An artificial neural network for proton identification in HERMES data*

WANG Si-Guang(王思广)[1)] MAO Ya-Jun(冒亚军) YE Hong-Xue(叶红学)

(School of Physics and State Key Laboratory of Nuclear Physics & Technology, Peking University, Beijing 100871, China)

Abstract The HERMES time-of-flight (TOF) system is used for proton identification, but must be carefully calibrated for systematic biases in the equipment. This paper presents an artificial neural network (ANN) trained to recognize protons from Λ^0 decay using only raw event data such as time delay, momentum, and trajectory. To avoid the systematic errors associated with Monte Carlo models, we collect a sample of raw experimental data from the year 2000. We presume that when for a positive hadron (assigned one proton mass) and a negative hadron (assigned one π^- mass) the reconstructed invariant mass lies within the Λ^0 resonance, the positive hadron is more likely to be a proton. Such events are assigned an output value of one during the training process; all others were assigned the output value zero.

The trained ANN is capable of identifying protons in independent experimental data, with an efficiency equivalent to the traditional TOF calibration. By modifying the threshold for proton identification, a researcher can trade off between selection efficiency and background rejection power. This simple and convenient method is applicable to similar detection problems in other experiments.

Key words artificial neural network, particle identification, TOF

PACS 07.05.Mh, 82.80.Rt

1 Introduction

The HERMES (HERA MEeasurement of Spin) experiment uses 27.6 GeV/c electrons or positrons from the HERA accelerator (DESY laboratory, Hamburg) to study the quark-gluon spin structure of nucleons by deep inelastic scattering[1, 2].

The HERMES spectrometer uses two detectors to identify hadrons. A dual-radiator ring-imaging Čerenkov detector (RICH)[3] takes care of hadrons with momenta above 2.0 GeV/c. The other detector is a time-of-flight (TOF) system capable of identifying hadrons below this limit. As the TOF system both improves statistics and increases the overall range, tuning it carefully is certainly worthwhile.

The TOF system has already been calibrated in the traditional manner[4], but the process is complicated and time-consuming. As an alternative, we propose a method using artificial neural networks (ANN). The goal is to distinguish protons from other particles based only on track characteristics such as momentum and trajectory.

An ANN must be provided with a training sample: data where the identities of the particles are known. Monte Carlo (MC) data are often used for this purpose, since event parameters for protons and non-protons can easily be simulated. If the MC data do not correctly model the detector performance, however, systematic errors will be introduced. To avoid this problem, the method introduced here uses raw experimental data to train the ANN.

2 HERMES spectrometer

The structure of the HERMES spectrometer is described in Fig. 1.

Received 3 July 2008, Revised 30 July 2008

* Supported by National Science Foundation of China (10775006, 10375002, 10675004) and Doctoral Program Foundation of Institutions of Higher Education of China (20070001008), and China Postdoctoral Science Foundation

1) E-mail: siguang@hep.pku.edu.cn

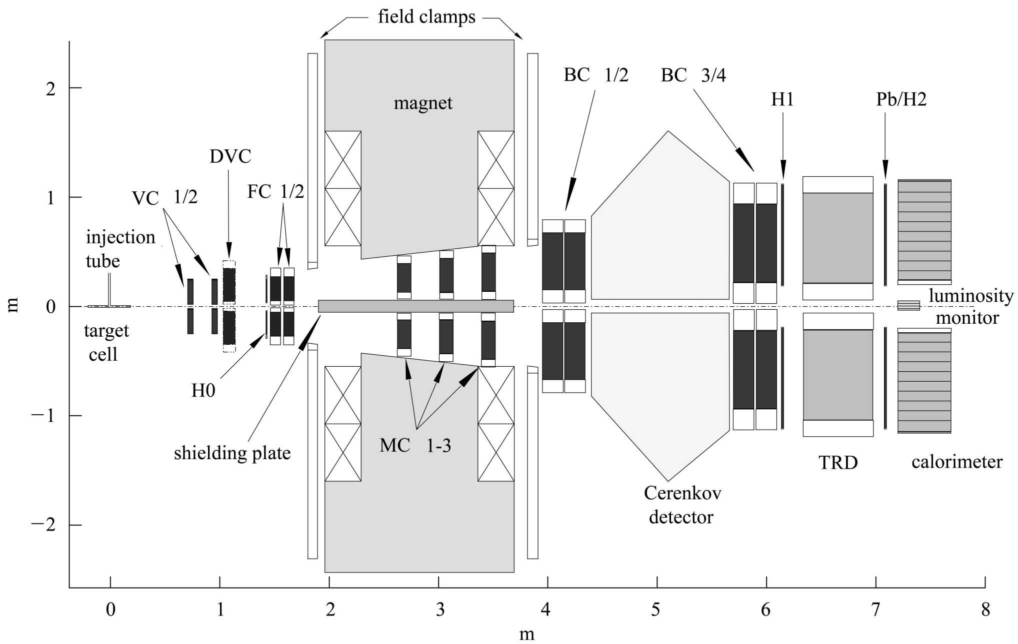

Fig. 1. Side view of the HERMES spectrometer. See the text for the meanings of the labels.

Physical models of particle data rely on the production vertices and momenta of decay products. The trajectory of the particle is measured by microstrip gas chambers (MSGC), also referred to as vertex chambers (VC), as well as drift chambers in front of the magnet (DVC, FC1/2) and behind the magnet (BC1/2 and BC3/4). In addition, there are three proportional chambers inside the magnet (MC1/3). These help match the front and back tracks, and also detect low-momentum particles that do not reach the spectrometer. The momentum of a charged particle can be calculated from the curvature of its trajectory inside the magnet, whose field is known. Several detectors act in concert to distinguish leptons from hadrons: a lead-glass calorimeter; a pre-shower detector (H2) consisting of two radiation lengths of lead followed by a plastic scintillator hodoscope; a transition radiation detector (TRD) consisting of six identical modules; and the RICH detector. The calorimeter and preshower detector are included in the trigger, along with a second hodoscope (H1) placed in front of the TRD[5].

Both H1 and H2 are composed of vertical scintillator modules (each has 84, divided equally between the upper and lower detectors). The scintillation light from each module is detected by a photomultiplier tube (PMT). The combined PMT signal from a hodoscope is passively split, with one output going to a LeCroy 1881M ADC and the other going to a LeCroy 3420 constant fraction discriminator (CFD). Individual CFD outputs are fed into time-to-digital converters (TDC)[4]. This result can be used as a stop time for the particle. The moment when the HERA bunch crosses the center of the target is used as the start time. Together, these two measurements are used to calculate the HERMES TOF. The HERA beam has a bunch length of 27 ps and the distance between bunches is 96 ns, so particles from different bunches are completely separated.

3 Hadron identification by TOF

We now describe how to calibrate the TOF system for particle identification.

3.1 Traditional calibration

Previous calibration work done by our colleagues is based on the fact that electrons above 10 MeV are moving at essentially the speed of light. Any measured deviations from this speed must therefore be due to artificial limitations of the experiment, and should be corrected. Several corrections have been taken into account: different cable lengths, variance in the time needed for light created inside a scintillator to reach the PMT, different PMT response times, and others. This method places an upper limit of 2.9 GeV/c on the momentum of protons that can be identified by TOF [4].

One fact has not been corrected in the current calibration. The magnetic field has a deflecting power of $\int B\mathrm{d}l = 1.3\ \mathrm{T\cdot m}$[5], so low-momentum charged particles follow a curved trajectory inside the magnet. Within the magnetic field, they are deflected through a total angle $|\phi|$:

$$|\phi| = \frac{0.2998}{p}\int B\mathrm{d}l\,. \tag{1}$$

The angle is expressed in radians, and the momentum in GeV/c. These particles can be identified by the TOF system, but calculation of their actual trajectories would be very complicated. Instead, the traditional calibration method approximates the trajectory with two straight lines: one from the front vertex to the center of the spectrometer magnet, and the other from this point to the H1 and H2 detections. This procedure clearly introduces some error into the TOF calculation.

To avoid such corrections and difficulties, we have developed a new calibration method using artificial neural networks (ANN).

3.2 ANN approach

This section describes the structure of the ANN and its performance based on training and experimental data.

3.2.1 Structure of the ANN

The momentum of a relativistic particle can be written as

$$p = m_0\frac{\beta}{\sqrt{1-\beta^2}}\,, \tag{2}$$

where m_0 is the rest mass and $\beta = v/c$ can be calculated from the time of flight (TOF) and trajectory length (between the target cell and the position of hodoscope H1 or H2).

Likewise, the rest mass can be expressed as

$$m_0^2 = p^2\left(\frac{1}{\beta^2}-1\right). \tag{3}$$

The TOF system uses this measure of m_0^2 to identify particles.

The ANN takes primary event parameters (momentum, time of flight and trajectory) as inputs, and classifies the particle as either a proton or a non-proton. All these input parameters are available in HERMES spectrometer records. The ANN must be trained on events where the identity of the particle is known.

Consider an event with two particles: one is a positively charged hadron (assigned one proton mass), and the other is a negative hadron (assigned one π^- mass). If their reconstructed invariant mass falls within the Λ^0 resonance, there is a high probability that the positive hadron is a proton. If the reconstructed invariant mass falls outside the Λ^0 peak, the probability is considerably lower.

Our ANN code is taken from the ROOT platform[6]. The network has two hidden layers, as shown in Fig. 2.

The ANN is composed of "neurons" (the nodes in Fig. 2), which are characterized by a threshold activation function. Every neuron is linked to all the neurons in the two adjacent layers (we call these links "synapses"). Each synapse is characterized by a single numerical weight, so each neuron receives a weighted sum of the prior layer's outputs. The first layer is composed of inactive neurons, and just serves to pass on the input data. The hidden neurons use a sigmoid activation function in our work. The output

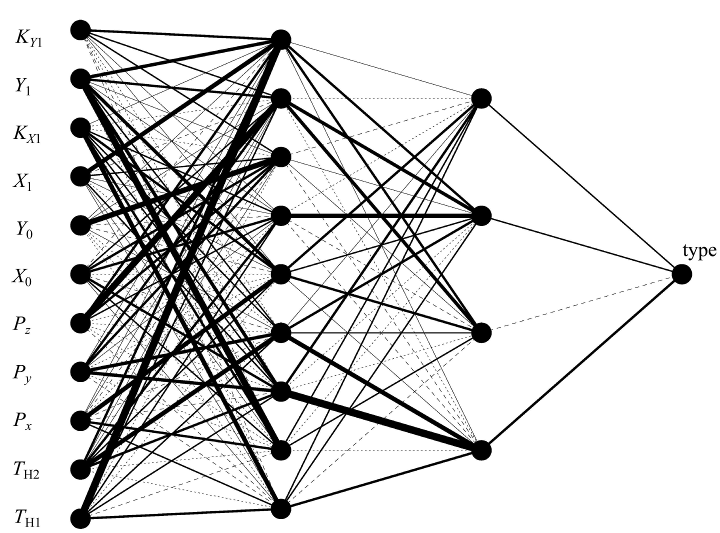

Fig. 2. The artificial neural network used for proton identification. Line widths are used to indicate synapse weights.

neuron produces a signal directly proportional to its inputs. We use the learning method of Broyden, Fletcher, Goldfarb and Shanno (BFGS) to train the network (this is an algorithm for adjusting the weights in response to training data). For more information about ANNs, we refer the reader to the "TMultiLayerPerceptron" class description in the ROOT platform[6].

The input parameters are: $T_{\rm H1}$, the time of flight measured by H1; $T_{\rm H2}$, the time of flight measured by H2; P_x, P_y and P_z, momentum components of the particle in the laboratory frame; X_0 and Y_0, coordinates of the target cell centroid in the $z=0$ plane, perpendicular to the HERA beam; X_1 and Y_1, coordinates of the point where the particle passed through the z-plane in the middle of the spectrometer magnet; K_{x1} and K_{y1}, the slopes dx/dz and dy/dz of the particle trajectory after passing through the z-plane in the middle of the magnet ($z=275$ cm).

The inputs are automatically normalized by the ANN before beginning the training process. For each parameter x the normalized values x' are $x' = \frac{x_i - \text{mean}}{\text{RMS}}$, where the mean and RMS (root mean square, or standard deviation) are calculated over all examples in the training data.

3.2.2 Training the ANN

The ANN was trained using HERMES data taken in the year 2000 (run numbers less than 5000). We select events with two oppositely charged tracks, assign a proton mass to one and a π^- mass to the other, then reconstruct the invariant mass. If the reconstructed mass falls under the Λ^0 resonance we assign its Type to 1, meaning that the positively charged particle is a proton. Otherwise we assign its Type to 0. To suppress background events, we impose two cuts: (1) the Λ^0 candidate must have flown at last 10 cm from the interaction point before decaying, and (2) the distance of closest approach (DCA) for the two tracks must be less than 1 cm. Due to background events, it is still possible for a positively charged particle to be misidentified under this system. The particles which have been assigned Type 1, however, are much more likely to be protons than those which have been assigned Type 0. (This statement will be quantified in the next section.) The ANN is used to figure out which patterns of input variables correspond best to the two particle types.

Once the ANN has been trained, the same event parameters can be used by the ANN to classify other particles. The ANN output is a continuous number between 0 and 1. Since all the training events are assigned values of either 0 or 1, the actual outputs should be close to one of these extremes. The closer the output is to 1, the higher the probability that the particle is a proton. If the value is close to 0, it is more probable that the particle is not a proton. For

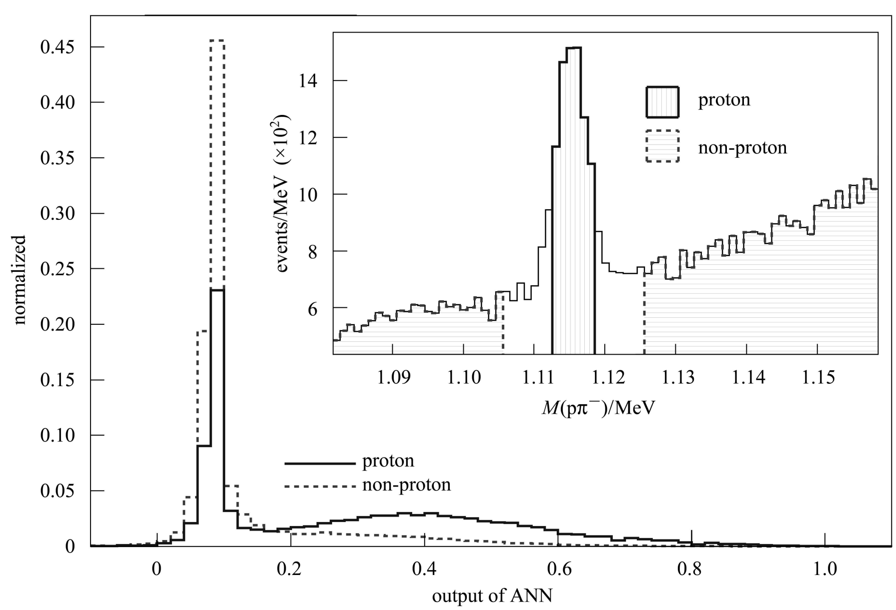

Fig. 3. Output of the trained ANN. The inset histogram shows how the training sample was selected. Events with a reconstructed invariant mass under the Λ^0 resonance (the region filled with vertical lines) were assigned Type 1 (proton). Events with a reconstructed invariant mass outside the resonance (the region filled with horizontal lines) were assigned Type 0 (non-proton). The outputs produced by the trained ANN on these same events are described by the main histogram and normalized separately.

more details we refer the reader to Fig. 3, which presents the normalized distribution of training data outputs obtained after training the ANN. The mean value of the output distribution for particles which have been assigned type 1 is clearly higher than that for particles assigned 0. The strong peak in this distribution corresponds to misclassified background events, while the low, broad peak corresponds to actual protons. If the output of the trained ANN is greater than 0.2 (see Fig. 3), the particle is more than 50% likely to be a proton. These distributions can be used to calculate the probability as a function of output value.

The various inputs have different impacts on the ANN output, as shown in Fig. 4. The influence of a variable on the result is evaluated by shifting its value by ±0.1RMS while holding the others constant, then calculating the difference between the two outputs. This is done for each event in the training sample. The resulting distributions are drawn in Fig. 4, and their summary statistics are listed in Table 1.

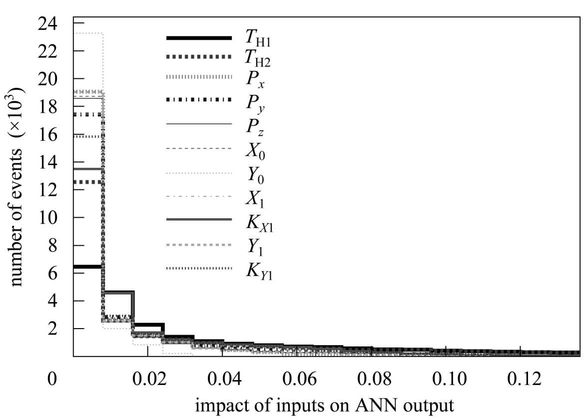

Fig. 4. Distributions of the change ("impact") in network output resulting from a small variation (±0.1RMS) in each input, while holding other variables constant. Summary statistics of the histograms are listed in Table 1.

Table 1. Mean and RMS values of the impact distributions in Fig. 4.

inputs	mean	RMS
T_{H1}	0.06480	0.09211
T_{H2}	0.03356	0.04785
P_x	0.00886	0.01344
P_y	0.01296	0.02153
P_z	0.01134	0.02022
X_0	0.01054	0.01813
Y_0	0.00289	0.00498
X_1	0.01381	0.02176
K_{X1}	0.01937	0.02681
Y_1	0.00912	0.01603
K_{Y1}	0.01616	0.02385

Flight time values (H1 and H2) have the greatest impact on the ANN, which is reasonable because the particle mass is correlated with momentum and with therefore flight time (see Eq. (3)). The least important variable is Y_0 (at $z = 0$ m), because this point is very far from the two detectors (which lie around $z \approx 7$ m, as shown in Fig. 1). Most of the Y_0 inputs are therefore distributed in the range ±2 cm, and this coordinate has little impact on the total length of the particle trajectory.

3.2.3 Applying the trained ANN

An independent sample of data taken in the year 2000 (run numbers from 5001 to 10000) was used to evaluate the trained ANN. The event selection process was much the same: positively charged hadrons were assigned a proton mass and negatively charged hadrons within the same event were assigned a π^- mass. The invariant mass was reconstructed to identify Λ^0 candidates. We required a flight distance greater than 10 cm for the reconstructed Λ^0 particles, and a DCA less than 1.0 cm for the decay products. The expected number of protons was taken by fitting the reconstructed invariant mass spectrum with a Gaussian distribution and a linear background, then integrating the Gaussian within $\pm 2\sigma$ from the mean. Fig. 5 shows three spectra of the positive hadron masses obtained in this sample. The first panel (a) includes all selected events, without any particle identification. The middle panel (b) shows only those protons identified by the traditional TOF calibration. The third spectrum (c) includes protons identified by the trained ANN.

The two peaks shown in panels (b) and (c) of Fig. 5 are very similar in shape, as the fit parameters indicate. When we compare individual events in the invariant mass range 1110.89—1119.65 MeV ($\pm 1\sigma$), however, we find that only 90.77% of the events are classfied the same way by both methods. Furthermore, the ANN method is clearly as effective as the traditional method at suppressing background events.

Further investigation indicates that the higher we place the threshold f_0 for proton identification, the more background events are rejected. The selection efficiency also decreases, of course. Unfortunately, it is not possible to obtain a 100% pure data sample. To get a rough idea of the relationship between rejection power and selection efficiency, we ran f_0 through a range of values and recorded the following information:

1. The number of Λ^0 events obtained after TOF identification, compared to that obtained without any identification ("Selection Efficiency").

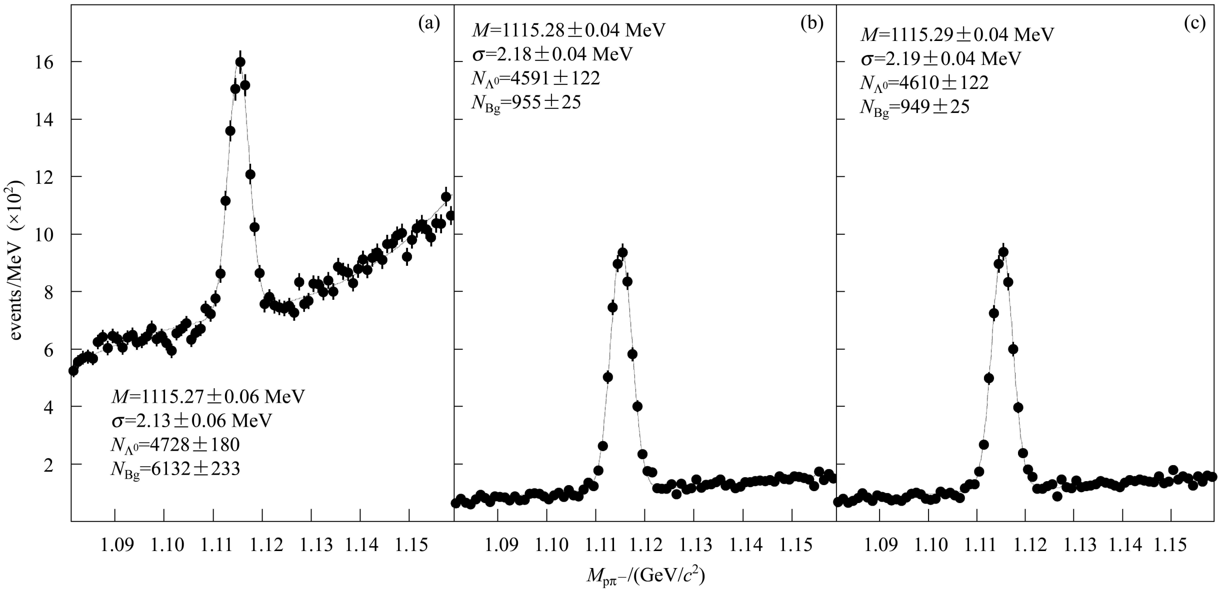

Fig. 5. Invariant spectra of $M(\mathrm{p}\pi^-)$ for events falling within the Λ^0 resonance. The fitting function is a Gaussian (peak) plus a first-order polynomial (background). The positive hadrons in panel (a) have NOT been identified; all events are included. Panel (b) contains only protons identified by the traditional TOF system. Panel (c) contains only protons identified by the ANN (any positive hadron with an output greater than 0.225 is judged a proton). The momenta of the positive particles are less than 2.9 GeV/c, the upper limit of the traditional TOF calibration method. Fit parameters are given for each plot. The number of Λ^0 events is calculated as $\sqrt{2\pi}A\sigma$, where A is the amplitude of the Gaussian fit. The number of background events is found by integrating the polynomial within $\pm 2\sigma$ of the Gaussian mean.

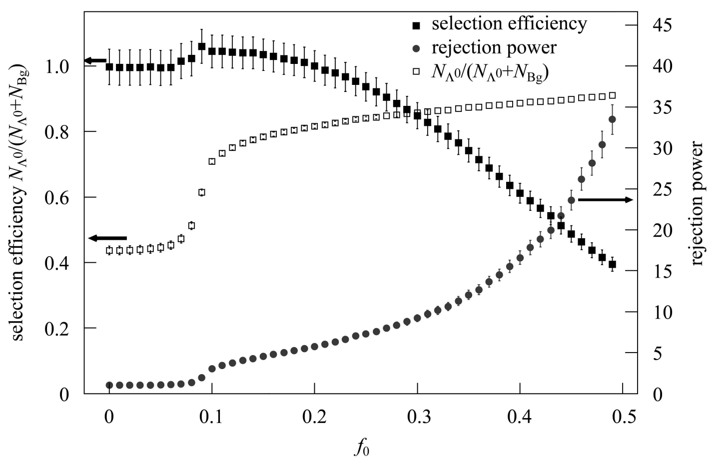

Fig. 6. Selection and rejection ratios as a function of the ANN proton identification threshold f_0. All three measures are defined in the text. The error bars are calculated from fitted parameters, and only include statistical contributions.

2. The ratio of background counts within $\pm 2\sigma$ of the peaks without and with ANN identification ("Rejection Power", calculated with $\dfrac{N_{\mathrm{Bg}}^{\mathrm{nonANN}}}{N_{\mathrm{Bg}}^{\mathrm{ANN}}}$, where $N_{\mathrm{Bg}}^{\mathrm{nonANN}}$ for the background count without ANN and $N_{\mathrm{Bg}}^{\mathrm{ANN}}$ with ANN identification), and

3. The fraction $\dfrac{N_{\Lambda^0}}{N_{\Lambda^0}+N_{\mathrm{Bg}}}$.

The results are shown in Fig. 6.

Around $f_0 = 0.1$, the "Rejection Power" and $\dfrac{N_{\Lambda^0}}{N_{\Lambda^0}+N_{\mathrm{Bg}}}$ curves increase abruptly. This corresponds to the edge of the non-proton histogram observed in Fig. 3. The slow increase in $\dfrac{N_{\Lambda^0}}{N_{\Lambda^0}+N_{\mathrm{Bg}}}$ after 0.12 reflects a reduced contamination from π^+ and K^+ events (which should only account for about 1% of the total, according to reference[4]). The "Selection Efficiency" decreases steadily, since more and more real Λ^0 decays are rejected by the network. We suggest setting the threshold to 0.225, which results in a "Selection Efficiency" of 0.955 ± 0.046 and a "Rejection Power" of 6.8 ± 0.9.

4 Conclusion

We have presented a new TOF calibration process based entirely on experimental data, using the ANN code provided by the ROOT platform. All of our work can easily be cross-checked and repeated.

This simple and convenient procedure to identify protons should be an attractive choice in experiments where a Λ^0 peak is evident. The method is not limited to proton identification—it is equally applicable to kaons and other particles provided an appropriate training sample is available.

We gratefully acknowledge the DESY management and staff for their support, as well as the personnel of collaborating institutions. Finally, we greatly appreciate the valuable suggestions of our HERMES colleagues.

References

1 HERMES Technical Design Report, DESY-PRC 93/06. 1993
2 Rith K. Progr. Part. Nucl. Phys., 2002, **49**: 245
3 Akopov N et al. Nucl. Instrum. Methods A, 2002, **479**: 511
4 Airapetian A et al. Nucl. Instrum. Methods A, 2005, **540**: 305—310
5 Ackerstaff K et al. Nucl. Instrum. Methods A, 1998, **417**: 230
6 Brun R et al. ROOT — An Object Oriented Data Analysis Framework. 2007. See Also http://root.cern.ch