DELPHI Collaboration

DELPHI 2001-003 PHYS 886
25 January, 2001

Use of a Neural Network for the determination of the $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ production cross-section at 189 GeV.

W. Van den Boeck [1]
(wvdboeck@hep.iihe.ac.be)
C. De Clercq
(declercq@hep.iihe.ac.be)

Inter-University Institute for High Energies (IIHE)
Brussels,Belgium

Abstract

The cross-section for the process $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ has been measured with the data sample taken by DELPHI at an average centre-of-mass energy of 189 GeV, corresponding with an integrated luminosity of $157pb^{-1}$.
Artificial intelligence, in the form of pattern recognition by means of a Feed Forward Neural Network has been used to select the events.
Based on the 1369 events selected as $e^+e^- \to W^+W^- \to q\bar{q}q\bar{q}$ candidates, the cross-section has been measured to be $7.63 \pm 0.21(stat) \pm 0.11(syst)$ pb.

[1] Now working at Alcatel Bell,Antwerp,Belgium

1 Introduction

In this paper a result is given for the value of the cross-section for the decay process of the interaction $e^+e^- \rightarrow W^+W^-$ where both W's decay into hadrons (called fully hadronic final states).
The data used was collected by the DELPHI experiment at a centre-of-mass energy of $188.63 \pm 0.06 GeV$ during the 1998 data taking period.
The total integrated luminosity corresponds to $157 pb^{-1}$ and has been measured by detecting Bhabha processes, which are described very well theoretically. Its systematic error is estimated to be 0.6%.
A detailed description of the DELPHI detector and an overview of its performance are given in ref. [1, 2].
The cross-section which has been determined refers to the processes described by the three doubly resonant tree-level diagrams ('CC03 diagrams' [3]) involving t-channel ν exchange and s-channel γ and Z exchange.
The selection has been done in 2 steps; a loose preselection based on sequential cuts, followed by a Neural Network classification.
The selection performance was expressed in efficiency and purity, selection parameters which where determined from Monte Carlo simulation samples for signal and backgrounds. These were made with the PYTHIA 5.7 [4] event generator followed by the DELSIM [5] full detector simulation program.
The backgrounds which had to be considered after loose precuts where $q\bar{q}(\gamma)$, ZZ and Ze^+e^- events and semileptonic WW events, in which only one W decays hadronically and the other one decays into a lepton and a neutrino [2].
Although the four-fermion final states can also be produced via other diagrams, involving either zero, one or two massive vector bosons, the correction which accounts for the interference between CC03 diagrams and these additional diagrams are negligible at this energy for the fully hadronic final state [6].
The CC03 Standard Model cross-section calculated by the program GENTLE [7] for a centre-of-mass energy of $\sqrt{s} = 188.63 GeV$ and $M_W = 80.39 GeV/c^2$ [8] is $\sigma_{e^+e^- \rightarrow W^+W^-} = 16.65 \pm 0.33 pb$, which gives a value of $\sigma_{e^+e^- \rightarrow W^+W^- \rightarrow q\bar{q}q\bar{q}} = 7.59 \pm 0.15 pb$.

2 Neural Network

2.1 Introduction

Artificial Neural Networks are mathematical models based on some key ingredients from biology and neurofysiology. Whether they are really similar is further a philosophical question [9]. The subclass of Neural Networks used and discussed here are Feed Forward Neural Networks. These are mostly used for pattern recognition and thus classification tasks.

[2] Here we mean a charged lepton and an antineutrino or a charged antilepton and a neutrino

2.2 Building blocks

The basic elements of a Neural Network are the neurons with outputs n_i, which can take values within the interval [0,1].
When a neuron is fed by the output of other neurons n_j through weighted connections (weights w_{ij}), the action of the neuron produces an output n_i :

$$n_i = g(\sum_j w_{ij}.n_j) \tag{1}$$

Here g is a non-linear transfer function, usually a sigmoid, as shown in Fig 1. and defined as:

$$g(x) = \frac{1}{1 + e^{\frac{-(x-0.5)}{C}}} \tag{2}$$

where C is an appropriate scaling constant.

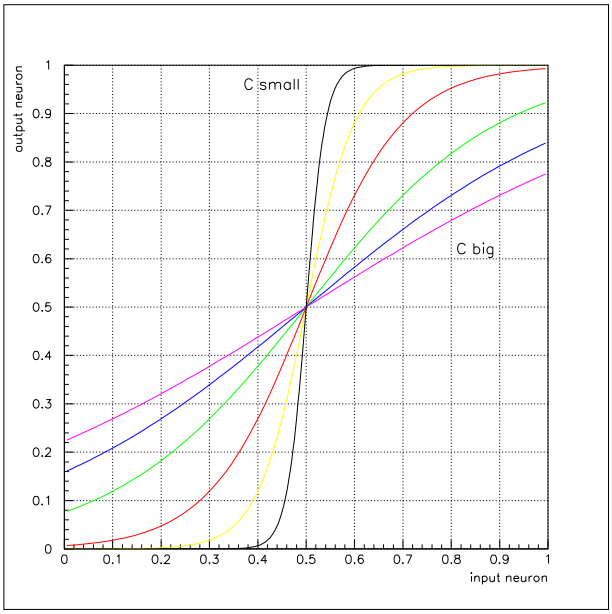

Figure 1: Sigmoid function for different values of C.

2.3 Feed Forward Neural Network

A collection of interconnected neurons is called a neural network. Feed Forward networks are layered and such that signals are processed from a set of input neurons in the bottom to output neurons in the top, layer by layer. In each neuron the local updating rule of equation 1 is used.

2.4 Training & testing

A pattern recognition system such as a neural network can be considered as a two stage device. First there is feature extraction, then classification.
Let's consider a feature vector or pattern: a vector that contains characteristic variables for the different classes that have to be separated. In our case it will be the physical variables which are typical or just not typical for the signal events. The 'summed square error function' is then defined as :

$$E = \frac{1}{2} \sum_{p} \sum_{i} (o_i^p - t_i^p)^2. \tag{3}$$

Where i is running over all the output neurons and p over all training patterns. The quantity t_i^p is the expected (known) network output on output neuron i for pattern p and o_i^p is the corresponding effective network output.

'Training' the network is then defined as minimizing this error function, by changing the interconnecting weights (w_{ij} in equation (1)) in the network in a systematic way. To do this, a wide range of imaginative algorithms ('learning rules') is available [10, 11].

Once this error function is minimized, the network weights are fixed and its performance must be determined by passing through independent samples of all classes of inputs that have to be classified. This is called 'testing'.

3 Selection of the $e^+e^- \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ candidates

3.1 Preselection

After a track selection, a loose sequential cut preselection was applied to reduce the main backgrounds.
The track selection was the following :

- charged particles :
  - track length $> 15cm$
  - $0.4GeV < momentum < 200GeV$
  - $abs(R_{impact}) < 4cm$
  - $abs(z_{impact}) < 10cm$
- neutral particles :
  - total energy $E > 0.4GeV$
  - $\triangle E/E < 1$

The preselection consisted of the following cuts :

- $\frac{\sqrt{s'}}{\sqrt{s}} > 0.80$
  $\sqrt{s}$ is the centre-of-mass energy
  $\sqrt{s'}$ is the effective centre-of-mass energy of the event calculated with the SPRIME package [12]
- $\frac{E_{vis}}{\sqrt{s}} > 0.16$
  E_{vis} is the total visible energy of the event
- $\frac{E_{ch}}{\sqrt{s}} > 0.06$
  E_{ch} is the total energy in the event carried by charged particles
- Number of charged particles ≥ 10
- Number of jets (using LUCLUS [13] with $d_{join} = 4.0 GeV/c^2$) ≥ 3

The preselection efficiency was $97.8 \pm 0.1\%$. In the data 4748 events were selected, while 4706 were expected from Monte Carlo if one considers $q\bar{q}(\gamma)$, ZZ, Ze^+e^-, $q\bar{q}l\nu$ and signal processes. This shows that all important backgrounds have been considered. The remaining background is given in Table 1.

Table 1: Result after preselection.

Type	Efficiency (in %)	cross-section left (pb)
$q\bar{q}l\nu$	40.7	2.94
$q\bar{q}(\gamma)$	18.5	18.4
ZZ	54.4	0.86
Zee	4.1	0.34
Total background	-	22.55
$q\bar{q}q\bar{q}$	97.8	7.42

After these cuts, there was a good agreement between data and Monte Carlo simulation for the variables that fill in the feature vectors for the network, as can be seen in Figs. 2, 3 & 4.

3.2 Neural Network classification

A fully connected Feed Forward neural network was used, consisting of an input layer of 10 neurons, 1 hidden layer of 8 neurons and 1 output neuron.

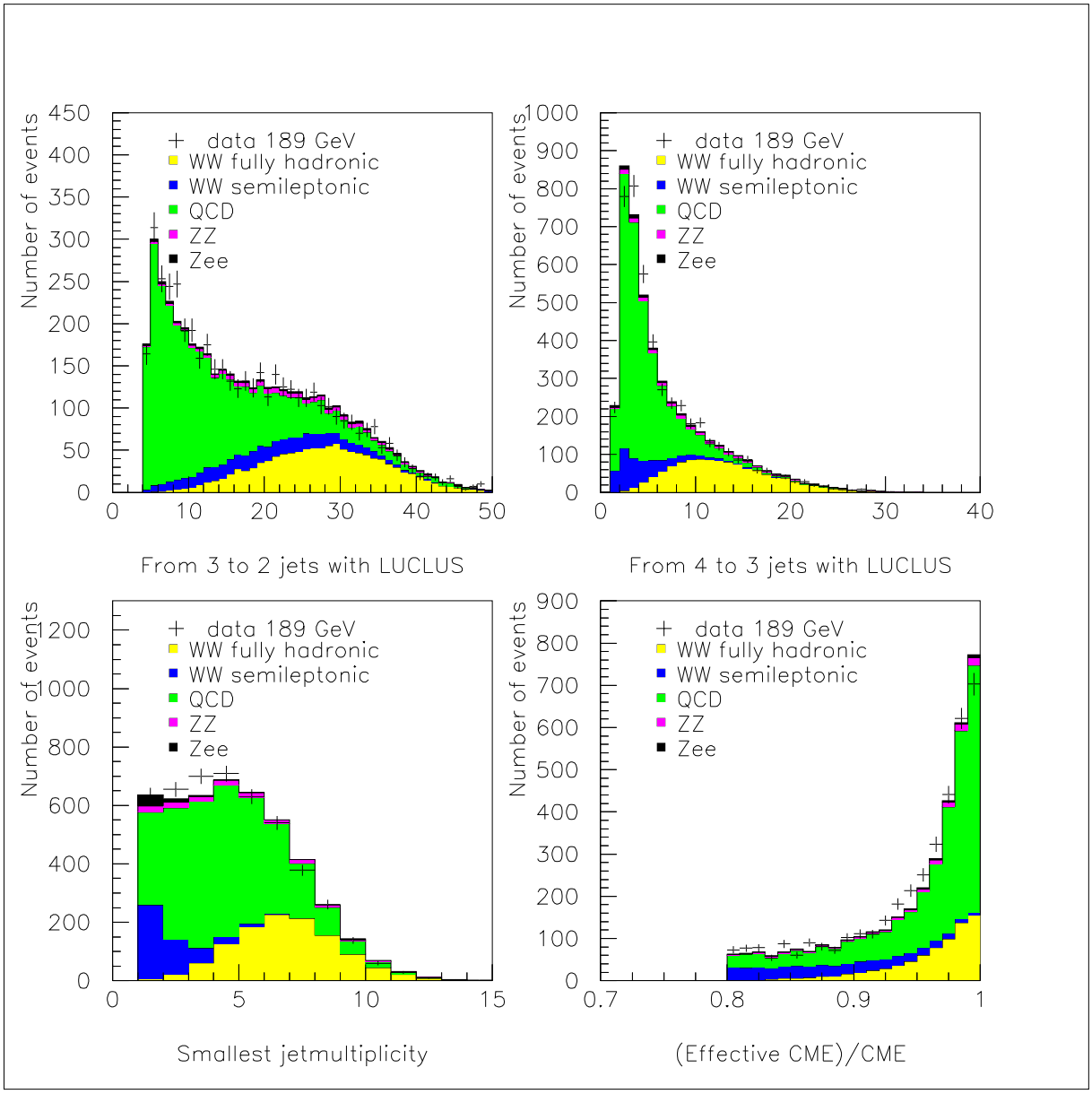

Figure 2: Data - Simulation comparison after preselection.

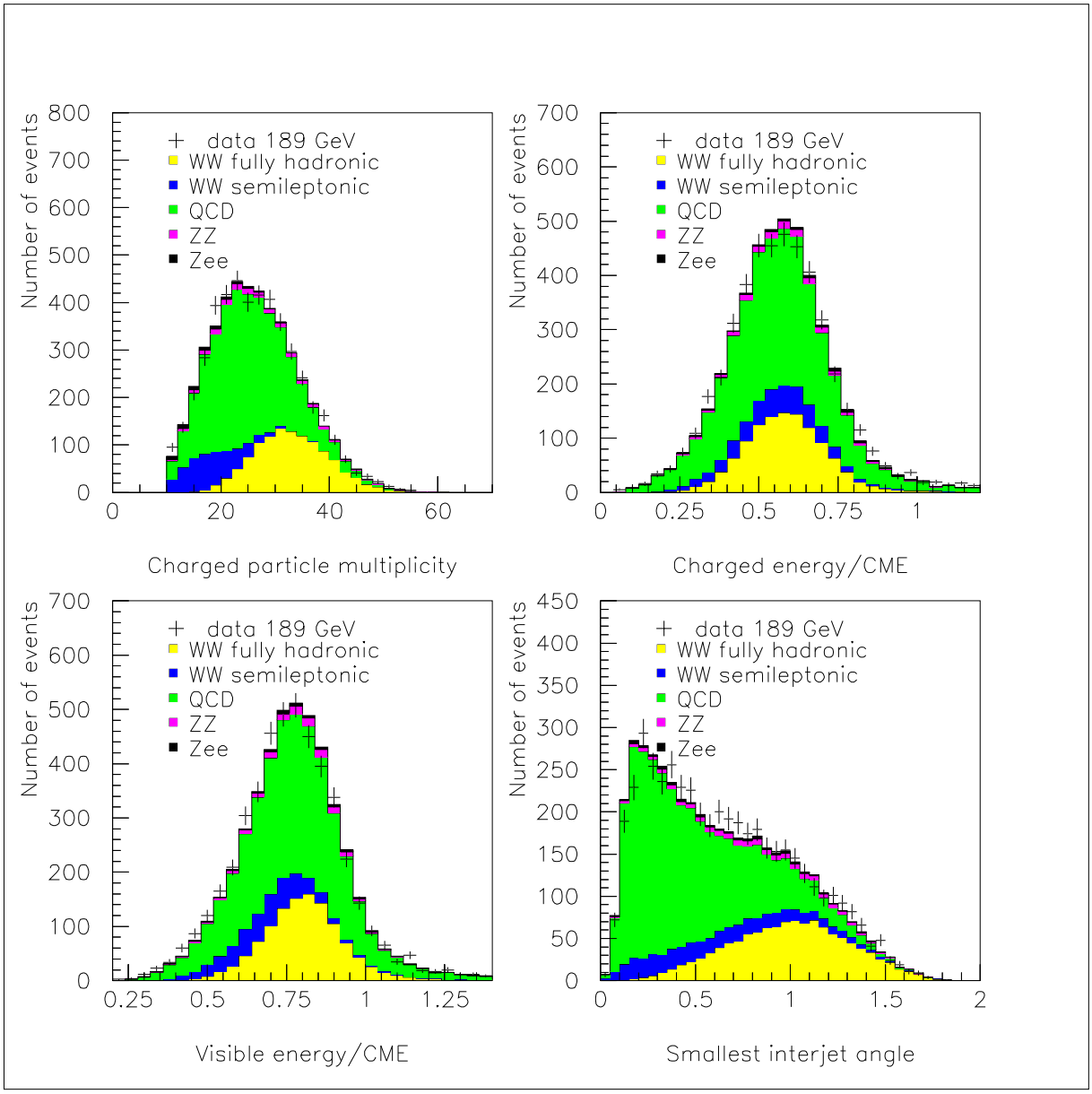

Figure 3: Data - Simulation comparison after preselection.

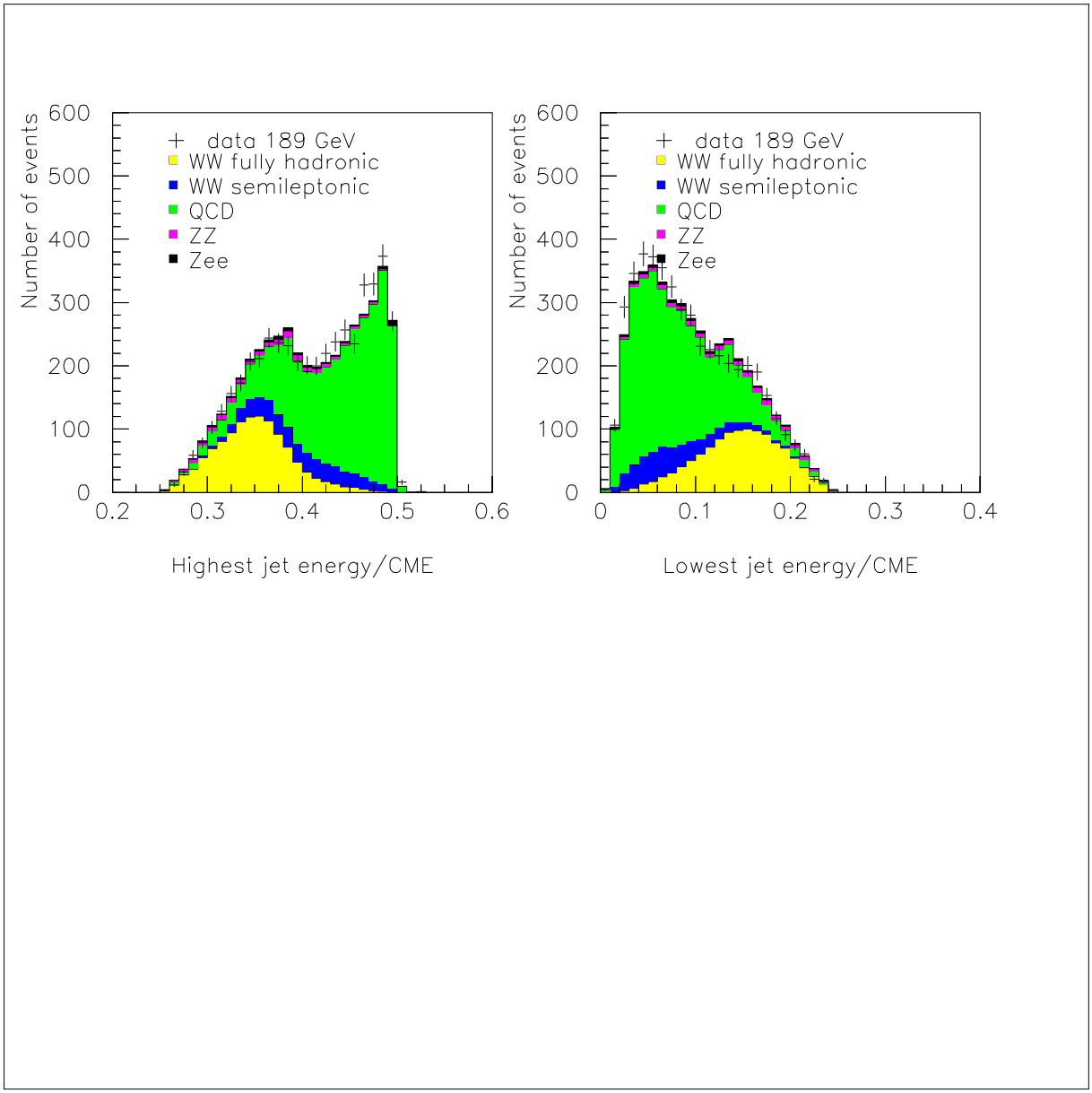

Figure 4: Data - Simulation comparison after preselection.

The network has been implemented in the 'Stuttgart Neural Network Simulator' [14] package.
The feature vectors for the events contained the following 10 variables:

- $\frac{\sqrt{s'}}{\sqrt{s}}$
- $\frac{E_{vis}}{\sqrt{s}}$
- $\frac{E_{ch}}{\sqrt{s}}$

  with $\sqrt{s'}$,$\sqrt{s}$,E_{vis} and E_{ch} defined as before.
- CPM : number of charged particles.
- $E_{jet,max}$, $E_{jet,min}$: maximum and minimum jet energy, after forcing the event in 4 jets and applying a 4C fit (four momentum conservation).
- $\theta_{jj,min}$: minimum interjet angle, after forcing the event in 4 jets and applying a 4C fit.
- $d_{join}(4 \rightarrow 3)$, $d_{join}(3 \rightarrow 2)$: the value of the LUCLUS parameter d_{join} for which the topology of the event changes from 4 jets to 3 jets and from 3 jets to 2 jets respectively.
- $PM_{jet,min}$: smallest jetmultiplicity, after forcing the event in 4 jets.

The training was done with a standard backpropagation algorithm. The weights were optimized by feeding the network 4000 signal and 4000 $q\bar{q}$ events (both preselected) 1500 times, being shuffled every cycle. For other backgrounds no training has been done.

For testing the network, independent samples of signal and all background events (remaining after preselection) were fed to the network to calculate the network performance. The network-output distribution for the simulated samples agreed well with that for the data, as shown in Fig. 5.

A different training however leads to another performance of the network. Different training-variations have been studied like :

- other training durations.
- other reweighting algorithms.
- using a pruning algorithm: this reduced the hidden layer of the network to 3 nodes instead of 8, with a similar performance. This is interesting since it shows that the network topology chosen might be more complex than necessary, but no further research has been done in this direction.

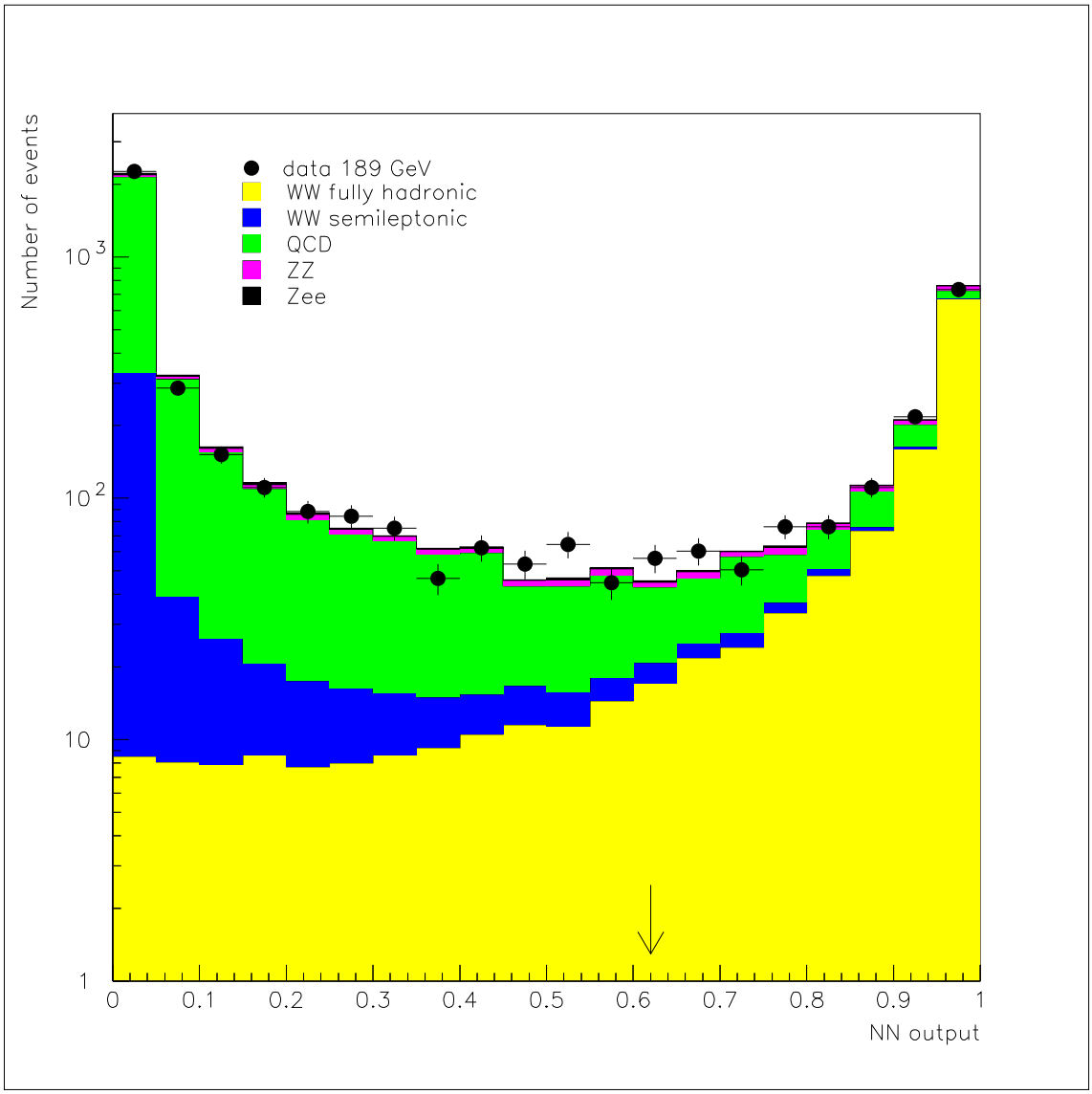

Figure 5: Output distribution of the network. The points are the output for the data and the histograms show the output for the simulated signal and the different backgrounds.

The network used for the analysis was optimized in function of the effects of these variations.

The ultimate selection was done by a cut on the network output, where the product of total efficiency and purity is maximal. Fig. 6 motivates the choise : $NNoutput > 0.62$.

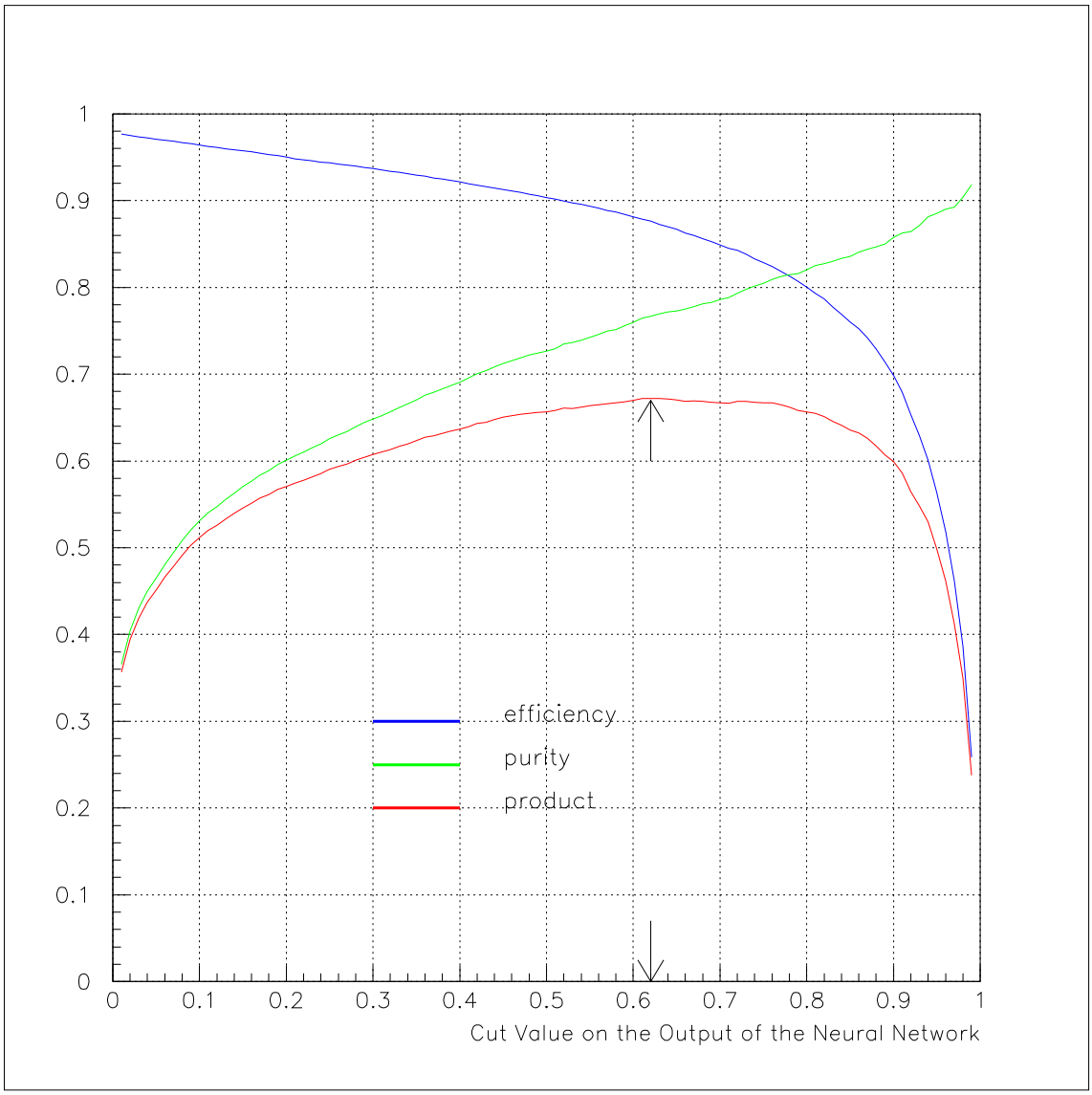

Figure 6: Evolution of efficiency, purity and the product of both as a function of the cut value on the network output.

The total selection efficiency and purity are $87.7 \pm 0.3\%$ and $76.7 \pm 0.6\%$ respectively. A total of 1369 candidates remained in the data, while we expected 1361 events from simulation. In Table 2 the remaining signal and background is

presented.

Table 2: Result after cut on NN output variable.

Type	Total efficiency (in %)	cross-section left (pb)
$q\bar{q}l\nu$	2.2	0.16
$q\bar{q}$	0.86	1.44
ZZ	24.6	0.39
Zee	0.39	0.03
Total background	–	2.02
$q\bar{q}q\bar{q}$	87.7	6.65

4 Result and systematic error

4.1 Cross-section

A total of 1369 events where selected out of the $157pb^{-1}$ of data. Taking into account the efficiency and purity of the selection, the following value is obtained for the cross-section:

$$\sigma_{e^+e^-\to W^+W^-\to q\bar{q}q\bar{q}} = \sigma^{tot}_{WW}.BR(WW \to q\bar{q}q\bar{q}) = 7.63 \pm 0.21(stat)pb \qquad (4)$$

where $BR(WW \to q\bar{q}q\bar{q})$ is the probability for the WW pair to give a purely hadronic final state. The error is statistical only.

4.2 Systematic error

The systematic error introduced by a neural network is very hard to determine. There is no consensus about standard procedures, but some guidelines exist. The problem is that small deviations in the distribution of the input variables might introduce non negligible deviations in the output distribution, thus in the number of selected events.
First the following remarks should be made :

- After the preselection described in paragraph 3.1, only small differences could be seen between data and simulation in the distributions of the variables filling the input or feature vectors.
- For the output distribution of the Neural Network there is good agreement between data and Monte Carlo simulation.

Considering this, only one study has been made to investigate the effect on the network behaviour due to small deviations in the input variables.

Each of the input variables, except CPM and $PM_{jet,min}$, has been smeared separately following a Gaussian distribution, with a width given by the experimental error on its measurement. For the two discrete variables, no study has been made.

The network performance has been calculated for every smeared variable, always using the same original events. The resulting shift in the cross-section has been determined. The results are given in Table 3. The biggest deviation in efficiency and purity was caused by the smearing of the maximum jet energy variable.

Table 3: Result after smearing of the input variables.

Spread Variable	eff.(%)	pur.(%)	Shift eff.	Shift pur.	Shift cross-sec.(pb)
$\frac{\sqrt{s'}}{\sqrt{s}}$	87.3	76.7	-0.4	0.0	0.03
$\frac{E_{vis}}{\sqrt{s}}$	87.5	76.6	-0.2	-0.1	0.00
$\frac{E_{ch}}{\sqrt{s}}$	87.6	76.8	-0.1	0.1	0.01
$E_{jet,max}$	87.1	75.9	-0.6	-0.8	-0.03
$E_{jet,min}$	87.6	76.6	-0.1	-0.1	-0.01
$\theta_{jj,min}$	87.7	76.5	0.0	-0.2	-0.02
$d_{join}(4 \rightarrow 3)$	87.5	76.6	-0.2	-0.1	0.00
$d_{join}(3 \rightarrow 2)$	87.6	76.5	-0.1	-0.2	-0.02
$total$					0.05

The variation of the cross-section with the cut on the Neural Network output variable has been examined and leads to an additional systematic error of 0.03 pb. Taking into account also the errors on the total integrated luminosity, on the efficiency and purity due to the limited Monte Carlo statistics and the theoretical error on the cross-sections for signal and $q\bar{q}$ events we have for the systematic error:

$$\sigma_{e^+e^- \rightarrow W^+W^- \rightarrow q\bar{q}q\bar{q}} = 7.63 \pm 0.21(stat) \pm 0.11(syst)pb \qquad (5)$$

5 Summary

For the data sample of $157pb^{-1}$ integrated luminosity taken by the DELPHI experiment in e^+e^- collisions at an average centre-of-mass energy of 188.63 GeV, the cross-section for the process $e^+e^- \rightarrow W^+W^- \rightarrow q\bar{q}q\bar{q}$ was measured to be $7.63 \pm 0.21(stat) \pm 0.11(syst)pb$. The first error is statistical and the second error is systematic. The systematic error comes from the uncertainty on the total integrated

luminosity and from the errors on the selection efficiency and purity determined in paragraph 4.2.
This result is compatible with the Standard Model prediction of $7.59 \pm 0.15 pb$ and is in agreement with the results published by the four LEP experiments [8].

References

[1] DELPHI Collaboration, P. Aarnio et al., Nucl. Instr. & Meth. A303 (1991) 23.

[2] DELPHI Collaboration, P. Abreu et al., Nucl. Instr. & Meth. A378 (1996) 57.

[3] W. Beenakker et al., WW cross-sections and distributions, Physics at LEP2, eds. G. Altarelli, T. Sjöstrand and F.Zwirner, CERN 96-01 (1996) Vol 1, 79.

[4] T. Sjöstrand, PYTHIA 5.719 / JETSET 7.4 Physics at LEP2, eds. G. Altarelli, T. Sjöstrand and F.Zwirner, CERN 96-01 (1996) Vol 2, 41.

[5] DELPHI Collaboration: DELPHI event generation and detector simulation - User Guide, DELPHI Note 89-67 (1989), unpublished.

[6] DELPHI Collaboration, P. Abreu et al., W pair production cross-section and W branching fractions in e^+e^- interactions at $189GeV$, CERN-EP-2000-035 (2000). (Accepted by Physics Letters B)

[7] D. Bardin et al., Nucl.Phys. (Proc.Suppl.) B37 (1994) 148; D. Bardin et al., GENTLE/4fan v. 2.0: A program for the Semi-Analytic Calculation of Predictions for the Process $e^+e^- \to 4f$, DESY 06-233, hep-ph/9612409.

[8] The LEP EW Working group, http://www.cern.ch/LEPEWWG

[9] R. Penrose, Artificial Intelligence. Setting the Scene : the Claim and the Issues, 1991 CERN School of Computing, CERN 92-02 (1992) 1.

[10] R. Beale and T. Jackson, Neural Computing : An Introduction. Adam Hilger (1990)

[11] C. Peterson and T. Rögnvaldsson, An Introduction to Artificial Neural Networks, 1991 CERN School of Computing, CERN 92-02 (1992) 113.

[12] DELPHI Collaboration, P. Abreu et al., Nucl. Instr. & Meth. A427 (1999) 487.

[13] T. Sjöstrand, PYTHIA 5.7 / JETSET 7.4, CERN-TH7112/93 (1993)

[14] University of Stuttgart, Institute for parallel and distributed high performance systems (IPVR), Stuttgart Neural Network Simulator SNNS, User manual 4.1.