CPC(HEP & NP), 2010, **34**(10): 1632–1635 **Chinese Physics C** Vol. 34, No. 10, Oct., 2010

Photon discrimination simulation with the PCA method*

ZHOU Jing(周静)[1)] LIU Zhi-Yi(刘志毅) MENG Qiu-Ying(孟秋英) ZHOU Shu-Hua(周书华)

China Institute of Atomic Energy, Beijing 102413, China

Abstract The shower shape of n, $\bar{\text{n}}$, p, $\bar{\text{p}}$, K^+, π^+ and photons, generated by JPCIAE code for 5.5 TeV/A ^{208}Pb+^{208}Pb collisions, incident on the ALICE photon spectrometer (PHOS), is analyzed with the principal component analysis (PCA) method. The efficiency dependence of purity for the photon discrimination is simulated for the deposited energy ranges 0.5–2 GeV, 2–10 GeV, 10–50 GeV and 50–100 GeV. The result shows that in the energy range of 0.5 to 100 GeV, the efficiency of the photon identification can reach 90% with purity of 90%.

Key words photon discrimination, PCA, ALICE, PHOS

PACS 02.90.+p, 29.40.Vj, 29.40.Gx

1 Introduction

The shower shape analysis is an important technique in processing data in high energy experiments, where an electromagnetic calorimeter (EMC) is used to measure the photon 4-momentum in a high particle multiplicity environment. Two methods - topology and principal component analysis (PCA) - can be employed to discriminate photons from other particles by analyzing their shower shapes in the calorimeter. In principle, using the topology method to define selection criteria on the multi-dimensional space composed of shower shape parameters should result in good identification. However, one usually needs to use as many as 7 or more parameters for the shower shape description and these parameters are often correlated with each other, thus the topology method is a tedious task with no guarantee of finding the absolute minimum in the multi-dimensional space. The PCA is a dimensionality reduction method, which has found many applications in the field of pattern recognition, information transmission and data classification [1, 2]. The basic idea of PCA is that by a linear transformation of the original parameters to a new set of uncorrelated and optimized ones, one can choose the most significant parameters (principal components) which contain most information, and drop the rest parameters to reduce the dimension of data.

In our earlier work [3], we compared the topology (7 parameters) method with the PCA (2 components) method in the analysis of shower shapes in a simple case, where the showers were produced by n, $\bar{\text{n}}$, p, $\bar{\text{p}}$, K^+, π^+ and photons, 10000 each with random energy distributions from 0.5 GeV to 100 GeV, incident on the ALICE photon spectrometer (PHOS). The results show that the two methods are comparable in the photon identification.

In order to study the photon identification power of PHOS in the ALICE experiment at the Large Hadron Collider (LHC) environment, we have simulated the response of PHOS to different kinds of particles generated by JPCIAE code [4, 5] for 5.5 TeV/A $^{208}\text{Pb} + {}^{208}\text{Pb}$ collisions and applied the PCA method to the shower shape analysis.

2 The structure of PHOS

Details about the ALICE experiment can be found elsewhere [6, 7]. Briefly, ALICE is a large ion collider experiment at the LHC at CERN, aiming to study the behavior of matter under extreme conditions of high temperature and high density, probing a new state of matter - quark gluon plasma (QGP) and investigat-

Received 2 November 2009

* Supported by National Natural Science Foundation of China (00121140488)

1) E-mail: jzh@ciae.ac.cn

ing the mechanism of QCD confinement and chiral symmetry restoration.

The PHOS is composed of time of flight (TOF), a charged particle veto detector (CPV) and an electromagnetic calorimeter (EMC). The EMC is divided into five modules each equipped with 64 (across the beam direction) by 56 (along the beam direction) Lead-Tungstate ($PbWO_4$) scintillator crystals. Each crystal is an 18 cm long parallelepiped, equivalent to 20 units of radiation length, and has a squared cross-section of 22 mm×22 mm, corresponding to 90% of the Molière radius. The PHOS is located inside a magnet at a distance of 4.6 m to the collision point. Its role is to detect and identify real photons and measure with high resolution their 4-momentum with the prospect of performing the physics programs exploiting direct photon and light neutral mesons as privileged probes. The unique ability of the PHOS to measure identified photons over a broad dynamic range in transverse momentum will enable one to access key information about the soft and hard processes occurring in pp and AA collisions at LHC energies.

3 Simulation

The simulation is performed based on the ALIROOT platform, and consists of three steps: simulation of the output signals of PHOS, reconstruction algorithm and particle identification. Heavy particles and charged particles are suppressed by setting proper cuts for TOF and CPV. But still some heavy particles and charged particles can enter EMC and develop showers. Then the PCA shower shape analysis is used to identify photons from other particles.

In the simulation of the PHOS output signals, we make seven kinds of particles: n, $\bar{\mathrm{n}}$, p, $\bar{\mathrm{p}}$, K^+, π^+ and photons generated by JPCIAE for 5.5 TeV/A ^{208}Pb+^{208}Pb collisions incident on the PHOS. The energy spectra of these particles are shown in Fig. 1. For each particle, the deposited energy, position, time signal of the fired cell and incident primary particle number are tracked. We define the EMC Hit as a summed deposited energy by a given primary particle in a single cell. Hits below 1 MeV lose their reference to the primary particle and are called SDigits. Gaussian noise of 4 MeV (EMC) and 0.01 (CPV) is added to all cells. SDigits in the same cell are merged into one Digit. Digits below 12 MeV (EMC) and 0.09 (CPV) are removed.

The reconstruction algorithm processes the EMC and CPV digits in three steps to produce reconstructed points, track segments and reconstructed particles. The first step is performed separately on EMC digits and CPV digits, whereas the two last steps combine the information collected by the two detectors.

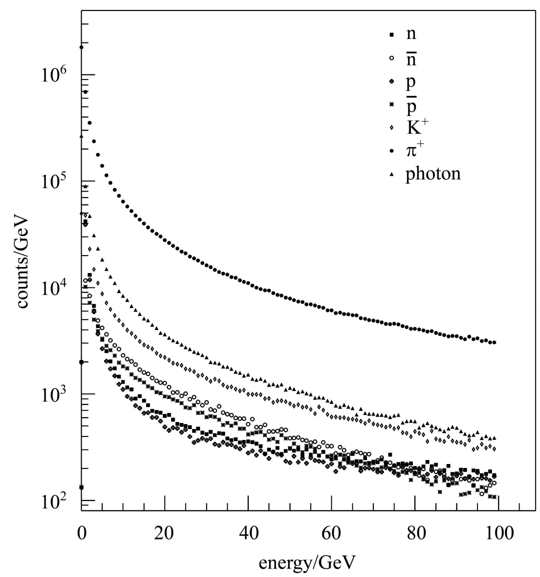

Fig. 1. Energy spectra of n, $\bar{\mathrm{n}}$, p, $\bar{\mathrm{p}}$, K^+, π^+ and photons generated by JPCIAE for 5.5 TeV/A ^{208}Pb+^{208}Pb collisions.

Cells with a common edge or corner are grouped into a cluster (reconstructed point). Clusters with several local maxima are unfolded. The energy e and the position of the clusters in (x, z)-plane of the module reference frame are calculated, respectively, as the sum of the digit energies e_i, and the center of gravity with a logarithmic weight,

$$\omega_i = \max\left[0, p + \lg\frac{e_i}{e}\right], \tag{1}$$

where e_i is the ith digit energy and p is an empirically determined parameter.

Track segments are made up from at most two reconstructed points: one in CPV and one in EMC within 10 cm.

Reconstructed particles are constructed from track segments and reconstructed points.

The shower shape in the calorimeter is characterized by the following 7 parameters:
lateral dispersion d,

$$d = \frac{\sum\limits_{\text{digits}} \omega_i[(x_i - x)^2 + (z_i - z)^2]}{\sum\limits_{\text{digits}} \omega_i}, \tag{2}$$

where x_i and z_i are the coordinates of the ith cell fired by the shower.
λ_1 and λ_2 being the two major axes of the ellipse

of the intersection of the shower cone and the EMC surface.

Sphericity parameter S,

$$S=\frac{|\lambda_1-\lambda_2|}{\lambda_1+\lambda_2}. \tag{3}$$

The largest fraction of energy deposited in a single crystal core energy corresponding to the summed energy of digits within a given radius ($R_{\rm core}$=3 cm) around the largest digit.

Multiplicity of digits $M_{\rm Digits}$ in the shower.

Obviously, not all of the 7 parameters are independent. A set of 7 statistically independent parameters, referred to as the principal components, is obtained by diagonalizing the covariance matrix of the original 7 parameters listed above. Thus, showers produced by photons can be recognized in the space of the principal components with the PCA approach. The two most significant principal components, i.e. those two components which correspond to the largest eigenvalues of the covariance matrix p_0 and p_1, are used to identify the showers in a 2-dimensional space spanned over these 2 principal components. The resulting ellipse like 2-dimensional distributions (see Fig. 2) are fitted by a 2-dimensional Gaussian function,

$$f(p_0) = \exp(-R(p_0,p_1)), \tag{4}$$

$$R(p_0,p_1) = \left(\frac{p_0-x_0}{a}\right)^2+\left(\frac{p_1-x_1}{b}\right)^2 +C\left(\frac{(p_0-x_0)(p_1-x_1)}{ab}\right), \tag{5}$$

where parameters a, b, C, x_0 and x_1 depend on the reconstructed energy. x_0 , x_1 are the coordinates of the center and C is the orientation of the ellipse.

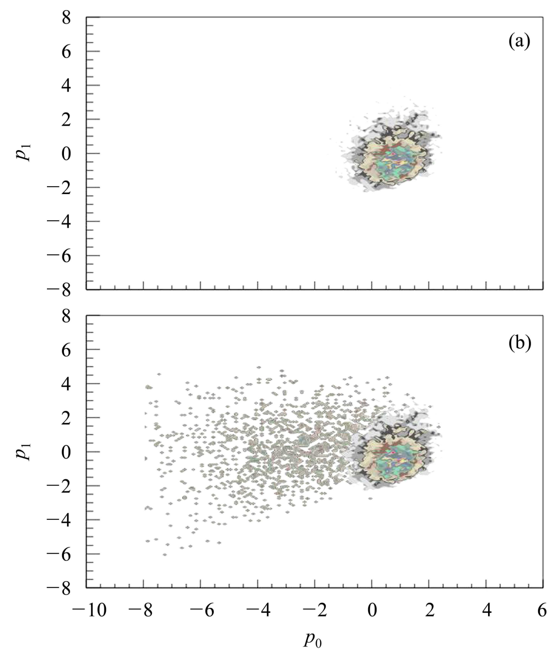

Fig. 2. Shower shape in the principal component space for photons only (a), and for n, $\bar{\rm n}$, p, $\bar{\rm p}$, $\rm K^+$, π^+, plus photons (b).

4 Results and conclusion

We define the photon identification efficiency as the number of incident photons identified as photons

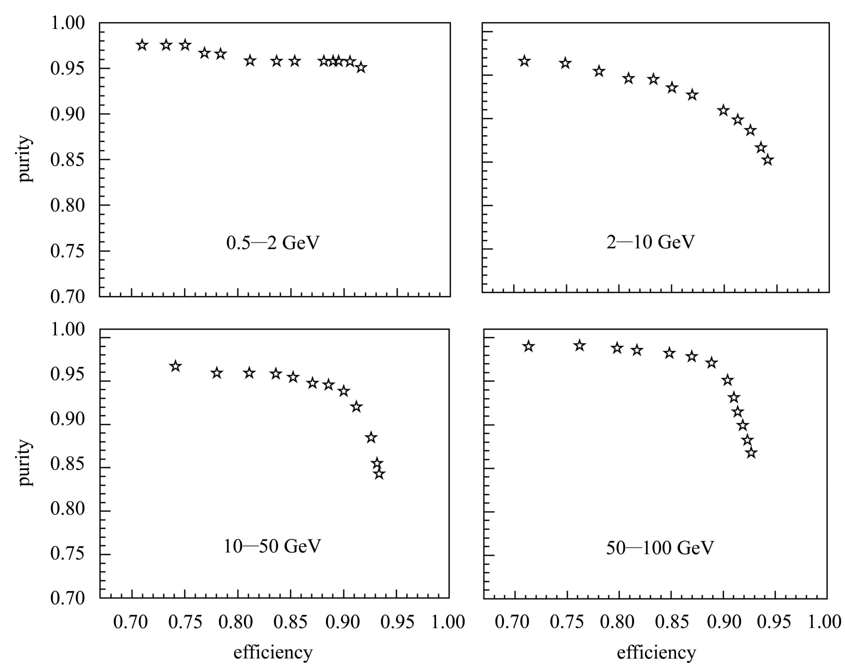

Fig. 3. Efficiency dependence of purity for the photon identification of the PHOS in 4 different energy regions.

divided by the total number of incident photons, and the purity as the number of incident photons identified as photons divided by the number of the seven kinds of incident particles (n, n̄, p, p̄, K^+, π^+, and photons) identified as photons. The photon identification purity as a function of the efficiency is obtained in 4 deposited energy regions, 0.5–2 GeV, 2–10 GeV, 10–50 GeV and 50–100 GeV, as shown in Fig. 3. The results are different from that for particles with random energy distributions [3]. In particular, for particles generated by JPCIAE, in 0.5–10 GeV energy regions the purity is much higher and in the 50–100 GeV region the purity decreases drastically when the efficiency exceeds 0.9. So, using realistic energy distribution in the LHC environment is necessary in evaluating the PHOS performance.

The authors wish to thank Prof. Sa Benhao for his help in using JPCIAE code and to Prof. Li Xiaomei in using ALIROOT.

References

1 Gonzalez Rafael C, Woods Richard E. Digital image processing. Addison Wessley Publishing Company, 1992
2 Wind H. Function Parameterization, CERN 72–21
3 ZHOU J et al. Nuclear Physics Review, 2004, **21**: 12
4 SA B H. Phys. Rev. C, 1999, **59**: 2728
5 SA B H, TAI A. Phys. Rev. C, 2000, **62**: 044905
6 ALICE collaboration. Technical Proposal [Z], CERN/ LHCC/ 95-71, 1995
7 ALICE collaboration. Photon Spectrometer (PHOS) [Z]. CERN/LHCC/ 99-4, ALICE TDR 2, 5, March 1999