Bulletin of PFUR Series *Mathematics. Information Sciences. Physics.* No 2 (2). 2010. Pp. 76–84

UDC 004.93'1:519.237:51-72

Methods of e/π Identification with the Transition Radiation Detector in the CBM Experiment

E. P. Akishina, T. P. Akishina,
O. Yu. Derenovskaya, V. V. Ivanov

Laboratory of Information Technologies
Joint Institute for Nuclear Research
Joliot-Curie 6, 141980 Dubna, Moscow region, Russia

A problem of e/π identification using n-layered transition radiation detector (TRD) in the CBM experiment is considered. With this aim, we elaborated algorithms and implemented various approaches. We discuss the characteristic properties of the energy losses by electrons and pions in the TRD layers and special features of applying artificial neural networks (ANN) and statistical methods to the problem under consideration. A comparative analysis is performed on the power of the statistical criteria and ANN.

Key words and phrases: general statistical methods, multivariate analysis, pattern recognition, CBM experiment, transition radiation detector TRD.

1. Introduction

The CBM Collaboration [1, 2] builds a dedicated heavy-ion experiment to investigate the properties of highly compressed baryon matter as it is produced in nucleus-nucleus collisions at the Facility for Antiproton and Ion Research (FAIR) in Darmstadt, Germany. Fig. 1 depicts a general layout of the CBM experiment. There is a Silicon Tracking System (STS) inside the dipole magnet. Ring Imaging Cherenkov (RICH) is designed to detect electrons. TRD arrays identify electrons with momentum above 1 GeV/c. TOF provides time-of-flight measurements needed for hadrons identification. ECAL measures electrons, photons and muons.

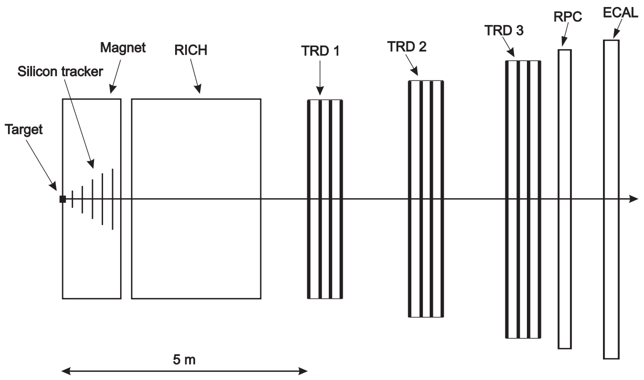

Figure 1. **Schematic view of the CBM experimental setup**

The measurement of charmonium is one of the key goals of the CBM experiment. For detecting J/ψ meson in its dielectron decay channel the main task is the e/π separation. One of the most effective detectors for solving this problem is the Transition Radiation Detector.

The problem of particle identification (PID) using n-layered TRD consists in the following: having a set of energy losses measured in n layers of the TRD, one has to estimate to which particle, e or π, this set is relative. To estimate the efficiency of particle identification, we used different approaches.

Received 28th November, 2009.

2. Traditional Statistical Criteria: MV and LFR Methods

In the mean value (MV) method the PID is based on a variable $\overline{\Delta E} = \frac{1}{n}\sum_{i=1}^{n} \Delta E_i$ (where ΔE_i is a particle energy loss in the i-th TRD layer and n is the number of layers in the TRD). Fig. 2 shows distributions of variable $\overline{\Delta E}$ for π (a), e (b), and a summary distribution for e and π (c).

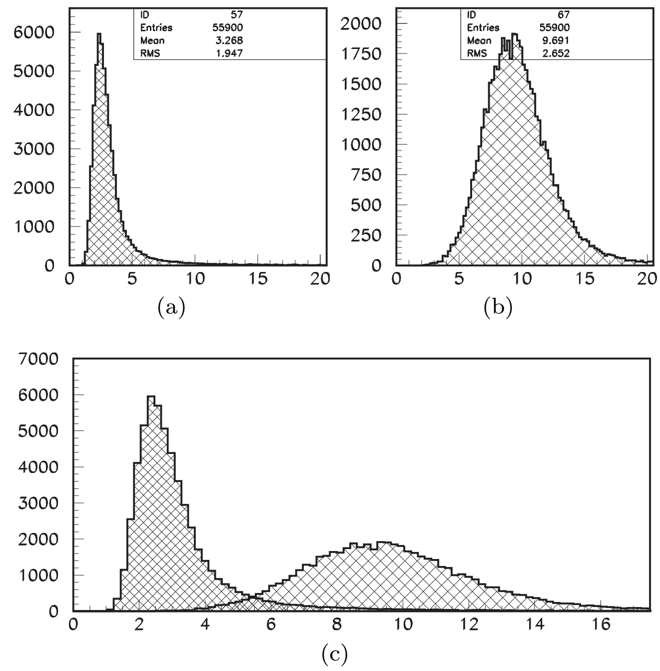

Figure 2. **Distributions of variable $\overline{\Delta E}$ for π (a) and for e (b) events; the summary distribution for π and e events (c)**

While applying the likelihood functions ratio (LFR) test [3] to the PID problem, the value

$$L = L = P_e/(P_\pi + P_e) \quad P_e = \prod_{i=1}^{n} p_e(\Delta E_i), \quad P_\pi = \prod_{i=1}^{n} p_\pi(\Delta E_i),$$

is calculated for each event (see Fig. 3), where $p_\pi(\Delta E_i)$ is the value of the density function p_π in the case when the pion loses energy ΔE_i in the i-th absorber, and $p_e(\Delta E_i)$ is a similar value for electron.

We have found that the distribution of ionization losses of pions in the TRD is well approximated by a log-normal density function (see Fig. 4)

$$f_1(x) = \frac{A}{\sqrt{2\pi}\sigma x} \exp^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, \tag{1}$$

where σ is the dispersion, μ is the mean value, and A is a normalizing factor.

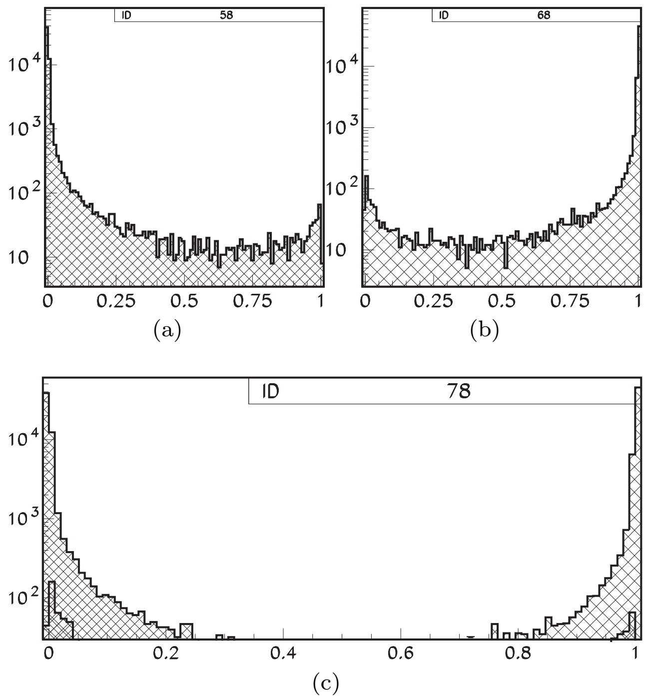

Figure 3. **Distributions of variable L for π (a) and for e (b); the summary distribution for π and e events (c)**

The distribution of energy losses by electrons in the TRD is approximated with a good accuracy by the density function of a weighted sum of two log-normal distributions (see Fig. 5)

$$f_2(x) = B\left(\frac{\alpha}{\sqrt{2\pi}\sigma_1 x}\exp^{-\frac{1}{2\sigma_1^2}(\ln x-\mu_1)^2} + \frac{\beta}{\sqrt{2\pi}\sigma_2 x}\exp^{-\frac{1}{2\sigma_2^2}(\ln x-\mu_2)^2}\right), \tag{2}$$

where σ_1 and σ_2 are dispersions, μ_1 and μ_2 are mean values, α and β are contributions of the first and second log-normal distributions, correspondingly, and B is a normalizing factor.

3. Nonparametric Goodness-of-Fit ω_n^k-criterion [4, 5]

This test is based on the comparison of the distribution function $F(x)$ corresponding to the preassigned hypothesis (H_0) with empirical distribution function $S_n(x)$:

$$S_n(x) = \begin{cases} 0, & \text{if} \quad x < x_1; \\ i/n, & \text{if} \quad x_i \leqslant x \leqslant x_{i+1}, \quad i = 1, \ldots, n-1. \\ 1, & \text{if} \quad x_n \leqslant x, \end{cases}$$

Here $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ is the ordered sample (*variational series*) of size n constructed on the basis of observations of variable x.

The test statistics measures the “distance” between $F(x)$ and $S_n(x)$. Such statistics are known as *non-parametric*. We suggested a new class of non-parametric statistics

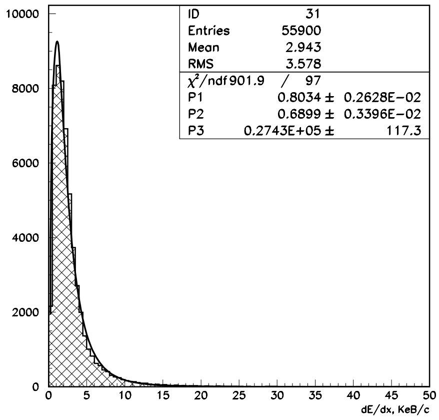

Figure 4. **Distribution of pion energy losses in one layer of the TRD and its approximation by a log-normal function (1)**

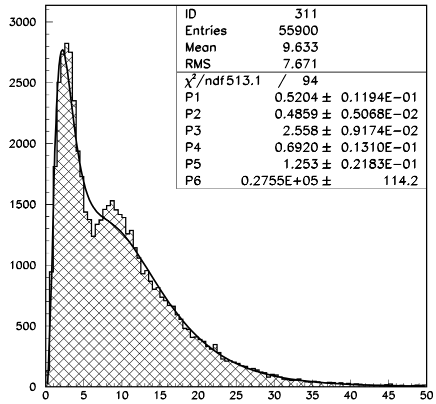

Figure 5. **Distribution of electron energy losses in one layer of the TRD and its approximation by a weighted sum of two log-normal functions (2)**

(with $k \geqslant 3$):

$$\omega_n^k = -\frac{n^{\frac{k}{2}}}{k+1}\sum_{i=1}^{n}\left\{\left[\frac{i-1}{n} - F(x_i)\right]^{k+1} - \left[\frac{i}{n} - F(x_i)\right]^{k+1}\right\}.$$

The goodness-of-fit criteria constructed on the basis of these statistics are usually applied for testing the correspondence of each sample to the distribution known *apriori*.

Energy losses for π have a form of Landau distribution. We use it as H_0 to transform the initial measurements to a set of a new variable λ:

$$\lambda_i = \frac{\Delta E_i - \Delta E_{mp}^i}{\xi_i} - 0.225, \quad i = 1, 2, \ldots, n, \tag{3}$$

ΔE_i — the energy loss in the i-th absorber, ΔE_{mp}^i – the value of most probable energy loss, $\xi_i = \frac{1}{4.02}$ FWHM of distribution of energy losses for π.

In order to determine ΔE_{mp}^i and FWHM of the distribution of pion energy losses in the i-th absorber, the indicated distribution was approximated by the density function of a log-normal distribution (see Fig. 4).

The obtained λ_i, $i = 1, \ldots, n$ are ordered due to their values (λ_j, $j = 1, \ldots, n$) and used for calculation of ω_n^k:

$$\omega_n^k = -\frac{n^{\frac{k}{2}}}{k+1}\sum_{j=1}^{n}\left\{\left[\frac{j-1}{n} - \varphi(\lambda_j)\right]^{k+1} - \left[\frac{j}{n} - \varphi(\lambda_j)\right]^{k+1}\right\}. \tag{4}$$

Here the values of Landau distribution function $\varphi(\lambda)$ are calculated using the DSTLAN function (from the CERNLIB library). Fig. 6 shows the distributions of ω_{12}^6 values for π (a) and e (b); the summary distribution is shown in the (c).

4. Combined Method

This approach is based on a successive application of two statistical criteria: 1) the mean value method, and 2) the ω_n^k test.

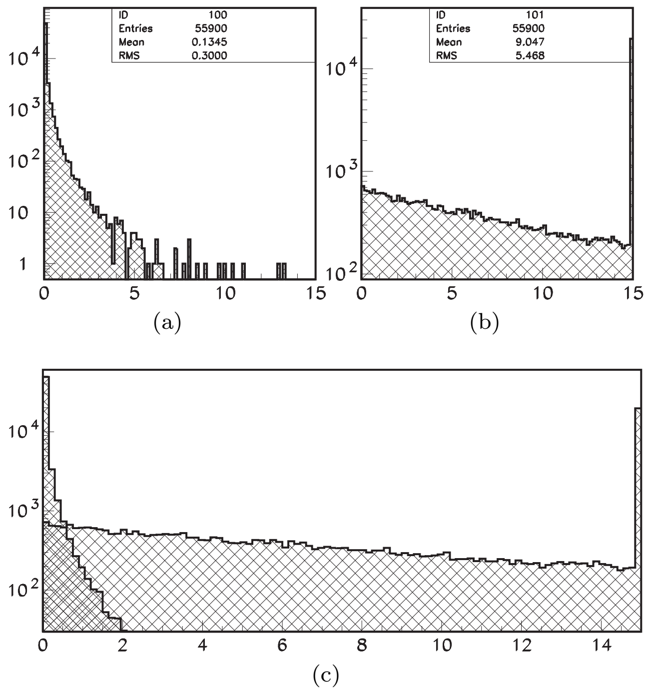

Figure 6. **Distributions of ω_{12}^{6} values for π (a) and for e (b) events; the summary distribution for π and e events (c)**

The main idea of this scheme consists in the following:

– using the mean value method, we collect in the admissible region the main part of electrons together with a small admixture of pions,
– then we apply the ω_n^k test to the events selected in the admissible region; this way, we loose a small part of electrons together with additional suppression of pions.

5. Modified ω_n^k Test

Fig. 7 shows the dE/dx distribution of e, and Fig. 8 presents the distribution of electron energy losses on the transition radiation (TR).

Fig. 9 shows the probability of events with a different number of TR counts. We see that the most probable value of TR counts in the TRD with 12 layers is 6, and we almost do not have the events with TR counts in all 12 layers.

Let us turn back to the distribution of electron energy losses on the transition radiation (see Fig. 8). Fig. 8 shows that when e passes the i-th layer with TR=0, then its energy loss follows the distribution of dE/dx losses (Fig. 7). In this case, it is practically impossible to distinguish electrons from pions on the basis of their energy losses. In the opposite case, when we have the TR count in the i-th layer, the electron energy loss will correspond to the sum of $\mathrm{d}E/\mathrm{d}x + TR$. Only such counts in TRD layers may permit us to distinguish electrons from pions.

When calculating ω_n^k, in formula (4), one uses a sample of λ_i values (see Eq. 3), which are ordered due to their values. The λ_i value is directly proportional to the energy loss by a particle registered in the i-th layer of the TRD. In this connection and taking into account that the most probable value of TR counts in the TRD with 12 layers is 6 (Fig. 9), we may use in the ω_n^k test only that part of $\{\lambda_i\}$ sample which corresponds to indexes $i > 6$, i.e. to large values of particle energy losses.

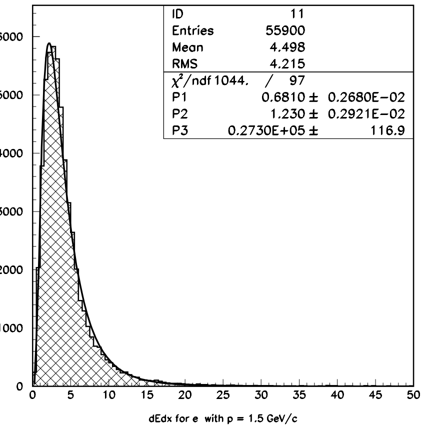

Figure 7. **Distribution of electron energy losses on the ionization and its approximation by a log-normal function**

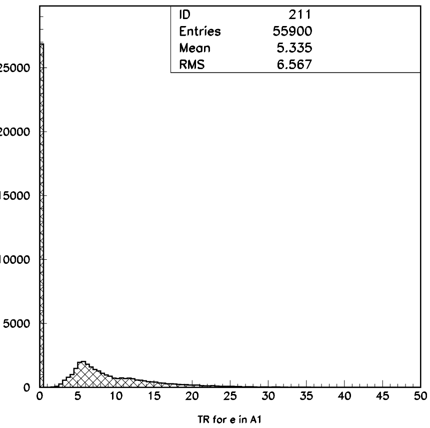

Figure 8. **Distribution of electron energy losses on the transition radiation**

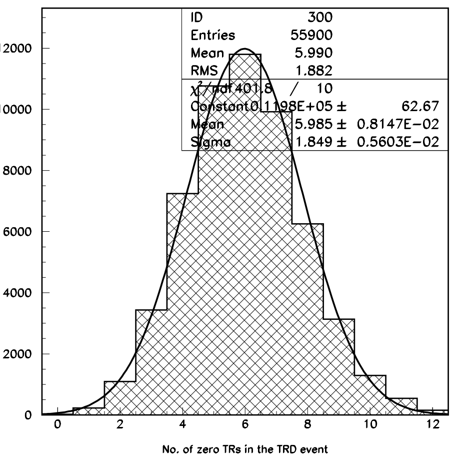

Figure 9. **Distribution of events with different number of TR counts and its approximation by Gaussian distribution**

Fig. 10 shows the distributions of ω_n^k values for GEANT samples: $i = 7$, $n = N_{layers} - i + 1$, $k = 6$.

6. Method Based on Artificial Neural Network (ANN)

In [6] we applied a three-layered perceptron from the packages JETNET3.0 and ROOT to estimate the efficiency of PID. Initially the training patterns were formed using the set of energy losses ΔE_i, $i = 1, \ldots, n$ corresponding to the passage through the TRD pions or electrons.

In spite of the fact that the distribution of the energy losses by electrons significantly differs from the character of the distribution of the energy losses by pions, for such a choice of input data the training process was going on very slow (see bottom

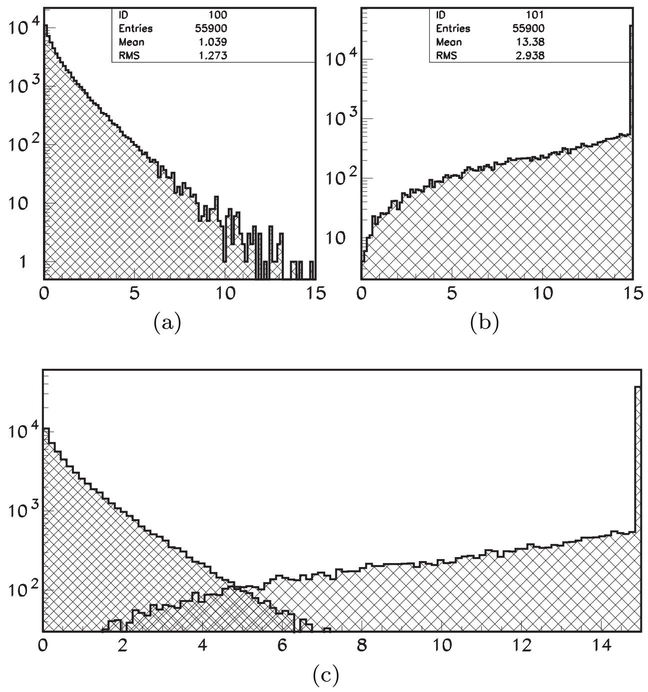

Figure 10. **Distributions of modified ω_n^k values calculated for π (a) and for e (b) events; the summary distribution (c)**

curve in 11), there were large fluctuations (against the trend) of the efficiency of events identification by the network. Moreover, one cannot reach the needed level of pions suppression.

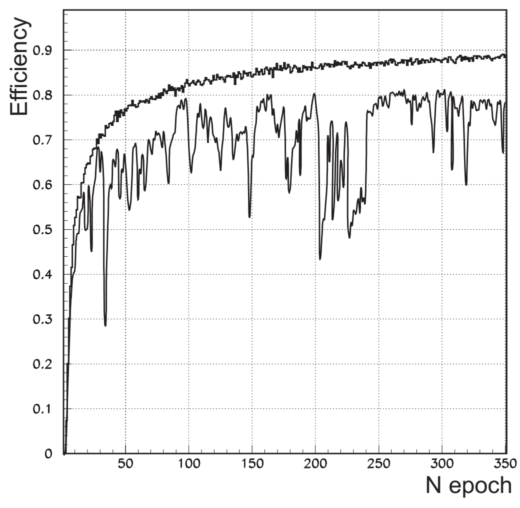

Figure 11. **The efficiency of pion/electron identification by the MLP for original (bottom curve) and transformed (top curve) samples**

Aiming to improve the situation, we decided to apply as input sets the sets of variable λ (3): see top curve in Fig. 11. Fig. 12 shows distributions of the MLP output signals obtained at the training (b) and testing (d) stages; the left plots show the distributions of errors between the target value and the MLP output signal at the training (a) and testing (c) stages.

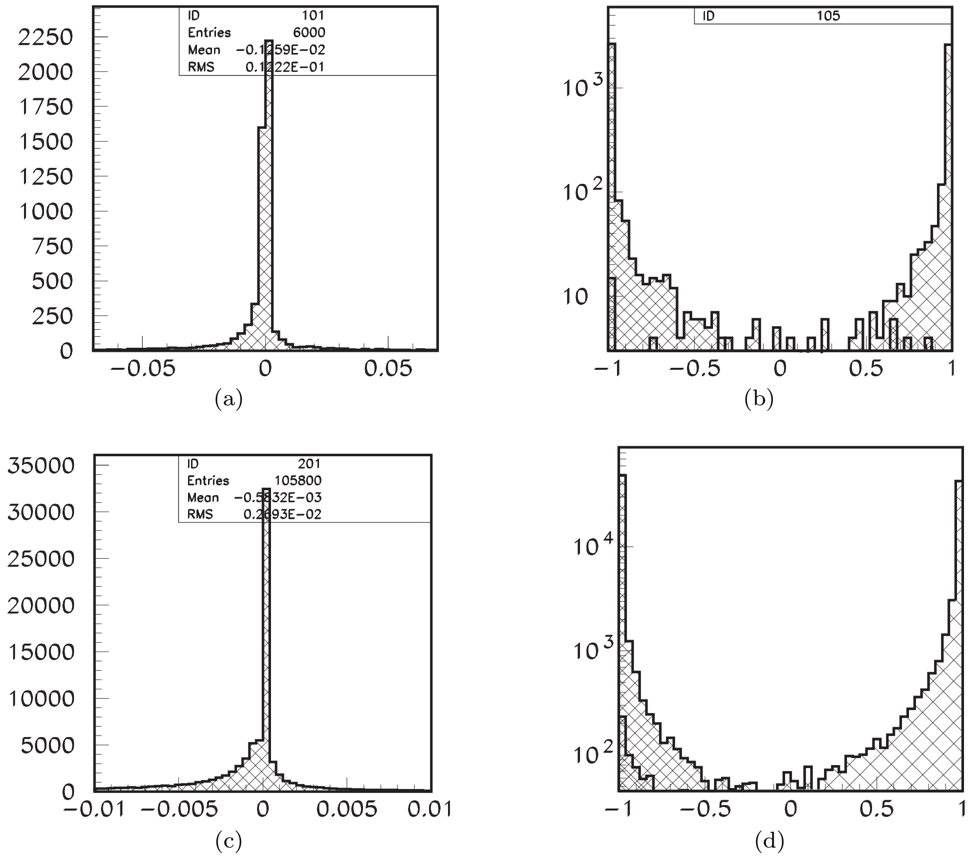

Figure 12. **Distributions of errors (a and c); the distributions of the MLP output signals (b and d)**

7. Conclusion

Table 1 shows the results of comparison of the given methods for different particle (pion and electron) momenta.

Table 1

Comparison of the given methods: pion suppression factor for different momenta

$p, GeV/c$	1	2	3	4	5	7	9	11
LFR	1273	1315	1581	1480	936	861	800	749
MV	15	17	17	16	16	16	16	16
ω_n^k	104	96	73	55	44	35	29	25
MV + ω_n^k	284	271	249	242	203	198	157	159
$\mathrm{mod}\,\omega_n^k$	296	621	628	776	650	745	588	537
MV + $\mathrm{mod}\,\omega_n^k$	311	515	518	493	438	606	398	413
root	1219	1400	1112	1446	730	1054	610	882
jetnet	1857	1837	1378	1713	1446	1317	1045	1089

This table demonstrates that the MV method and the ω_n^k criterion do not provide a required level of the pion suppression ($\sim$ 100-150). The main cause is that the electron energy losses are not described by a single distribution, but bear a character of a composite hypothesis.

The criteria simple from a practical viewpoint (the modified ω_n^k criterion and the composite criteria based on MV + ω_n^k criterion) provide high levels of the pion suppression.

One succeeds in reaching the best pion suppression level using: a) ANN when transmitting from the initial energy losses in the TRD layers to a new variable typical for the ω_n^k criterion, and b) LFR method with the energy losses approximated by a lognormal distribution for pions and by a weighted sum of two lognormal distributions for electrons.

References

1. Letter of Intent for the Compressed Baryonic Matter experiment. — `http://www.gsi.de/documents/DOC-2004-Jan-116-2.pdf`.
2. The CBM Collaboration. CBM Compressed Baryonic Matter Experiment. Technical Status Report: Techrep / GSI. — Darmstadt, 2005. — `http://www.gsi.de/onTEAM/dokumente/public/DOC-2005-Feb-447e.html`.
3. Statistical Methods in Experimental Physics / W. T. Eadie, D. Dryard, F. E. James et al. — Amsterdam-London: North-Holland Pub. Comp, 1971.
4. *Zrelov P. V., Ivanov V. V.* The Relativistic Charged Particles Identification Method based on the Goodness-of-Fit ω_n^3-Criterion // Nucl. Instr. and Methods in Phys. Res. — 1991. — Vol. A310. — Pp. 623–630.
5. Electron/Pion Identification in the CBM TRD Applying a ω_n^k Goodness-of-Fit Criterion / E. P. Akishina, T. P. Akishina, V. V. Ivanov et al. // Particles and Nuclei, Letters. — 2008. — Vol. 5, No 2. — Pp. 202–218.
6. *Акишина Т. П., Дереновская О. Ю., Иванов В. В.* Об идентификации электронов и пионов с помощью многослойного перцептрона в детекторе переходного излучения эксперимента CBM // Вестник РУДН, Серия «Математика. Информатика. Физика». — 2010. — № 1. — С. 100–109.

УДК 004.93'1:519.237:51-72

Методы идентификации e/π с помощью детектора переходного излучения в эксперименте CBM

Е. П. Акишина, Т. П. Акишина, О. Ю. Дереновская, В. В. Иванов

Лаборатория информационных технологий
Объединённый институт ядерных исследований
ул. Жолио-Кюри, д.6, Дубна, Московская область, 141980, Россия

Рассмотрена задача идентификации электронов и пионов в детекторе переходного излучения (TRD) эксперимента CBM. С этой целью разработаны разные алгоритмы и подходы. Обсуждаются характерные свойства энергетических потерь электронов и пионов в слоях TRD и особенности применения искусственных нейронных сетей (ИНС) и статистических методов для решения рассматриваемой проблемы. Проведен сравнительный анализ мощностей статистических критериев и ИНС.

Ключевые слова: статистические методы анализа данных, многомерные методы анализа данных, методы распознавания образов, эксперимент CBM, детектор переходного излучения TRD.