EUROPEAN ORGANIZATION FOR NUCLEAR
RESEARCH

Delphi 94-xxx Phys

November 16, 1994

Analysis of τ 1-prong hadronic
inclusive branching ratio using
neural networks

F. Matorras, A. Ruiz (Universidad de Cantabria)

Abstract

A neural net has been used to classify the 1-prong τ decays. The net is able to separate hadronic decays from leptonic with 90% efficiency and 93% purity. Applied to 92 data the τ inclusive 1-prong hadronic branching ratio has been measured to be

$$B_{1h} = 0.5050 \pm 0.0032^{+0.0046}_{-0.0031}$$

1 Introduction

The $\tau^+\tau^-$ channel is one of the most interesting in LEP experiments. Its analysis is one of the most effective in the determination of the Z^0 electroweak couplings. The τ has a short lifetime and decays close to the interaction vertex, allowing to obtain physics information from its decay products. In particular, the hadronic τ decays amounts to about 65% of the total; in those decays only one neutrino is missing, so the hadronic channels provide interesting and sensible information about the τ itself.

The identification of τ decays is a difficult task in a detector from LEP as it is the case of DELPHI and involves all subdetectors. Many of the hadronic channels contain π^0 particles decaying into two photons; so, the granularity of the electromagnetic calorimeters is essential to separate them from photons and to distinguish hadronic channels from electron ones. Also, a good performance of the hadronic calorimeters is needed in order to separate the hadronic channels from the leptonic ones.

Recently, the Neural Network technique [1] has proved to give very good results classifying different channels by optimising the global information of the different subdetectors.

We have applied this technique to separate τ hadronic from leptonic decays, in particular considering 1-prong decays. To do that we have preselected $\tau^+\tau^-$ events which decay to one prong and we have trained a feed-forward perceptron neural network to classify them as hadronic or leptonic.

We intend to show how the use of this technique can be efficient separating the decay channels of the τ

In particular, the method has been applied to the data collected in the 1992 run with the DELPHI experiment. From that we obtained the branching ratio of the τ to 1-prong hadronic.

The second section is dedicated to the method of preselection of the $\tau^+\tau^-$ channel. The third section show the characteristics of the network and the training. The fourth section explains how the 1-prong hadronic channel is selected. The fifth section is dedicated to the calculation of the 1-prong hadronic Branching Ratio and the next one to the analysis of the systematic errors. In section 7 the conclusions are presented.

2 $e^+e^- \rightarrow \tau^+\tau^-$ preselection

The initial sample of $\tau^+\tau^-$ is selected with the standard criteria used in team3 (described in [2]). The cuts applied are summarized here:

- $2 \leq N_{trk} \leq 6$
- $\theta_{iso} \geq 160^\circ$
- $E_{vis} > 8 \text{ GeV}$
- $P_{rad} < 1.0$
- $E_{rad} < 1.0$
- $|z_1| < 4.5 \text{ cm}$ and $|z_2| < 4.5 \text{ cm}$
- $|r_1| < 1.5 \text{ cm}$ and $|r_2| < 1.5 \text{ cm}$

for two prong events

- $\theta_{acol} > 0.5^\circ$
- $|z_1 - z_2| < 3 \text{ cm}$
- $|\vec{P}_T| > 0.4 \text{ GeV}$

Where N_{trk} is the charged multiplicity, θ_{iso} the isolation angle (minimum angle between tracks in opposite hemisphere), E_{vis} the total energy measured as charged energy or neutral electromagnetic energy, $P_{rad} = \sqrt{p_1^2 + p_2^2}/E_{beam}$, with p_1 and p_2 the momenta of the most energetic tracks in each hemisphere, $E_{rad} = \sqrt{e_1^2 + e_2^2}/E_{beam}$ with e_1 and e_2 the electromagnetic energy measured in a cone of 30° around the track, z and r refer to the coordinates of the point of closest approach of the track extrapolation to the beam spot, θ_{acol} is the acollinearity and $|\vec{P}_T|$ the transverse momentum.

The event is divided into two hemispheres, according to a plane perpendicular to the thrust axis, that are treated independently in the following analysis.

Only one prong decays are retained. We thus require to have only one track in the hemisphere with associated hits in the vertex detector, avoiding in that way losing 1-prong events seen as multiprongs in the TPC, due to photon conversions.

3 The Network Structure

We have used a feed-forward neural network [3] with one input layer, one hidden layer and one output layer.

The input layer has 12 neurons, each one corresponding to a tau physical variable, which have been chosen in order to facilitate the network training. They are essentially the same variables used in the standard method applied to measure the polarization of the τ through its hadronic decays [6] The variables are the next ones:

1. **MOM**: The momentum of the charged track.
2. **THE**: The polar angle of this track
3. **PHIX**: The azimuthal angle of this track extrapolated to the HPC entry, module 15 (to analyze HPC cracks).
4. **MINV**: The invariant mass of all the particles seen in the τ hemisphere
5. **EHPC**: The energy deposited by the track in the first four layers of the HPC.
6. **EEM**: The electromagnetic energy observed in a cone of 30 degrees around the track
7. **EH1**: The hadronic energy deposited in the first layer of the hadronic calorimeter.
8. **EH2**: The hadronic energy deposited in the second layer of the hadronic calorimeter.
9. **EH3**: The hadronic energy deposited in the third layer of the hadronic calorimeter.
10. **EH4**: The hadronic energy deposited in the fourth layer of the hadronic calorimeter.
11. **NMUG** : The number of muon chambers hit, including as μ -chamber the last layer of the hadronic calorimeter if the polar angle is such that the track go to a region without them.
12. **NLHA** : The number of hadronic calorimeter layers hit by the track

All the input variables have been normalised to the range $[-1,1]$.

In the figures 1, 2, 3 it is shown the distribution of some of those variables for the hadronic and leptonic channels, from Monte Carlo simulation. It is shown the discriminant power to separate hadronic from leptonic channels.

Other variables have been introduced in order to increase the acceptance range; for example, the variable PHIX will inform if the track crosses a dead zone of the HPC.

The hidden layer consists of 12 neurons. We have seen that the separation power does not improve when we increase the number of neurons but the CPU for training grows. In fact we have observed that new neurons added in the hidden layer have weights an order of magnitude smaller than the others showing that their contribution is negligible.

Also, a neural network with two hidden layers was tested, with similar conclusions. The output layer consists on one neuron, whose assigned value is -1 in the case of leptonic decay and +1 in the case of hadronic one.

We have also analysed the case of two outputs, showing the consistency of the network, since the identification as hadron was equivalent to non-identification as lepton.

3.1 Training

The network has been trained using $e^+e^- \rightarrow \tau^+\tau^-$ events generated by the program KORALZ [4], and simulated with DELSIM [7]. About 4000 leptonic and 4000 hadronic τ decays were selected as explained above.

We have analysed the optimal parameters of the network in order to gain in efficiency and CPU time. The activation function chosen is $\tanh(x)$ and the cost function is the standard euclidean square distance from target. The method to optimise the weights is the gradient descent and the learning strength parameter η , which corresponds to the distance to move along the gradient of the cost function has been chosen 0.4, which provides us with a fast convergence [1].

In order to avoid oscillations during the learning phase we used a smoothing term which contains the momentum parameter α . This value is multiplied by the last weight increase and the result is added to the other term. The value chosen for this parameter is 0.9.

So, in each step of the training the variation of the weights is given by:

$$\Delta_n w_{ji} = -\eta \frac{\delta E_n}{\delta w_{ji}} + \alpha \Delta_{n-1} w_{ji} \quad (1)$$

where E is the cost function which gives the squared deviation of the output values from the desired ones, summed over the training sample and w_{ji} are the weights between neuron i and j.

The updating of the weights is made after reading all the input events. As we see in figure 4 the convergence is reached after about 300 epochs.

The test of the training has been made using about 80000 hadronic and 65000 leptonic τ decays.

The efficiency versus purity obtained for the training and testing are indicated in the figure 5, in which it is possible to observe the effect of overtraining.

In order to know the relative importance of the weights, we have calculated for each input variable the "effective weight" defined as the sum of the input to hidden weights multiplied by the hidden to output weights over all hidden nodes, which in some way show the importance of that variable. From the figure 6 it could seem that there are some variables with negligible sensitivity, but they provide very useful information for a small set of events (for instance, the information of the extrapolated azimuthal angle in the HPC entry).

The results of the network on the test sample are shown in the figure 7. A very good separation is achieved (for instance for a purity of 94 % the efficiency is close to 90 %). It can be compared with the standard cuts method [6] that has for similar purity an efficiency below 80% for tracks not pointing to an HPC crack ($\pm 1^\circ$). This is due to the fact that the network takes into account the correlation between the input variables and extracts the maximum of information. The method used to separate the two types of decay in real data is to chose a cut on the target value, which must be near zero, where the distribution is flat. There we have a good efficiency and purity, and small systematic errors due to the small discrepancies in the target values between the MonteCarlo and real data.

4 Selection of 1-prong hadronic inclusive channel

The optimal weights obtained in the training of the network have been used to classify real data, preselected with the criteria of the Chapter 2.

The real data corresponds to the 1992 DELPHI run, selecting only the cases in which the detector has operated correctly in all the subdetectors which contribute to the input variables chosen. The preselected one prong τ events amount to 26207. The external background from bhabha events, dimuon events and hadronic decays of the Z^0 , as well as two-photon events, after the preselection is less than 1% of the total.

We have generated Monte Carlo background events and they have been classified by the network in order to subtract their contribution from the data.

As we intend to classify 1-prong hadronic decays of the τ , we have also taking into account the contribution of hadronic events with 3 or more tracks which have been preselected. They act as a background on the classified events and must be also subtracted.

In order to see the general agreement between the MonteCarlo events and the real data, which is very important in order to use the test results to compute the efficiency-purity as a function of the target value cut, we have represented in the figures 8, 9, 10 the distribution of the input variables normalized to 1 for the two samples. The MonteCarlo does not include the external background.

Small corrections were applied to energies to improve MonteCarlo-data agreement. To avoid very low energetic showers seen in data and not in MonteCarlo, the low energies (EEM less than 0.4 or EHi less than 0.3) were reset to 0. The corresponding changes produced in the classification are used to evaluate the systematic errors.

5 Branching ratios calculation

The branching ratio of τ decay to 1-prong hadronic channel can be obtained from the next expression:

$$\begin{aligned} N_h^o &= 2N_{\tau^+\tau^-}(B_{1h}\epsilon_{1h}^s\epsilon_{1h}^o + B_l\epsilon_l^s\epsilon_l^o + B_{>1h}\epsilon_{>1h}^s\epsilon_{>1h}^o + \sum_i \delta_i\epsilon_i^s\epsilon_i^o) \\ N_{tot} &= 2N_{\tau^+\tau^-}(B_{1h}\epsilon_{1h}^s + B_l\epsilon_l^s + B_{>1h}\epsilon_{>1h}^s + \sum_i \delta_i\epsilon_i^s) \end{aligned}$$

and from there

$$B_{1h} = \frac{\frac{N_h^o}{N_{tot}}[B_l\epsilon_l^s + B_{>1h}\epsilon_{>1h}^s + \sum_i \delta_i\epsilon_i^s] - [B_l\epsilon_l^s\epsilon_l^o + B_{>1h}\epsilon_{>1h}^s\epsilon_{>1h}^o + \sum_i \delta_i\epsilon_i^s\epsilon_i^o]}{(\epsilon_{1h}^s\epsilon_{1h}^o - \epsilon_l^s\epsilon_l^o) - \frac{N_h^o}{N_{tot}}(\epsilon_{1h}^s - \epsilon_l^s)}$$

where N_h^o refers to the total number of events classified in data as hadronic from N_{tot} preselected τ ; the efficiencies are given by ϵ_a^b , where the index a refers to the decay channels (l, leptonic; 1h, 1-prong hadronic; $>1h$, multi-prong; i, each external background channel) and the index b refers to preselection (s) or classification (o) after the neural net analysis. The classification efficiencies can be obtained from the Monte Carlo test sample after the neural net analysis. The preselection ones are obtained from the original MonteCarlo sample after doing the cuts corresponding to the preselection as explained in Chapter 2.

The δ_i are the fraction of external background in each channel corresponding to the total number of τ generated. Finally, the B_b are the branching ratios of the τ decay which satisfy the relation:

$$\begin{aligned} B_{1h} + B_l + B_{>1h} &= 1 \\ B_1 &= B_{1h} + B_l \end{aligned}$$

The error on the 1-prong hadronic branching ratio can be calculated from the expression of B_{1h} , by propagation. The errors on the classification efficiencies have a statistical part which can be obtained from the corresponding binomial distributions; the systematic errors are explained below.

The topological branching ratios are taken from [5]:

$$\begin{aligned} B_1 &= 85.07 \pm 0.37\% \\ B_{>1} &= 14.93 \pm 0.40\% \end{aligned}$$

From the preselection cuts we get the next results of the preselection efficiencies and external background:

$$\begin{aligned}
\epsilon_l^s &= 0.510 \pm 0.001 \\
\epsilon_{1h}^s &= 0.455 \pm 0.001 \\
\epsilon_{>1h}^s &= 0.0212 \pm 0.001 \\
\delta_{qq}\epsilon_{qq}^s &= 0.00068 \pm 0.00012 \\
\delta_{\mu\mu}\epsilon_{\mu\mu}^s &= 0.0016 \pm 0.00014 \\
\delta_{ee}\epsilon_{ee}^s &= 0.0040 \pm 0.00016 \\
\delta_{\gamma\gamma\rightarrow eeqq}\epsilon_{\gamma\gamma\rightarrow eeqq}^s &\leq 0.00001 \\
\delta_{\gamma\gamma\rightarrow ee\tau\tau}\epsilon_{\gamma\gamma\rightarrow ee\tau\tau}^s &\leq 0.0001 \\
\delta_{\gamma\gamma\rightarrow ee\mu\mu}\epsilon_{\gamma\gamma\rightarrow ee\mu\mu}^s &= 0.0006 \pm 0.0001 \\
\delta_{\gamma\gamma\rightarrow eeee}\epsilon_{\gamma\gamma\rightarrow eeee}^s &= 0.001 \pm 0.00015
\end{aligned}$$

Using those results and taking 0. as the reference output value to classify the events we got the next value for B_{1h} :

$$B_{1h} = 0.5050 \pm 0.0040$$

where the error includes the statistical errors on data and MonteCarlo as given above.

Another possibility to obtain B_{1h} is from a fit of the output-neuron values distribution in data (figure 11) to the expected ones using the efficiencies given by the network on the MonteCarlo sample, taking as free parameter the 1-prong hadronic branching ratio gives the result:

$$B_{1h} = 0.5064 \pm 0.0031$$

with a $\chi^2/99 = 1.4$. The quoted error is the statistical one on data. The Montecarlo statistical error is the same as above.

The agreement between the two values shows the consistency of the method.

6 Systematic errors

The main systematic errors are related to the discrepancies between the MonteCarlo (used to get the classification efficiencies) and the real data in the input variables, specially those which are more effective in the separation of the decay channels.

The small discrepancies are related to the calibration of the calorimeters and the punch-through simulation in the hadronic calorimeters. So, we have used data to correct the MonteCarlo before training the network.

The observed variation on the network output induced by the uncertainties on those corrections contributes to the systematical errors.

In the figure 12 it is shown the value of the 1-prong hadronic branching ratio as a function of the chosen cut on the target value. As it is seen there is a small dependence, which is mainly flat in the central zone, where the separation between the classes is good. We have taken 0. as the reference value to measure the branching ratio, the differences obtained when correcting MonteCarlo, according to the target differences, being compatible with the errors.

Other sources of systematic errors are related to the topological branching ratios introduced in the MonteCarlo. We have observed that this is negligible compared to the other ones. The same is true for the systematic errors associated to the variation of the parameters of the network.

When we add the systematic errors the final value we got for the Branching ratio is:

$$B_{1h} = 0.5050 \pm 0.0032^{+0.0046}_{-0.0031}$$

7 Conclusions

A feed-forward neural network has been used to show the effectiveness of the technique to separate between different τ decays. In particular it has been used to analyse the branching ratio of τ to 1-prong hadronic in the DELPHI experiment at LEP.

The neural network consists of 1 input layer with 12 neurons, one hidden layer with 12 neurons and an output layer with a neuron. The efficiency-purity of the selection of hadronic decays versus leptonic ones is in the order of 90%-95%, much better than what is obtained by the standard method of linear cuts.

The Branching ratio to 1 hadronic-prong has been measured for the data corresponding to the 1992 DELPHI run and amounts to:

$$B_{1h} = 0.5050 \pm 0.0032^{+0.0046}_{-0.0031}$$

We think that the method can be extended to separate the exclusive channels of the τ decay. As it has been shown in the previous section, the systematical errors can be controlled and are similar to those obtained by the standard cuts method.

References

- [1] D.E.Rumelhart et al. in "Parallel Distributed Processing", MIT Press (1986), p.318
B.Denby FERMILAB-Conf-92/121-E (1992)
- [2] F.Matorras, DELPHI 92-124 PHYS 225 (1992)
- [3] R.Odorico, NEURAL program, private communication.
- [4] S. Jadach and Z. Was, Comp. Phys. Commun. 36 (1985) 191.
- [5] DELPHI Coll. Results presented to ASPEN-94 Conf. (1994)
- [6] F.Matorras, DELPHI note in preparation
- [7] DELSIM user guide DELPHI 89-67,89-68

Figure 1: MonteCarlo distributions of the input variables. The hatched area corresponds to the hadronic decays. The other to the leptonic ones

Figure 2: MonteCarlo distributions of the input variables. The hatched area corresponds to the hadronic decays. The other to the leptonic ones

Figure 3: MonteCarlo distributions of the input variables. The hatched area corresponds to the hadronic decays. The other to the leptonic ones

Figure 4: The evolution of the cost-function versus the epoch number

Figure 5: The efficiency versus purity for the test sample of MonteCarlo events. The points are the results for the training sample; the overtraining is clearly seen

Figure 6: The effective weights for each one of the input variables

Figure 7: The value of the output neuron in the MonteCarlo test sample, for the two types of events. The hatched area corresponds to the hadronic decays and the other one to the leptonic ones. Also it's shown the data results superposed to the MonteCarlo ones; the global agreement is excellent

Figure 8: The comparison between data and MonteCarlo events for the input variables. Histograms are MonteCarlo events and the dots are data ones.

Figure 9: The comparison between data and MonteCarlo events for the input variables. Histograms are MonteCarlo events and the dots are data ones.

Figure 10: The comparison between data and MonteCarlo events for the input variables. Histograms are MonteCarlo events and the dots are data ones.

Figure 11: The distribution of the output-neuron value for data, superposed to the expected one using the MonteCarlo efficiencies got by the network

Figure 12: The branching ratio of the τ to 1-prong hadronic as a function of the chosen cut in the output neuron value