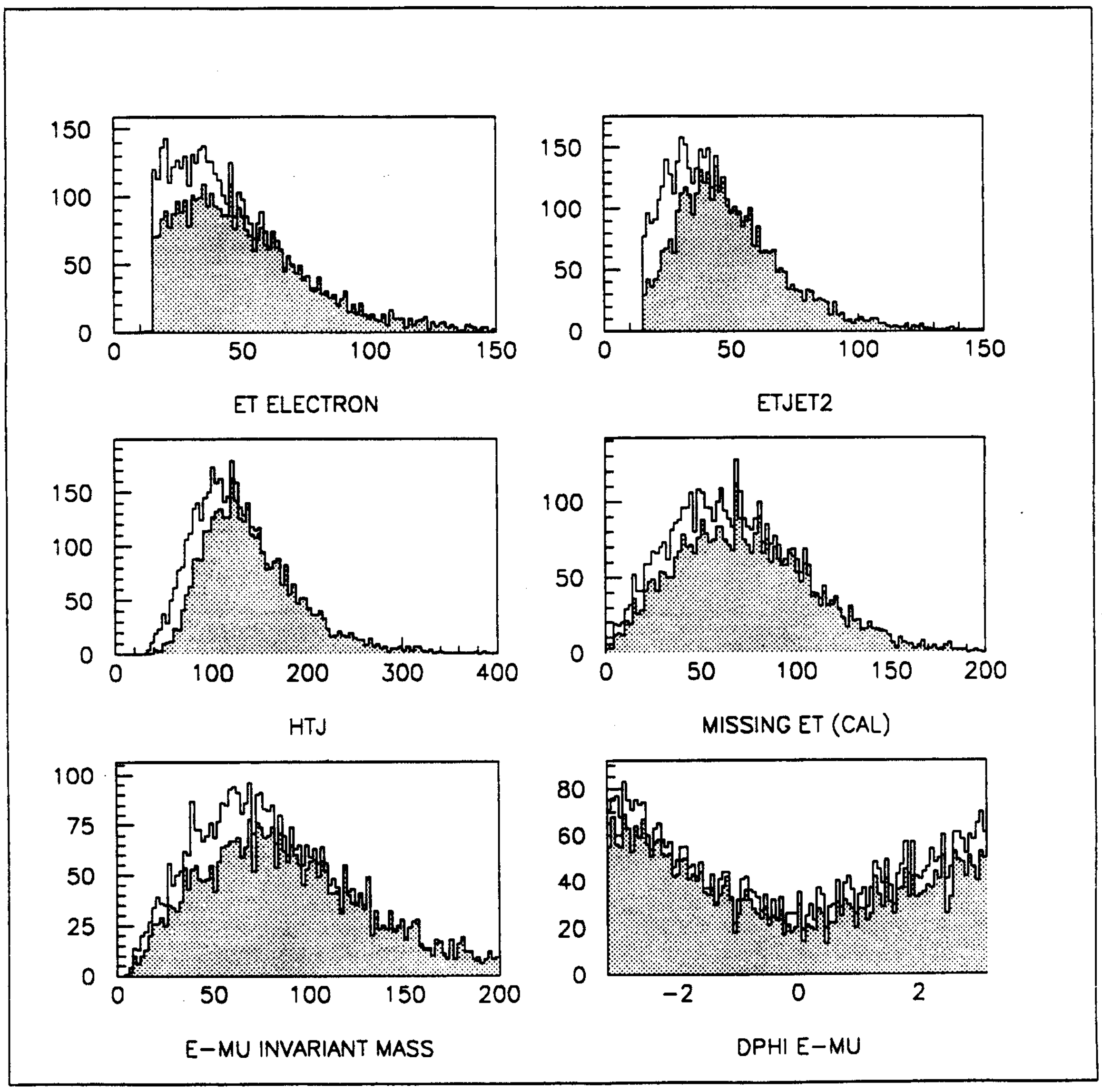

Figure 7: Distributions of six variables for top170 before and after (shaded histograms) neural cut.

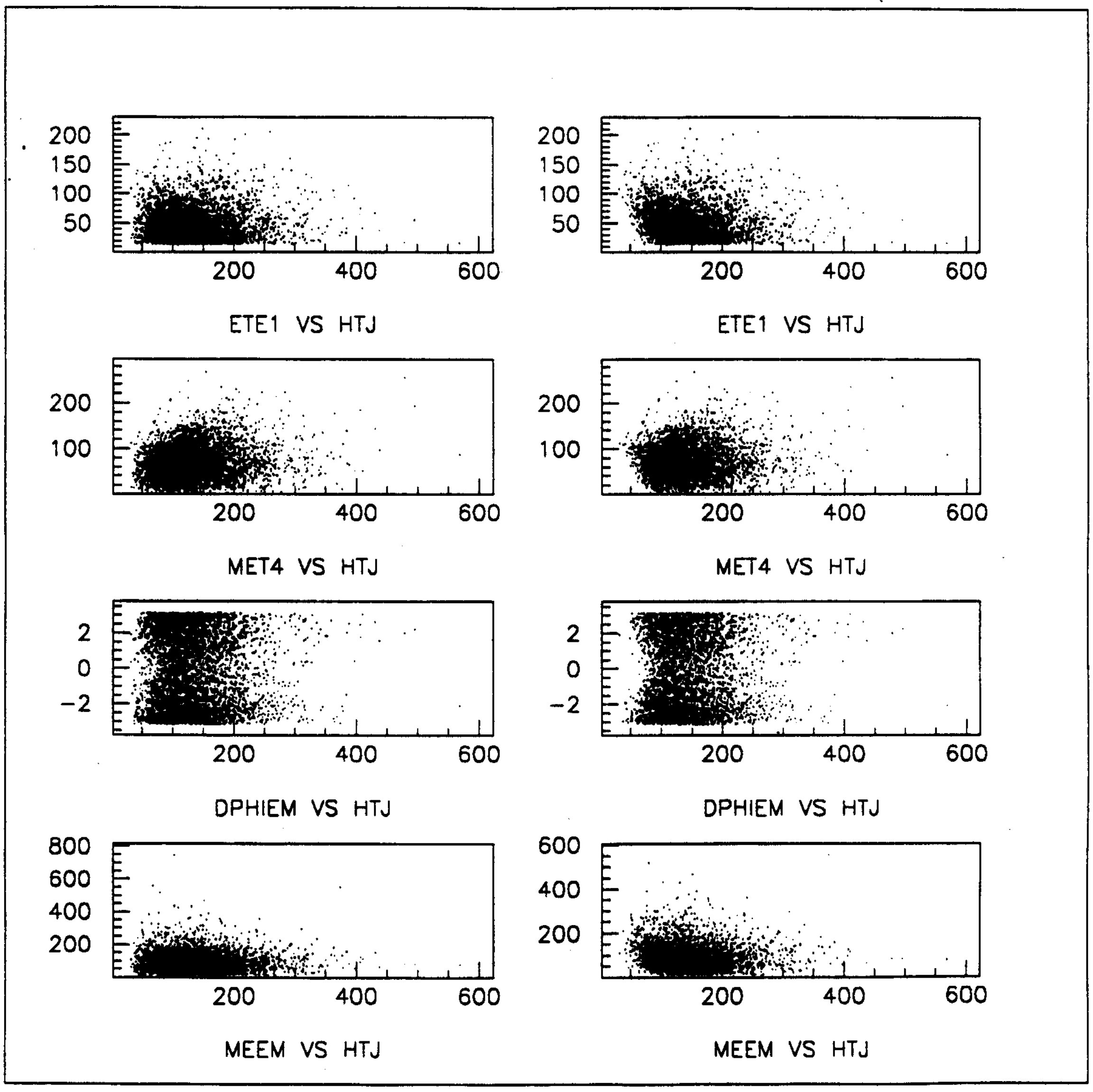

Figure 8: Scatter plots for top170 before (left) and after (right) NN cut

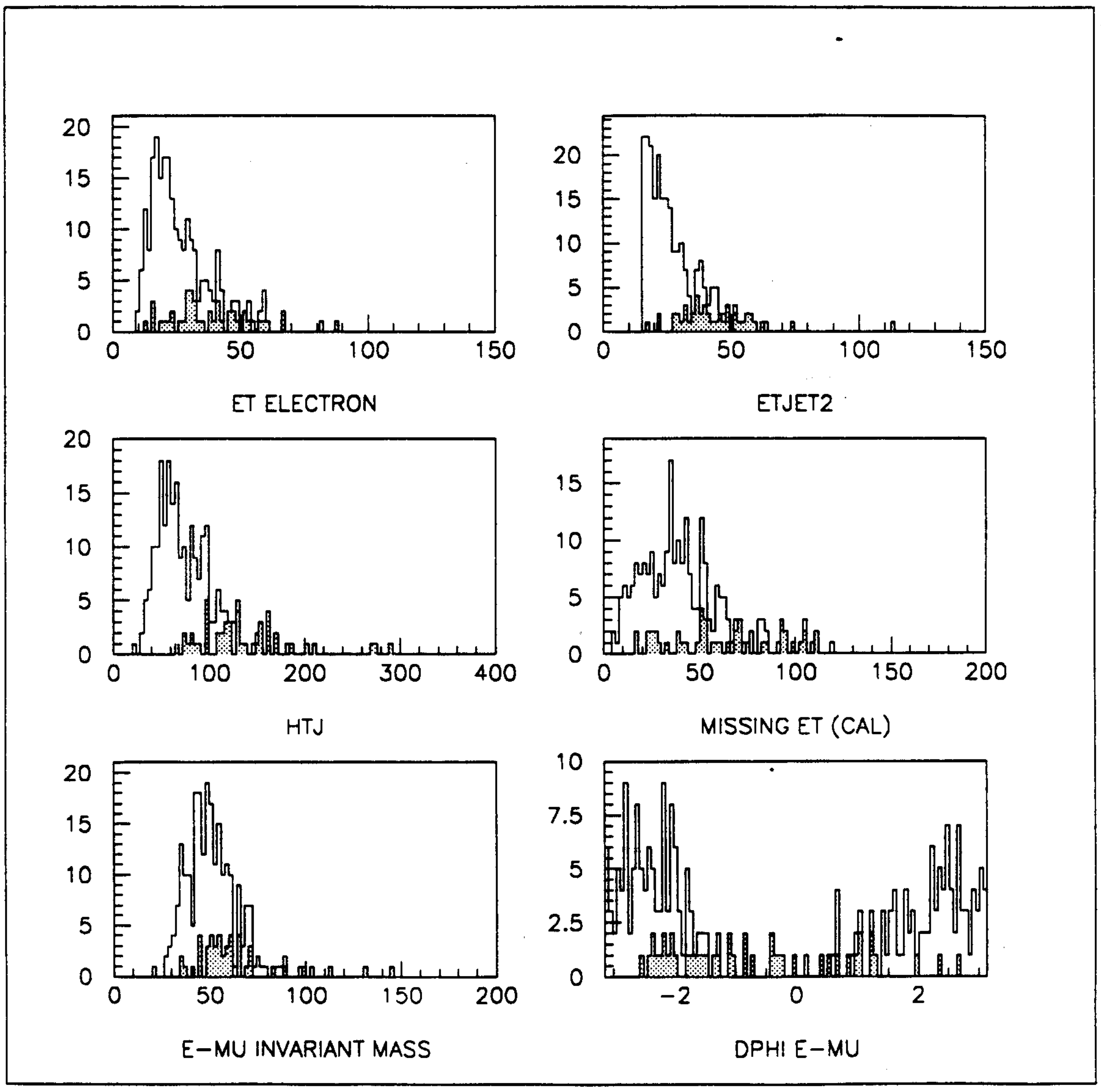

Figure 9: Distributions of six variables for $Z \to \tau\tau \to e\mu$ before and after (shaded histograms) neural cut.

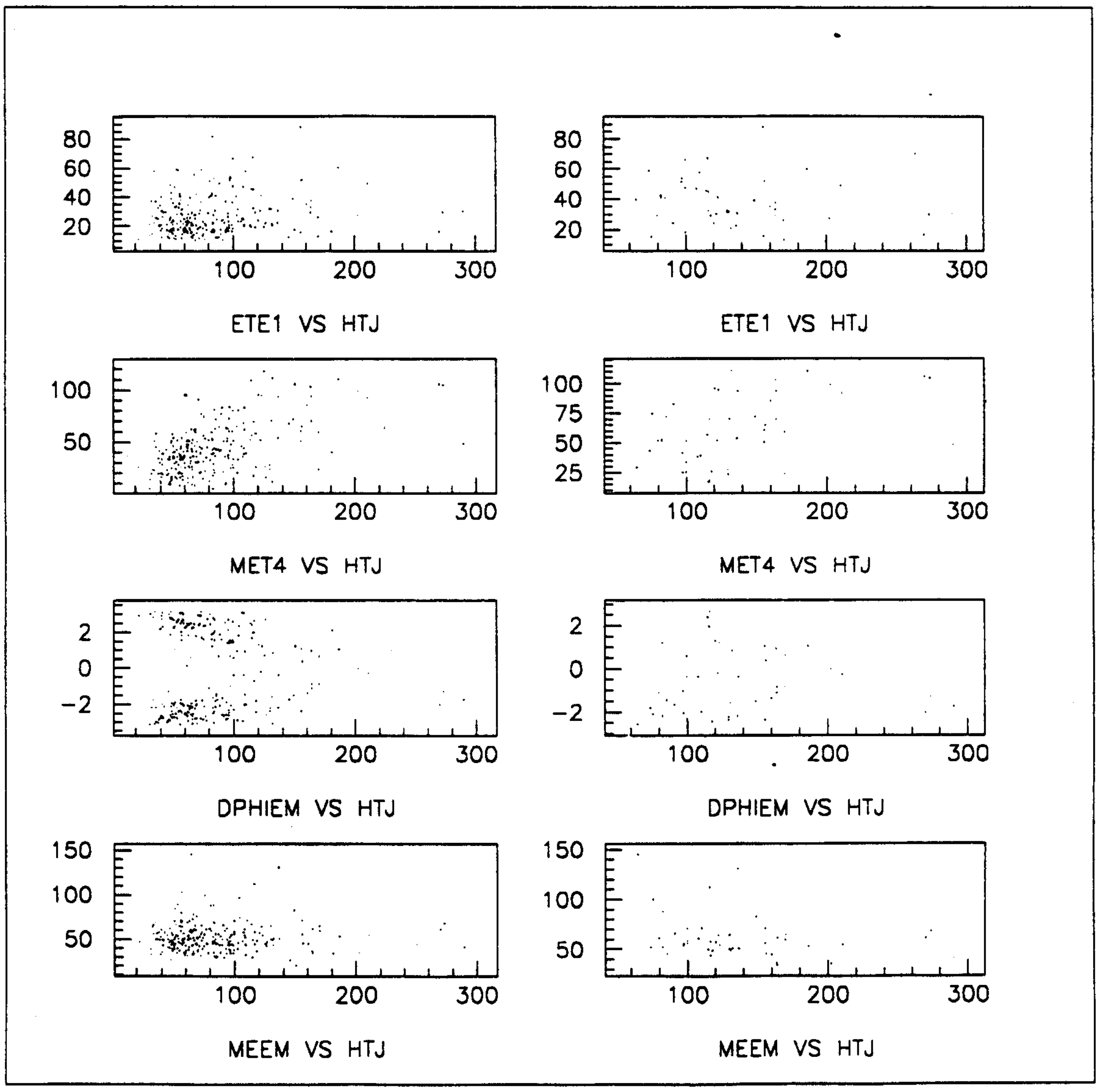

Figure 10: Scatter plots for $Z \to \tau\tau \to e\mu$ before (left) and after (right) NN cut