UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE FÍSICA
CAIXA POSTAL 66318
05315-970 SÃO PAULO - SP
BRASIL

PUBLICAÇÕES

IFUSP/P-1301

e^+/e^- IDENTIFICATION IN A WIDE MOMENTUM RANGE WITH A TRANSITION RADIATION DETECTOR

N. Kuropatkin and **R. Zukanovich Funchal**
Instituto de Física, Universidade de São Paulo

Março/1998

e^+/e^- identification in a wide momentum range with a transition radiation detector

N. Kuropatkin [1,2] and R. Zukanovich Funchal [3]

Instituto de Física da Universidade de São Paulo
05389-970 C.P. 66318 - São Paulo, SP
Brazil

Abstract

We discuss and compare the performance of three popular methods (cluster sum cut, likelihood ratio and artificial neural network) for e^+/e^- identification using a transition radiation detector in the environment of a hadroproduction experiment.

We evaluate each method with respect to its efficiency, background suppression and robustness. Our goal was to find the best method for lepton identification in a wide range of lepton momenta, from 10 to 300 GeV/c.

1 Introduction

In this work we compare the performance of three different methods that provide e^+/e^- identification using the Transition Radiation Detector (TRD) of E781 experiment at Fermilab [1]. In our experiment we measure the momentum of the particle tracks that enter the TRD region and this will be exploited in our classification schemes.

The track classification using transition radiation detectors has to meet two requirements: good efficiency in selecting the signal and good background rejection. The first requirement is crucial to ensure good acceptance for leptonic events, the second one is vital specially when working in the presence of big electromagnetic background.

⋆ Submitted to Nucl. Instr. Meth. A.

[1] On leave of absence from Petersburg Nuclear Physics Institut, Gatchina, Russia.
[2] E-mail: kuropat@charme.if.usp.br.
[3] E-mail: zukanov@charme.if.usp.br.

The TRD we are working with is made of 6 blocks each one composed by a radiator (210 foils of CH_2) and a Xe-CH_4 filled proportional chamber. It is practically the same detector as described in Ref. [3], except that the size of the strips was reduced to 4 mm.

The momentum region of the e^+/e^- tracks under investigation is from 10 to 300 GeV/c, which corresponds to the growing part of the Transition Radiation (TR) yield curve for pions [2]. Above this momentum it is more difficult to distinguish pions from electrons using this detector.

To achieve appropriate background suppression as well as good signal efficiency we have studied three different approaches. The first approach, known as cluster sum cut, exploits the difference in the detector response for radiative and non-radiative particles; the second approach is the likelihood ratio which explicitly takes into account the $\gamma = E/m$ factor dependence of the detector response function and finally we have investigated a Feed Forward Error Back Propagation artificial neural network algorithm.

All these methods are simple enough to be suitable for use in the on-line/off-line software of the experiment but as we will see in the following one of them is more robust with respect to changes in the experimental conditions.

In Sec. 2 we describe the cluster sum cut method, in Sec. 3 we discuss the likelihood ratio technique and in Sec. 4 we present the neural network approach. In Sec. 5 we compare the performance of each one of the methods with real experimental data. Finally the last section is devoted to our conclusions.

Although we have studied these identification techniques with a certain detector configuration in mind, our conclusions are valid for virtually any TRD in the cluster counting operating mode, specially when operating under heavy background conditions and used for identification of particles in a wide momentum range.

2 Cluster sum cut method

Our TRD works in the cluster counting mode [2] in which there is a natural way to distinguish e^+/e^- from other particles - the cluster sum cut method. This is a statistical method based on the fact that charged particles with high γ factor ($> 4\,000$), the so called radiative particles, will produce significantly more ionization clusters due to photoelectrons than non-radiative ones. The production of clusters for non-radiative particles will be dominated by $\delta-$electrons. Due to the heavy background conditions in hadroproduction experiments the cluster counting has to be performed in a narrow corridor

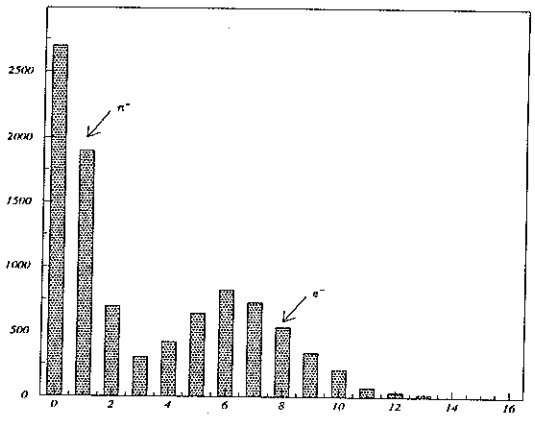

Fig. 1. Distribution of the sum of the number of cluster in TRD.

along the track direction [3]. The distribution of the sum of the number of clusters produced by a track in all TRD planes will obey in either case a Poisson distribution with a different mean. Usually a simple cut in the sum of the number of clusters is enough to identify e^+/e^- from other particles if these means are well separated, see Fig. 1. This is possible because the electrons in our momentum region are on the detector response function plateau while other particles, with momentum less than $\approx$ 60 GeV/c, are still non-radiative particles.

This method has an intrinsic limitation due to the fact that pions with high momentum begin to produce transition radiation which will reduce the rejection power of the method. Furthermore the fact that the cut to be optimized is an integer value makes it rather difficult to establish an optimal signal to background ratio.

3 Description of TRD likelihood method

This method was proposed for TRD detectors by M. L. Cherry et al. [4] and was demonstrated to give better performance than traditional methods in Ref. [5–7].

Because we would like the method to work in a wide momentum range we have explicitly put the γ dependence in the definition of the likelihood function and redefined the method in the following way.

The likelihood function is built in a way to classify particles into two categories : e^+/e^- (type 1) and others (type 2). So if a particle of type $i = 1, 2$ with Lorentz factor γ_i generates a sample $X=\{x_1, x_2, ..., x_6\}$ of TR clusters along its track we can define the probability $P(X|i, \gamma_i)$ of this event as

$$P(X|i, \gamma_i) = \prod_{k=1}^{6} P(x_k|\gamma_i). \tag{1}$$

The probability density function $P(x_k|\gamma_i)$ can be calculated using the detector response function. The response function is defined as a modified Poisson distribution function with mean depending on the γ factor, that is,

$$P(x_k|\gamma_i) = [C(x_k)\frac{a_k(\gamma_i)^{x_k} \exp -a_k(\gamma_i)}{x_k!} + f_{\rm cor}(x_k)] \cdot \epsilon_k, \tag{2}$$

where $a_k(\gamma_i)$ is the mean of the cluster multiplicity distribution for the TRD block k which depends on the γ factor of particle type i, $C(x_k)$ are correction constants taken from experimental data, $f_{\rm cor}(x_k)$ is the probability of a non-radiative particle to produce x_k clusters by ionization, also taken from experimental data, and ϵ_k - is the k chamber efficiency, which permits us to take into consideration the possible variations in the experimental conditions.

We can define $P(0|\gamma_i)$ which is the probability of a particle i to produce zero number of clusters in the block k from the probability normalization condition as

$$P(0|\gamma_i) = 1 - \sum_{x_k=1,4} P(x_k|\gamma_i). \tag{3}$$

We consider the maximum number of clusters produced by a track in a TRD chamber to be four, the probability to make more than four clusters, even for radiative particles, is negligible and included in $P(4|\gamma_i)$.

The mean $a_k(\gamma_i)$ was defined for each TRD block k as

$$a_k(\gamma_i) = \bar{N}_k \exp(\frac{3.52 \ln(\frac{\gamma_i}{\gamma_i + 1200})}{1 + (\frac{\gamma_i}{1500})^2}), \tag{4}$$

where $\bar{N}_k$ is the mean number of clusters for block k at the plateau of the detector response function [3].

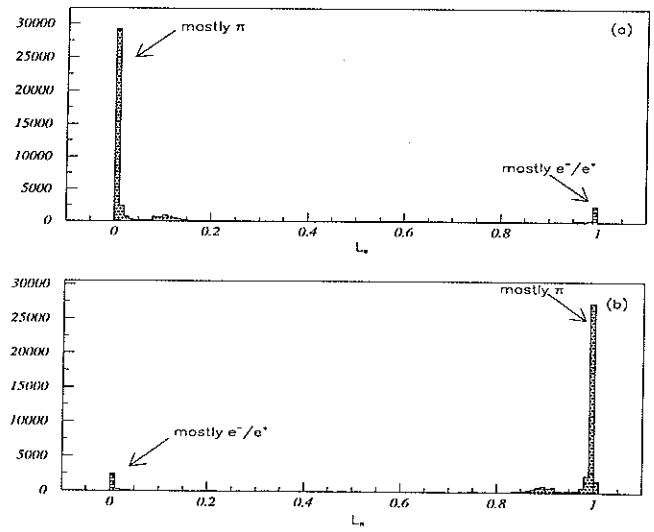

Fig. 2. Distribution of (a) $\mathcal{L}_e$ and (b) $\mathcal{L}_\pi$ for all particle tracks.

We define the likelihood ratio for particle of type i as

$$\mathcal{L}_i(X|\gamma_i) = \frac{P(X|i,\gamma_i)}{\sum_{j=1}^{2} P(X|j,\gamma_j)}, \tag{5}$$

which is restricted to the interval $0 \leq \mathcal{L}_i \leq 1$, and use this ratio as an indicator of the particle type. For each track one can calculate the above ration in two possible hypothesis. We expect that this ratio will be close to 1 whenever the hypothesis is correct and closer to zero whenever it is wrong. This can be seen in Fig. 2.

Of course there will be some region where the method will not be able to distinguish one hypothesis from the other. We will exclude this region, which will be the main contribution to the inefficiency of this method.

4 Description of TRD neural network

As another attempt to improve the background suppression in the electron identification in TRD we have developed an algorithm using a Feed Forward Error Back Propagation artificial neural network [8]. A similar application

was developed in Ref. [9] for 10 modules of TRD used for cosmic ray lepton identification. In this work it was proposed a network structure with 10 + 21 + 2 nodes. The network was feed with a normalized vector containing cluster counts from each of the TRD blocks. The authors have demonstrated that the network solution was preferable in comparison with the likelihood method for several fixed energies from 1 to 4 GeV.

In our experimental conditions we have to, as explained before, count the clusters along the predicted track direction and provide solution on the track by track bases in a large momentum region. As we have 6 modules of TRD we will use 6 input nodes that will receive the cluster sum along the track in each TRD block normalized to unity. To explicitly take into account the γ dependence of the detector response and consequently increase the momentum region in which the algorithm can provide an efficient classification of tracks, an extra node was introduced. This node was fed with a normalized to one γ factor calculated in the pion hypothesis,i.e.

$$\text{node 7 activation} = \frac{E}{m_\pi \gamma_{\text{cut}}}, \tag{6}$$

where E is the energy of the particle, m_π the mass of the pion and γ_{cut} was chosen to be 3000, which corresponds to pions of about 400 GeV/c. For γ greater than γ_{cut} the node 7 activation is equal to one.

We start with 7 input nodes and would like to have a similar classification for particles here as in the previous method, that is, two output nodes. So according to Kolmogorov theorem [10] we can approximate our classification function with 15 nodes in the hidden layer. This defines the structure of the network as 7+15+2 nodes, as presented in Fig. 3.

The first layer of nodes are fully connected to the second layer of 15 nodes with a sigmoid response function which is followed by the output layer of 2 nodes (res 1 and res 2). Each neuron of one layer receives as input the outputs of all neurons from the previous layer with weights defined by the synaptic matrix W. The activation level of the output nodes will provide the track classification.

The sigmoid function we have used is defined as following:

$$F(x,b) = \frac{1}{1+\exp(-2.(x-b))},$$

where b is the neuron threshold.

The training process was performed with a standard back-propagation tech-

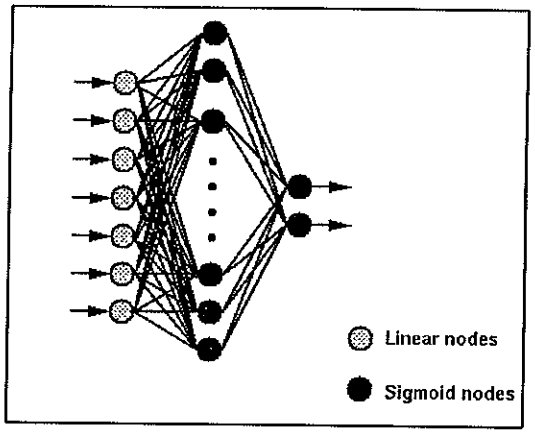

Fig. 3. The TRD neural network structure.

nique during which the corrections to the synaptic matrix elements W_{lk} were calculated according to the rule :

$$\Delta W_{lk}(i) = -S\frac{\partial E[W]}{\partial W_{lk}} + M\Delta W_{lk}(i-1),$$

where $\Delta W_{lk}(i)$ is the correction to the synaptic matrix elements W_{lk} after i steps, E[W] is the summed square error function, S is the learning rate parameter and $M\Delta W_{lk}(i-1)$ is the momentum term used to avoid sudden oscillations. We have used in our implementation $S = 0.1$ and $M = 0.3$.

To train our network we have used data simulated by GE781 [11], the E781 GEANT [12] based Monte Carlo. It is in principle also possible to use experimental data for this purpose, but for this we need some independent tagging of electrons.

In Fig. 4 we display the general behavior of the two output nodes, after the training process, for the classification of particles in the TRD.

5 Comparison of methods

To compare the performance of these three methods we have used data from a special calibration run in which the beam composition was about 50% e^-, 50% π^- with momentum around 25 GeV/c; Monte Carlo generated data, here we know exactly the particle type and hadronic interaction data.

To choose a suitable efficiency to contamination ratio working point we can

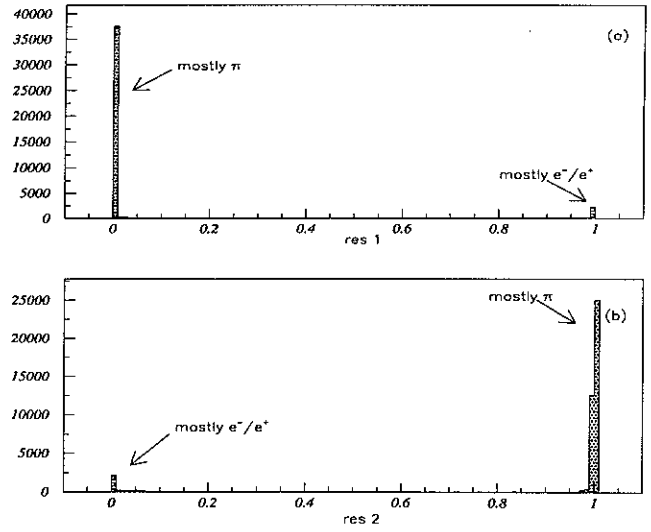

Fig. 4. Network response of the output nodes (a) res 1 and (b) res 2 for all particle tracks.

apply correlated cuts in the feature space of $\mathcal{L}_e$ (in the e^+/e^- hypothesis) and $\mathcal{L}_h$ (in the other particles hypothesis) for the likelihood method and in the feature space of res 1 and res 2 for the artificial neural network. This can be done in the following way:

$$\mathcal{L}_e > \text{cut} \quad \text{and} \quad \mathcal{L}_h \leq 1. - \text{cut}, \tag{7}$$

for the likelihood ratio case and

$$\text{res } 1 > \text{cut} \quad \text{and} \quad \text{res } 2 \leq 1. - \text{cut}, \tag{8}$$

for the network, where cut is any real value from zero to one. Changing the value of the cut we can build plots of the hadronic contamination as a function of electron detection efficiency for these two methods.

To make a similar curve for the cluster sum method we have to change the cluster sum cut by an integer value which gives only a few points on the plot.

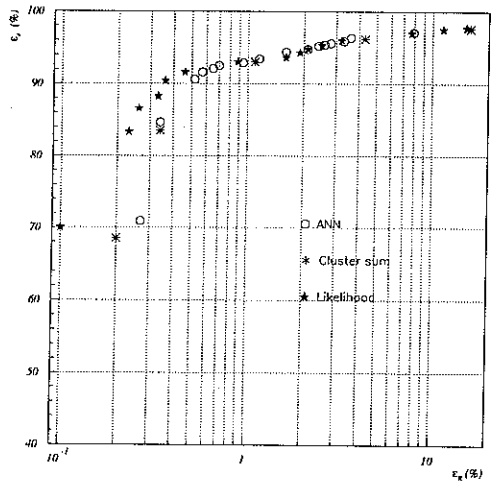

Fig. 5. Electron identification efficiency (ϵ_e) versus pion contamination (ϵ_π): a comparison of the three methods with calibration run data.

5.1 *Calibration run data*

For the detector calibration purpose a special beam was prepared with composition about 50% e^-, 50% π^- at about 25 GeV/c. The particle type of this beam was tagged with the help of another TRD, the beam TRD, a similar detector to the one described in Ref. [2].

We have built a plot of electron identification efficiency (ϵ_e) versus the pion contamination (ϵ_π) for each one of the three methods as well as the ratio between ϵ_e and ϵ_π versus ϵ_e. This is presented in Fig. 5 and in Fig. 6 respectively.

In the conditions of fixed beam momentum, one should not expect to find differences among the three techniques. Nevertheless one can see that the likelihood ratio method gives the best performance.

5.2 *Monte Carlo generated data*

The calibration data were taken in a fixed narrow momentum bite, so we cannot use these data to investigate the acceptance and contamination of the methods for a wide range of particle momenta. We can use Monte Carlo generated data, which is in good agreement (up to a few %) with experimental results, to study this.

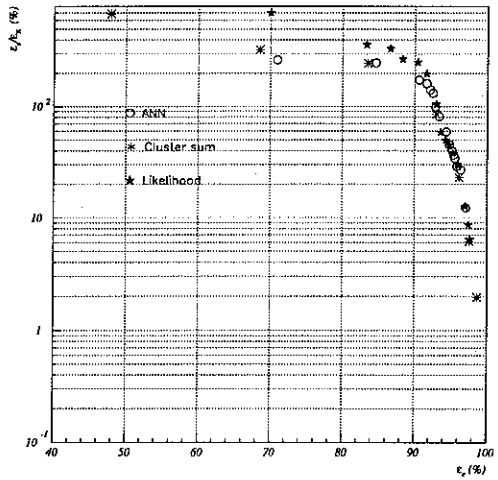

Fig. 6. Ratio between electron identification efficiency (ϵ_e) and pion contamination (ϵ_π) versus ϵ_e, for the three methods under study.

As one can see in Fig. 7 the efficiency for radiative particles is rather independent of the particle momentum for the artificial neural network algorithm. The efficiency decreases slightly after about 130 GeV/c which partly compensates the growth of the contamination. For this classification scheme the contamination remains low up to momentum around 130 GeV/c and than starts to rise up to 35% at 280 GeV/c. In the case of the likelihood ratio one observes that the efficiency decreases with almost constant slope reaching about 50% at 280 GeV/c. This gives rather good compensation of the contamination growth. The contamination remains very low regardless of the momentum region. This feature of the method is a crucial advantage, as we need to keep the contamination as low as possible. The cluster cut method's efficiency is obviously momentum independent but as expected contamination rises very fast for particles with momentum greater than 60 GeV/c.

We also can simulate the momentum distribution of particles in our experiment to estimate the integrated efficiency and contamination. This will be used to compare our contamination calculations with experimental data in the next section. To emphasize the difference among methods we have plotted the electron efficiency over the hadronic contamination ratio as a function of the electron efficiency.

The likelihood method achieves acceptable background level at higher electron efficiency and the neural network gives very close result for $\epsilon_e > 80\%$. The cluster sum method looses efficiency very rapidly as we change the in-

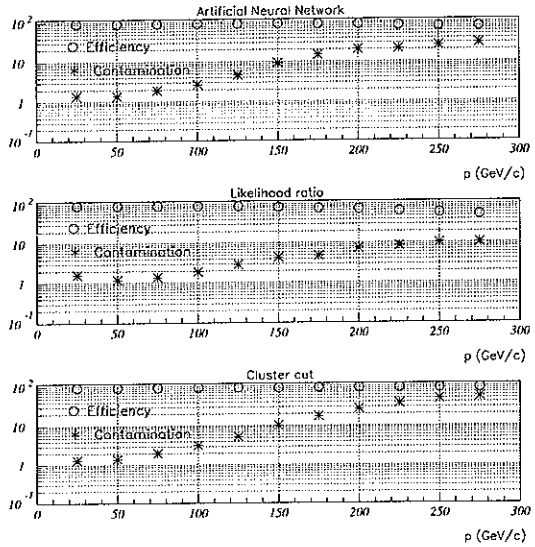

Fig. 7. Efficiency and contamination for the three algorithms as a function of particle momenta.

teger value of the cut which makes it difficult to select an optimal signal to background ratio.

5.3 Hadronic interaction data

This is a sample of data with the real background conditions, which permits us to estimate the background suppression achieved by each algorithm. This is clearly not possible to calculate with calibration data – which were taken in special conditions (very low particle multiplicity, low momenta and narrow momentum bite); nor with Monte Carlo simulated data which mimics real data general features but are not expected to reproduce the exact background conditions of the experiment.

In the case of hadronic interaction data we have no independent way to tag electrons, but we can estimate the number of e^+/e^- in this sample to be of order of 2%. Keeping this in mind we can use this data to calculate the real contamination level for each method.

To build the same distributions as for Monte Carlo data we have used the efficiencies obtained by the Monte Carlo simulation, which is quite reliable for electrons and get the contamination levels from the interaction data.

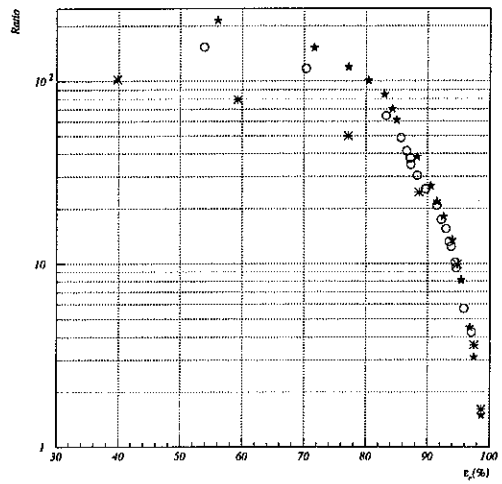

Fig. 8. Electron efficiency over hadron contamination ratio versus electron efficiency.

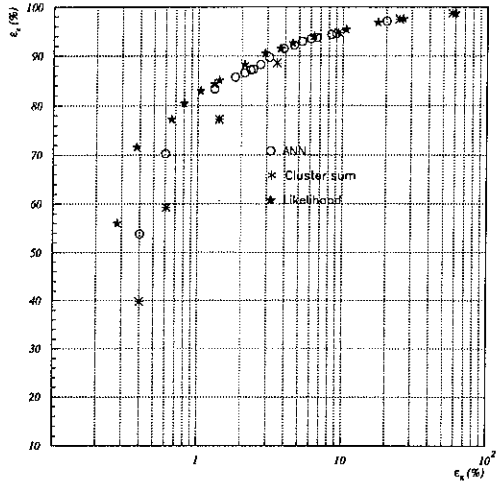

Fig. 9. Electron identification efficiency (ϵ_e), taken from Monte Carlo simulation versus hadronic contamination (ϵ_π), taken from hadroproduction data for the three methods under study.

We can see in Fig. 9 that our conclusions made from Monte Carlo data are confirmed by real hadroproduction data, so we can be confident that the momentum dependence of the three methods displayed in Fig. 7 is well understood.

6 Conclusions

We have studied three different methods to classify particles using TRD. Two of these methods directly uses momentum information of the track.

Our study shows that all three methods give almost the same performance in the case of low momentum particles, $\approx$ 60 GeV/c. For higher momentum the likelihood method looks preferable from the signal to contamination ratio point of view.

Although the cluster sum method is the simplest method for implementation it should not be used in wide momentum rage situation where contamination was shown to rise rapidly with momentum growth.

The comparison between the neural network method and the likelihood ratio has demonstrated the better performance of the last one. We should point out that the likelihood method permits to take into account changes in experimental conditions as the detector efficiencies are part of the algorithm. In the case of the artificial neural network to take into account big changes in the experimental conditions one would need to repeat the training process each time. The network performance depends on the training data and so depends on the quality of the Monte Carlo simulation. This robustness give an additional reason to choose the likelihood method as the preferable one.

Acknowledgement

We thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for financial support.

References

[1] FNAL E781 (SELEX) proposal (unpublished), July 1993; J. Russ, "SELEX – Hadroproduction of Charm Baryons out to Large x_F", presented at "Production and decay of Hyperon, Charm and Beauty Hadrons", Strasbourg, France (1995).

[2] B. Dolgoshein, Nucl. Instr. and Meth. A326 (1993) 434.

[3] A. Denisov et al., preprint Fermilab-Conf-84/134-E (1984); V. A. Andreev et al., preprint LENINGRAD-86-1186 (1986).

[4] M. L. Cherry et al., Nucl. Instr. and Meth. 115 (1974) 141.

[5] A. Büngener et al., Nucl. Instr. and Meth. 214 (1983) 261.

[6] K.-K. Tang, The Astrophysical Journal, 278 (1984), 881.

[7] R. D. Appuhn et al., Nucl. Instr. and Meth. A263 (1988) 309.

[8] C. Peterson and T. Rögnvaldson, "An Introduction to Artificial Neural Networks", LUTP 91-23, Lectures given at the 1991 Cern School of Computing, Ystad (1991).

[9] R. Bellotti et al., Computer Phys. Commun. 78 (1993) 17.

[10] A. N. Kolmogorov, Dokl. Akad. Nauk. USSR 114 (1957) 953.

[11] G. Davidenko et al., GE781: a Monte Carlo package for fixed target experiments, in: R. Shellard and T. D. Nguyen, eds., *Proceedings of the International Conference on Computing in High Energy Physics'95* (World Scientific, Singapore, 1996) p.832.

[12] R. Brun, F. Bruyant, M. Marie, A. C. McPherson, and P. Zanarini, GEANT3, DD/EE/81-1, CERN (1987); GEANT Detector Description and Simulation Tool, Application Software Group, CND, CERN Program Library Long Writeup W5013 (1994).