

## Particle Identification at BESIII

---

**Kanglin He<sup>\*a†</sup>, Jifeng Hu<sup>ab</sup>, Bin Huang<sup>a</sup>, Gang Qin<sup>a</sup>, Shengsen Sun<sup>a</sup>, Jianyong Zhang<sup>a</sup>, Yangheng Zheng<sup>c</sup>**

<sup>a</sup> *Institute of High Energy Physics, Chinese Academy of Sciences(CAS), Beijing 100049, China*

<sup>b</sup> *University of Science and Technology of China, Hefei 230026, China*

<sup>c</sup> *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*

*E-mail: hekl@ihep.ac.cn*

A correlation analysis and neural network (NN) method have been applied in the particle identification at BESIII. The measured time-of-flight's and their covariance matrices are subjected to a correlation analysis algorithm to combine the independent measurements with correlated errors. For electron and muon identifications, a multilayered perceptrons' neural network technique has been applied. The networks are trained in each sub-detector level. The NN output of sub-detectors can be sent to a sequential network or be constructed as PDFs for a likelihood.

*XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research  
April 23-27 2007  
Amsterdam, the Netherlands*

---

\*Speaker.

†This work is supported in part by the CAS Knowledge Innovation Project(U-602, U-34), National Natural Science Foundation of China (10491300, 10491303, 10605030) and 100 Talents Program of CAS (U-25 and U-54).

## 1. Introduction

Particle identification (PID) will play an essential role in most of the BESIII physics program[1]. Good  $\mu/\pi$  separation is required in precise  $f_D/f_{D_s}$  measurements. An excellent electron ID will help to improve the measurement precision of CKM elements  $V_{cs}$  and  $V_{cd}$ . The identification of hadronic ( $\pi/K/p$ ) particles is also the most common and crucial tool in the physics analysis.

Each part of the BESIII detector provides a vast amount of information, which determines the final efficiency of particle identification. The PID ability is quite different for each sub-detector in different momentum range. To improve the PID performance, a sophisticated technique is required to combine these information together, especially when some of these information are highly correlated. In recent years, a couple of PID algorithms have been developed: the likelihood method[2], the Fisher discriminator[3], the H-Matrix estimator[4], the Artificial Neural Network[5], and the Boosted Decision Tree[6], and so on.

## 2. The PID system of BESIII

The BESIII detector[1, 7] consists of a beryllium beam pipe, a helium-based small-celled drift chamber, a Time-Of-Flight (TOF), a CsI(Tl) crystal calorimeter, a super-conducting solenoidal magnet with the field of 1 Tesla, and a muon identifier of Resistive Plate Counters (RPC) interleaved with the magnet yoke plates. The preliminary version of the BES Offline Software System (BOSS)[8] has been implemented successfully. The detector simulation[9] is based on Geant4[10]. A tremendous amount of software work has been accomplished but much more remain to be done.

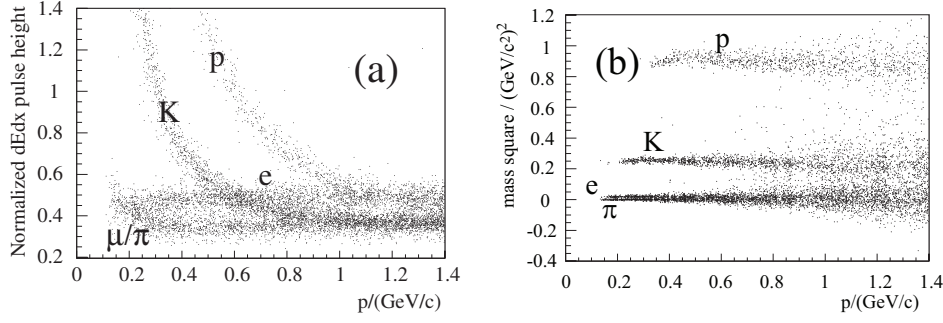
### 2.1 The dE/dx measurements

The Main Drift Chamber (MDC) consists of 43 layers of sensitive wires and works with a 60%/40% He/C<sub>3</sub>H<sub>8</sub> gas mixture. The expected momentum resolution  $\sigma_p/p$  is about 0.5% at 1 GeV/c. The energy loss in the drift chamber can provide additional information on particle identification. The normalized pulse height, which is proportional to the energy loss of incident particles in the drift chamber, is a function of  $\beta\gamma = p/m$ , where  $p$  and  $m$  are the momentum and the mass of a charged particle. Figure 1(a) shows the normalized pulse heights varying with the momentum for different particle species. The electron, muon and pion can not be well separated around 0.2 GeV/c by using dE/dx pulse heights. Similarly, the dE/dx pulse height will fail to separate electrons from kaons at the region of 0.5–0.6 GeV/c.

During the dE/dx measurements, a lot of factors have been considered [11], which include: the number of hits, the average pass lengths in each cell, the space charge and saturation effects, and the non-uniformity of electric fields. After the calibration, the resolution of dE/dx measurements is expected to be 6–7%. Using dE/dx information, a  $3\sigma$   $K/\pi$  separation can be achieved up to 0.6 GeV/c; A good  $e/\pi$  separation can be obtained above 0.4 GeV/c.

### 2.2 The TOF counter

Outside the MDC there is the TOF system, which consists of a two-layer barrel array and one layer endcap array. Two (one) readout PMTs attach on each barrel (endcap) scintillator. The



**Figure 1:** (a) The normalized pulse heights ( $dE/dx$ ) vs. momentum of charged particles; (b) The mass square distribution from TOF measurements.

expected time resolution for combing two layers is ranged from 100 to 110 ps for  $K$  and  $\pi$ , giving a  $2\sigma$   $K/\pi$  separation up to 0.9 GeV/c.

The physics goal of TOF system is to measure the flight time of charged particle. The velocity ( $\beta c$ ) and mass ( $m$ ) of the charged particle can be calculated by

$$\beta = \frac{L}{c \times t_{\text{mea}}}, \quad m^2 = p^2 \times \frac{1 - \beta^2}{\beta^2}, \quad (2.1)$$

where  $t_{\text{mea}}$  is the measured time-of-flight,  $L$  and  $p$  are the corresponding flight path and momentum of the charged particle given by the MDC measurements, and  $c$  is the velocity of light in vacuum. A typical mass square distribution for the electron, pion, kaon and proton in different momentum range is drawn in Figure 1(b).

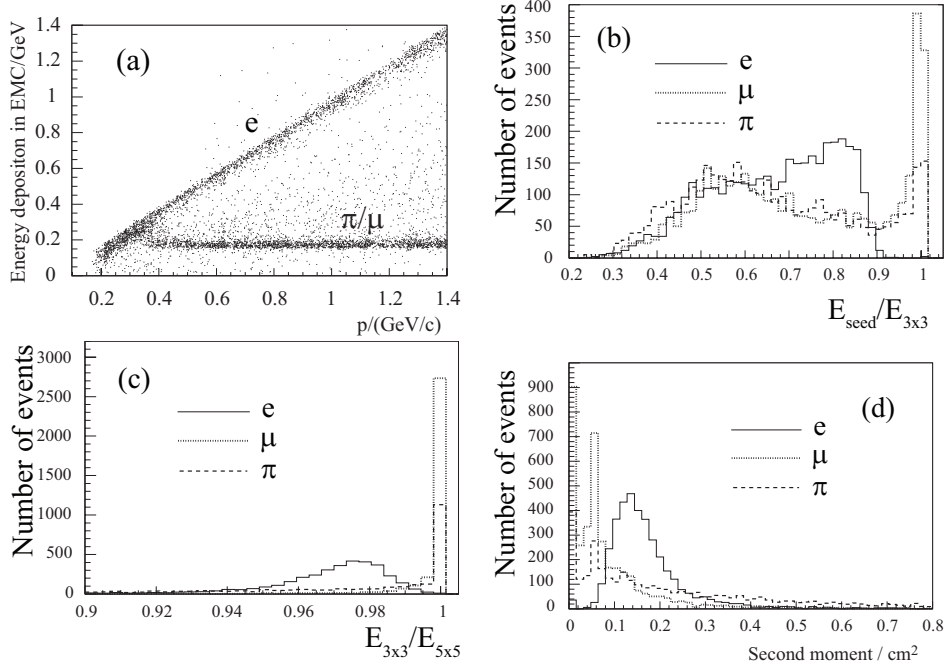
The TOF PID ability relies on the time resolution ( $\sigma_t$ ). The  $\sigma_t$  depends on the pulse height, the hit position, and the beam status. Usually the value of  $\sigma_t$  varies in different TOF counters due to the different performances of the scintillator, PMT, and electronics. Since the distinctive TOF measurements are correlated due to the common event start time, the weighted time-of-flight for two layers is obtained by a correlation analysis depicted in Ref. [12].

### 2.3 The CsI(Tl) Calorimeter

The CsI(Tl) crystal electromagnetic calorimeter (EMC) contains 6240 crystals, and is used to measure the energy of photons precisely. The expected energy and spatial resolutions are 2.5% and 0.6 cm at 1 GeV, respectively. The characteristic of electromagnetic shower is distinctive for electron, muon and hadron. The energy deposit and the shape of shower in calorimeter can be used as discrimination variables to do PID.

In the CsI(Tl) crystal, the energy deposit by ionization is about 0.165 GeV for the charged particles passing at normal incidence through the EMC. Electron and positron lose all of their energies in the calorimeter by producing the electromagnetic showers. The ratio of deposit energy to the track momentum ( $E/p$ ) will be approximately unity. Sometimes the energy deposit of hadrons will have an  $E/p$  ratio higher than that of the expected by ionization due to the nuclear interaction with materials. Figure 2(a) shows the energy deposit vs momentum for  $e$ ,  $\mu$  and  $\pi$  in EMC.

The “shape” of the shower can be characterized by the three energies:  $E_{\text{seed}}$ , the energy deposited in the central crystal;  $E_{3 \times 3}$ , the energy deposited in the central  $3 \times 3$  crystal array; and  $E_{5 \times 5}$ ,



**Figure 2:** (a) Energy deposit in EMC vs. the momentum for  $e$ ,  $\mu$  and  $\pi$ ; (b) Ratio of  $E_{\text{seed}}/E_{3\times 3}$  for  $e$ ,  $\mu$  and  $\pi$ ; (c) Ratio of  $E_{3\times 3}/E_{5\times 5}$  for  $e$ ,  $\mu$  and  $\pi$ ; (d) Second-moment  $S$  distribution for  $e$ ,  $\mu$  and  $\pi$ .

the energy deposited in the central  $5 \times 5$  crystal array. The ratios of  $E_{\text{seed}}/E_{3\times 3}$  and  $E_{3\times 3}/E_{5\times 5}$  for  $e$ ,  $\mu$  and  $\pi$  at 1 GeV/c are plotted in Figures 2(b) and 2(c), respectively.

The second-moment  $S$  is defined as

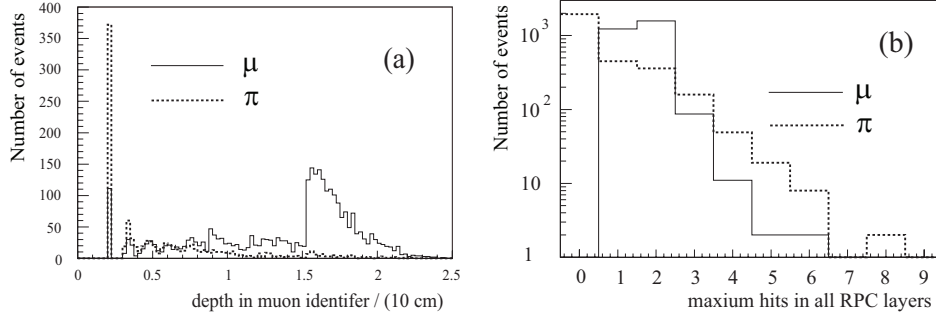
$$S = \frac{\sum_i E_i \cdot d_i^2}{\sum_i E_i}, \quad (2.2)$$

where  $E_i$  is the energy deposit in the  $i$ -th crystal, and  $d_i$  is the distance between the  $i$ -th crystal and the center position of reconstructed shower. The original idea of  $S$  was developed by the Crystal Ball experiment[13] to distinguish the cluster generated by  $\pi^0$  and  $\gamma$ . The  $S$  distributions for  $e$ ,  $\mu$  and  $\pi$  at 1 GeV/c are shown in Figure 2(d).

## 2.4 The muon system

The magnet return iron has nine layers of Resistive Plate Chambers (RPC) in the barrel and eight layers in the endcap to form a muon counter. All the RPCs have already been manufactured, tested and installed. The spatial resolution is obtained as about 16.6 mm.

The electron whose energy is exhausted in the calorimeter, cannot reach the muon counter. Most of the hadrons that passed the material of calorimeter and magnet, would be absorbed in the return irons. Muons have a quite strong punching through ability. Usually they will produce one hit in each layer, and hadrons may produce many hits in a certain layer if an interaction occurs. The distance of muon hit to the extrapolated position of inner track will be helpful to reduce the hadron contamination to a lower level, since the hits generated by the secondary muon from the decay of



**Figure 3:** (a) The travel depth of  $\mu$  and  $\pi$  in muon counter; (b) The maximum number of hits for  $\mu$  and  $\pi$  in all RPC layers.

$\pi/K$  cannot match the inner track very well. Figure 3 shows the distributions of travel depth and the maximum number of hits in all RPC layers for  $\mu$  and  $\pi$  with the momentum distributing in 0.8–1.5 GeV/c.

### 3. The correlation analysis in TOF PID

When one charged particle passes through the barrel array of scintillators, it will produce signals in one or two layers of TOF counter, corresponding to two or four measurements for time-of-flight. However, at BESIII the problem of averaging more than one TOF measurement becomes complicated since the distinctive measurements are correlated due to the common event start time. A better choice would be the weighted average of the different measurements.

#### 3.1 General algorithm

Suppose we have  $n$  measurements  $t_i$  of a particular time-of-flight. Since the measurements are correlated, we need more information than just the individual errors. Accordingly, let's define the covariance matrix  $V_t$ , whose terms are given by  $(V_t)_{ij} = \langle \delta t_i \delta t_j \rangle$ , where  $\delta t_i = t_i - \bar{t}$ ,  $\bar{t}$  is the average of  $t_i$ . The best linear estimator for the TOF which accounts for all measurements, including errors and correlations can be constructed generally as

$$\bar{t} = \sum_i w_i t_i, \quad \sum_i w_i = 1. \quad (3.1)$$

where the weights  $w_i$  must be found. Writing  $\delta \bar{t} = \sum_i w_i \delta t_i$  and using the definition of the standard deviation, we get

$$\sigma_{\bar{t}}^2 = \sum_{ij} w_i w_j (V_t)_{ij}. \quad (3.2)$$

To minimize  $\sigma_{\bar{t}}^2$  with the condition  $\sum_i w_i = 1$ , we use the Lagrange multiplier technique. Let's write

$$\sigma_{\bar{t}}^2 = \sum_{ij} w_i w_j (V_t)_{ij} + \lambda (\sum_i w_i - 1), \quad (3.3)$$

and set the derivative of Eq. (3.3) with respect to the  $w_i$  and Lagrange multiplier  $\lambda$  to zero. This give the solution

$$w_i = \frac{\sum_k (V_t^{-1})_{ik}}{\sum_{jk} (V_t^{-1})_{jk}}. \quad (3.4)$$

### 3.2 Errors and correlations of TOF measurements

The time resolution of TOF ( $\sigma_t$ ) can be factorized by the production of  $\sigma_t(Q) \cdot \sigma_t(z)$ , in which  $\sigma_t(Q)$  and  $\sigma_t(z)$  are functions of the pulse height  $Q$  and the hit position  $z$ [14]. The variation of  $\sigma_t(Q)$  is complicated, and needs the detailed study on the real data. Here, only the  $z$  dependent time resolution is taken into account, since the  $\sigma_t(z)$ 's are in similar manners for electrons, muons and hadrons[14]. Figure 4(a) shows a typical variation of  $\sigma_t(z)$  in one readout unit as a function of  $z$  from the Bhabha event. The time resolution becomes worse while the hit position is far from the readout end.

The  $t_{\text{mea}}$  is obtained by calculating the difference between the end-time and the start-time. The accuracy of end-time is limited by the detector resolution and the precision of start-time, which is controlled by the uncertainties of  $t_0$ . Thus, for a given barrel TOF counter, the  $t_{\text{mea}}$  in the left-end and the right-end readout units can be decomposed as

$$t_1 = t_c + (t_D)_1, \quad t_2 = t_c + (t_D)_2, \quad (3.5)$$

where  $t_1$  and  $t_2$  represent the  $t_{\text{mea}}$ 's in two readout PMTs,  $t_c$  represents the common part of time between  $t_1$  and  $t_2$ ,  $(t_D)_1$  and  $(t_D)_2$  represent the uncorrelated part of  $t_1$  and  $t_2$ . The covariance matrix for  $t_1$  and  $t_2$  can be expressed as

$$V_t = \begin{pmatrix} \sigma_1^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_2^2 \end{pmatrix}, \quad (3.6)$$

where  $\sigma_1$  and  $\sigma_2$  are the time resolution in the left-end and the right-end readout units,  $\sigma_c$  the fluctuation of  $t_c$ . According to the definition of covariance matrix, we have the following expressions

$$\begin{aligned} \sigma_1^2 &= \langle \delta t_1 \delta t_1 \rangle = \langle \delta t_c \delta t_c \rangle + \langle \delta (t_D)_1 \delta (t_D)_1 \rangle, \\ \sigma_2^2 &= \langle \delta t_2 \delta t_2 \rangle = \langle \delta t_c \delta t_c \rangle + \langle \delta (t_D)_2 \delta (t_D)_2 \rangle, \\ \sigma_c^2 &= \langle \delta t_1 \delta t_2 \rangle = \langle \delta t_c \delta t_c \rangle, \end{aligned} \quad (3.7)$$

by the fact that the correlations  $\langle \delta t_c \delta (t_D)_1 \rangle = 0$ ,  $\langle \delta t_c \delta (t_D)_2 \rangle = 0$  and  $\langle \delta (t_D)_1 \delta (t_D)_2 \rangle \approx 0$ .

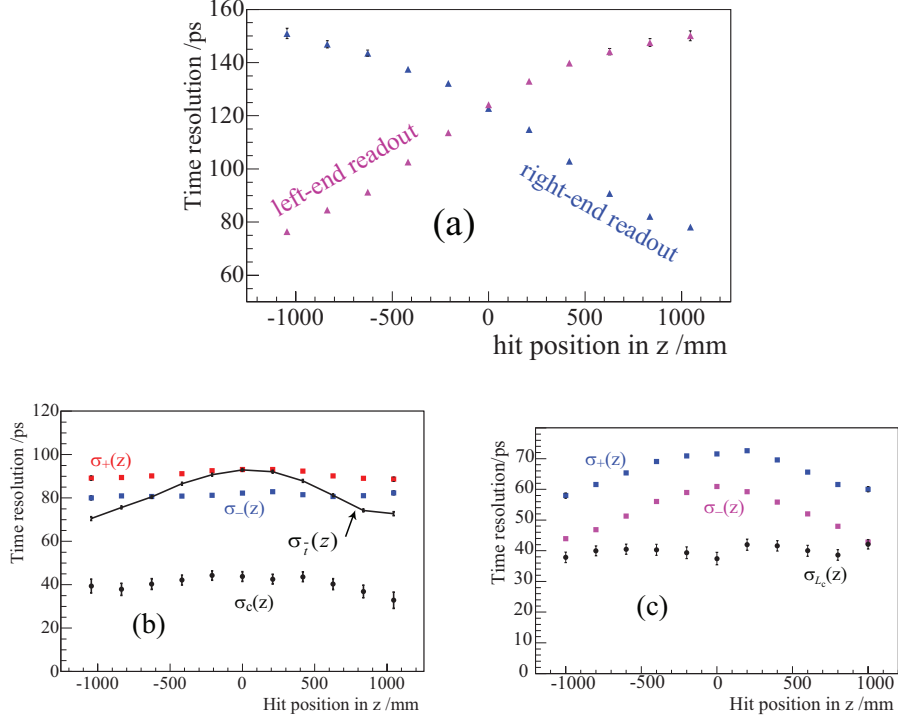
To get the  $\sigma_c$  conveniently, let's define two new time variables

$$t_+ = \frac{t_1 + t_2}{2}, \quad t_- = \frac{t_1 - t_2}{2}. \quad (3.8)$$

The fluctuations of  $t_+$  and  $t_-$  can be expressed as

$$\begin{aligned} \sigma_+^2 &= \langle \delta t_+ \delta t_+ \rangle = \frac{\sigma_1^2 + \sigma_2^2}{4} + \frac{\sigma_c^2}{2}, \\ \sigma_-^2 &= \langle \delta t_- \delta t_- \rangle = \frac{\sigma_1^2 + \sigma_2^2}{4} - \frac{\sigma_c^2}{2}, \end{aligned} \quad (3.9)$$

where  $\sigma_+$  and  $\sigma_-$  are the time resolution of  $t_+$  and  $t_-$ . The  $\sigma_c$  can be directly extracted by the calculation of  $\sigma_c = \sqrt{\sigma_+^2 - \sigma_-^2}$ . Figure 4(b) shows the distribution of  $\sigma_+(z)$ ,  $\sigma_-(z)$  and  $\sigma_c(z)$ ,



**Figure 4:** (a) The variation of  $\sigma_t(z)$  for the left-end and the right-end readout unit in barrel TOF counter; (b) Time resolution of  $t_+$ ,  $t_-$ ,  $t_c$  and the weighted time  $\bar{t}$  for one-layer of TOF measurement; (c) The correlations between two layer TOF measurements, where  $\sigma_{L_c}(z) = \sqrt{\sigma_+^2(z) - \sigma_-^2(z)}$ .

where  $\sigma_c(z)$  is approximately a constant, around 40 ps. Substituting the special expression of Eq. (3.6) into Eqs. (3.1)–(3.4), we get

$$w_1 = \frac{\sigma_2^2 - \sigma_c^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_c^2}, \quad w_2 = \frac{\sigma_1^2 - \sigma_c^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_c^2}, \quad (3.10)$$

and

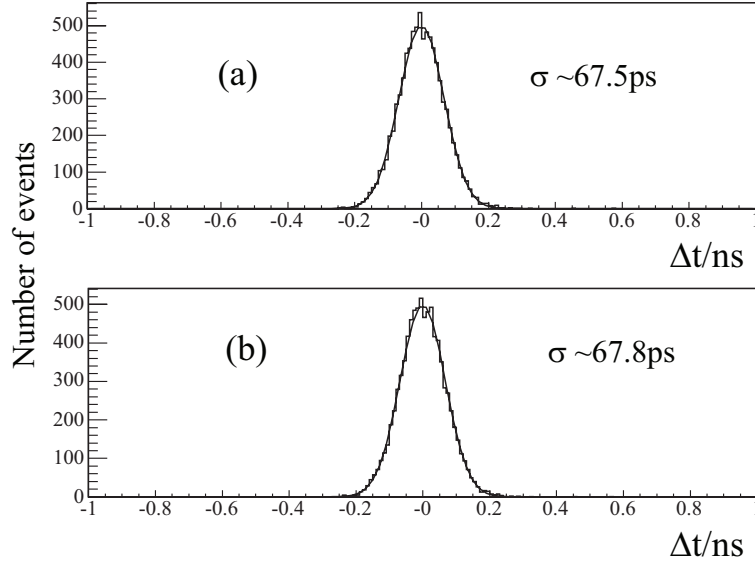
$$\sigma_{\bar{t}}^2 = \frac{\sigma_1^2 \cdot \sigma_2^2 - \sigma_c^4}{\sigma_1^2 + \sigma_2^2 - 2\sigma_c^2}. \quad (3.11)$$

The average time resolution for one-layer is about 86 ps.

### 3.3 Combining the time-of-flight from two-layer measurements

With the Bhabha events, we found that the correlation term  $\sigma_c$ 's in two layer measurements are the same value ( $\sim 40$  ps). The weighted time-of-flight and its error in each layer,  $t_{L_i}$  and  $\sigma_{L_i}$  ( $i = 1, 2$ ), can be determined from Eqs. (3.1), (3.10) and (3.11). Similar to the method adopted for one-layer measurement, we can construct the covariance matrix for two-layer measurements as follows

$$V_t = \begin{pmatrix} \sigma_{L_1}^2 & \sigma_{L_c}^2 \\ \sigma_{L_c}^2 & \sigma_{L_2}^2 \end{pmatrix}, \quad (3.12)$$



**Figure 5:**  $\Delta t = t_{\text{mea}} - t_{\text{exp}}$  distribution: (a)  $\bar{t}$  is weighted by  $\bar{t}_{L_i} (i = 1, 2)$ ,  $\bar{t}_{L_i}$ 's are the average time in each layer which is weighted by  $t_{E_i} (i = 1, 2)$ ,  $t_{E_i}$ 's are the TOF measurements in each end of readout units; (b)  $\bar{t}$  is directly weighted by  $t_{E_i} (i = 1, 2, \dots, 4)$ ,  $t_{E_i}$ 's are the TOF measurements in the four-end of readout units.

where  $\sigma_{L_c}$  is the correlation between two-layer measurements.

Substituting  $t_1, t_2$  with  $t_{L1}, t_{L2}$  in Eqs. (3.5) and (3.8), we can get the corresponding errors and correlations. The calculations of the  $\sigma_+, \sigma_-$  and  $\sigma_{L_c}$  are illustrated in Figure 4(c). The correlation between two-layer measurements is almost a constant,  $\sim 40$ ps. It agrees fairly well with the results from the two-end measurements in each layer. Thus, we have  $\sigma_{L_c} = \sigma_c$ .

The weighted time-of-flight of two-layer measurements can be easily obtained by applying the covariance matrix of Eq. (3.12) in Eq. (3.10). The resulting  $t_{\text{mea}} - t_{\text{exp}}$  are drawn in Figure 5(a). The average time resolution from two-layer measurements is about 68 ps, a little bit worse than  $86/\sqrt{2} \approx 61$  ps, where the value of 86 ps is the average time resolution of one-layer measurement.

The apparatus of barrel TOF array can be treated in such a way that the four independent readout PMTs can measure the time-of-flight for a charged particle. The covariance matrix of TOF measurements can be constructed as

$$V_t = \begin{pmatrix} \sigma_1^2 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_2^2 & \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 & \sigma_3^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 & \sigma_4^2 \end{pmatrix}. \quad (3.13)$$

In Eq. (3.13),  $\sigma_i (i = 1, 2, \dots, 4)$  are the resolution of all readout units, the correlations ( $\sigma_c$ ) between different measurements are taken as the same value ( $\sim 40$  ps) by the fact that  $c_{EE} = c_{LL}$ , where  $c_{EE}$ 's are the correlation between the two-end of readout units in each layer,  $c_{LL}$  is the correlation between two-layer measurements. Employing the covariance matrix Eq. (3.13) in Eqs. (3.1)–(3.4), the weight factors  $w_i (i = 1, 2, \dots, 4)$  can be easily calculated. The resulting  $t_{\text{mea}} - t_{\text{exp}}$  distribution is drawn in Figure 5(b).



As shown in Figure 5(a) and 5(b), the resulting time resolutions from two weighted methods are consistent. The standard weighted method adopted in TOF calibration/reconstruction software system of BESIII will be in two steps: combining the two-end TOF measurements in each layer; calculating the weighted time from the two-layer measurements.

#### 4. Applying the ANN technique in PID algorithm at BESIII

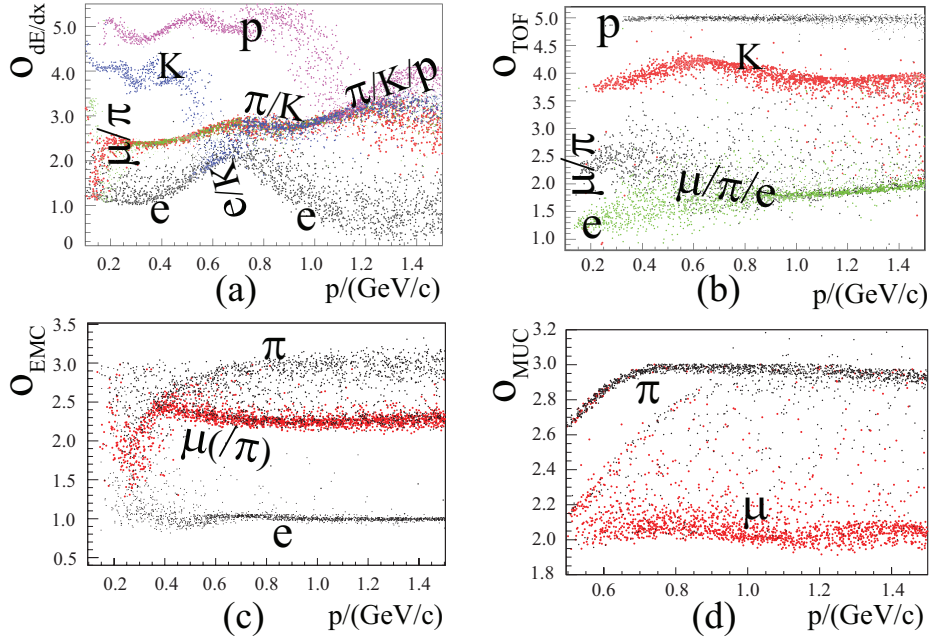
At present, a class of Multilayer Perceptrons (MLP)[15] network which is implemented in ROOT, has been applied to BESIII PID algorithm. The PID variables described in Section 2.3 and Section 2.4 are correlated each other. Without losing information, a cell analysis may not be sufficient for the likelihood method to get an optimal result. As we know, the correlations of PID variables among sub-detectors are reasonably small and could be ignored. For example, the  $dE/dx$  measurement and the energy deposit in EMC have almost no influence on the time of flight measurements. This allows us to configure the networks sequentially.

##### 4.1 The configuration of PID networks

The PID variables, selected from each sub-detector together with the incident momentum and the transverse momentum, have been grouped and trained separately. Each sub-detector (the barrel part and the endcap part) has one output. In this step, the network for each sub-detector is quite simple. Almost all sub-networks are configured with one hidden layer containing  $2N$  hidden neurons, where  $N$  is the number of the input neurons. 50,000 single track events for each particle species with momentum ranging from 0.1 GeV/c to 1.6 GeV/c and  $-0.83 < \cos\theta < 0.83$  are generated and subjected to networks, where  $\theta$  is the incident polar angle. The output values are constrained to be 1, 2, 3, 4 and 5 for electron, muon, pion, kaon and proton, respectively. The training results for each sub-detector are shown in Figures 6(a)–6(d).

The bands of muon and pion are merged into one single peak (around 2.5) in the outputs of  $dE/dx$ , TOF and EMC. The information in EMC and MUC is hard to be applied to kaon and proton samples. The output of EMC can help to separate muon and pion slightly. But the output of MUC could separate muon from hadrons clearly.

The outputs of neural network from sub-detectors can be combined in several ways to get the near-optimal discrimination variables. For example, the probability density functions (PDF) for the resulting variables can be used as the basis for a likelihood analysis, or can be used as the input variables for a sequential network. At present, a conventional likelihood analysis based on the neural output variables and a sequential network analysis are applied in parallel to the BESIII PID algorithm. Here, we are only concerned with the study of the sequential networks. Basically, the network consists of two input momentum variables and four input PID variables. The momentum variables are the incident momentum and the transverse momentum. The PID variables include the neural outputs from  $dE/dx$  ( $O_{dE/dx}$ ), TOF ( $O_{TOF}$ ), EMC ( $O_{EMC}$ ) and MUC ( $O_{MUC}$ ) system (the barrel part and the endcap part separately). The network is configured with one hidden layer of ten hidden neurons. Electron, muon, and hadron separations are studied with several simulated Monte Carlo samples through different configurations of networks. Cuts are put on the output of final discrimination variables (the output of sequential network  $O_{seq}$ ).



**Figure 6:** NN outputs of sub-detectors. (a) dE/dx output; (b) TOF output; (c) EMC output; (d) MUC output.

#### 4.2 Muon identification

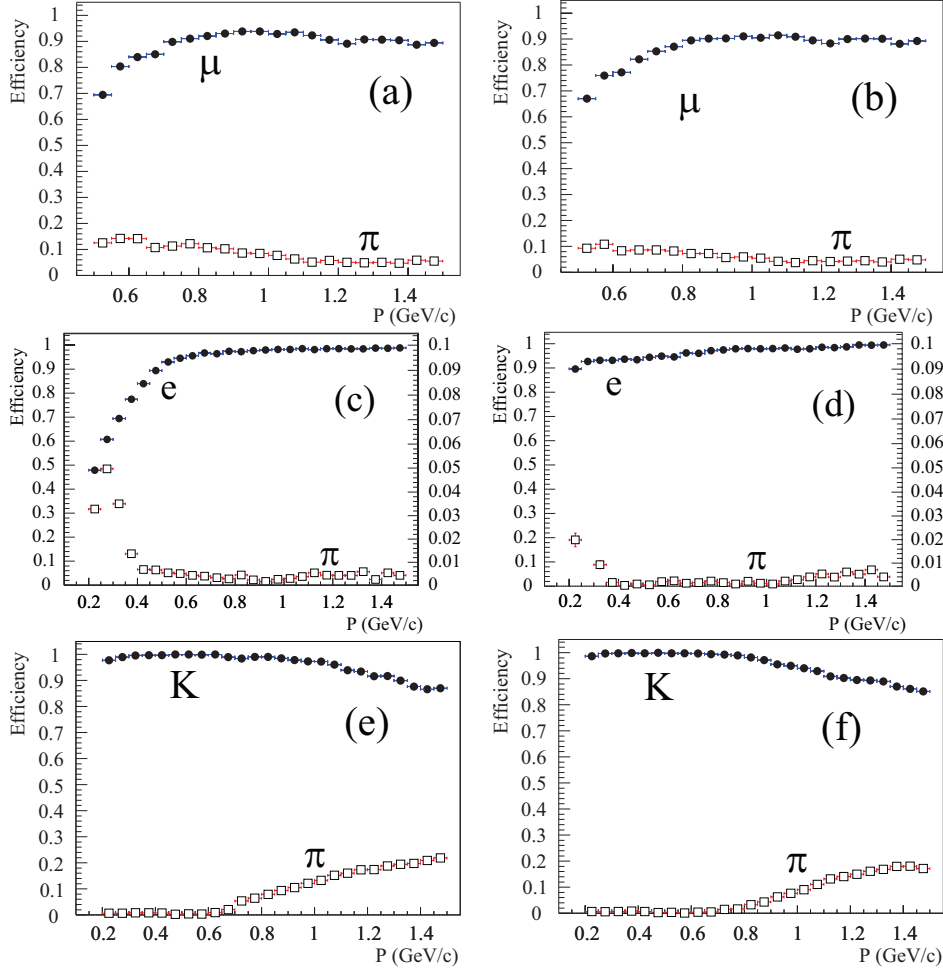
The muon candidate is required to have response in  $\mu$ -counter. The sequential network is trained with two PID variables: the  $O_{MUC}$  and the  $O_{EMC}$ . The  $\mu$ -ID abilities are studied in different momentum partitions by comparing the discrimination results from  $O_{MUC}$  and  $O_{seq}$ . Figures 7(a) and 7(b) show the variations of the muon identification efficiency and pion contamination rate as functions of incident track momentum, where the track momentum is required to be greater than the cut-off threshold ( $\sim 500$  MeV/c). Above 0.8 GeV/c, the muon identification efficiency is around 90%, and the pion contamination rate is about 5%. The additional information from EMC may help to improve the  $\mu$ -ID ability.

As experienced in BaBar experiment[16], the additional variables, e.g., the goodness of muon track fit and the goodness of the muon track matching to the extrapolation position from inner track system, may help to reduce the background contamination rates. These information will be studied at BESIII in the future.

#### 4.3 Electron identification

Figure 7(c) shows the variations of electron identification efficiency and pion misidentification rate at different momentum range by setting cuts on  $O_{EMC}$ . Above 0.6 GeV/c, one can see that the electron-ID efficiency is greater than 95% while the pion contamination rate can go down to the  $10^{-3}$  level. The  $e/\pi$  separation is quite bad for low momentum tracks (less than 0.4 GeV/c).

Both  $O_{dE/dx}$  and  $O_{EMC}$  are good discrimination variables to separate electron from pion above 0.4 GeV/c. The  $O_{TOF}$  can separate  $e/\pi$  very effectively below 0.3 GeV/c. The network is trained with  $O_{dE/dx}$ ,  $O_{TOF}$  and  $O_{EMC}$ . As shown in Figure 7(d), the network offers a nearly uniform accep-



**Figure 7:** PID efficiency and contamination rate in different momentum partitions. (a)  $\mu/\pi$  separation with  $O_{MUC}$ ; (b)  $\mu/\pi$  separation with  $O_{MUC}$  and  $O_{EMC}$ ; (c)  $e/\pi$  separation with  $O_{EMC}$ ; (d)  $e/\pi$  separation with  $O_{EMC}$ ,  $O_{TOF}$  and  $O_{dE/dx}$ ; (e)  $K/\pi$  separation with likelihood method; (f)  $K/\pi$  separation with neural networks. In (c) and (d), pion contamination rates are enlarged by a factor of 10.

tance and background contamination between 0.25 GeV/c and 1.6 GeV/c. It is an interesting result that the acceptance hole between 0.2 GeV/c and 0.4 GeV/c has almost vanished by adding an appropriate cut, where no detector has clear discrimination power for electron. The system obviously makes the inference that the particle has to be an electron if it is not one of the others. This is the combined contribution from the sub-detectors.

#### 4.4 $\pi/K$ separation

As a general view, the proton identification is extremely good with TOF and  $dE/dx$  information at BESIII. Here and below, only the  $K/\pi$  separation is focused on. As discussed in Section 2, the  $dE/dx$  can identify  $K/\pi$  very effectively below 0.6 GeV/c; the two-layer TOF can separate  $K/\pi$  up to 0.9 GeV/c.

Traditionally, a likelihood method which combines the TOF and  $dE/dx$  information will be applied to the hadron identification. To construct the PDFs, data are divided into several bins in momentum and  $\cos\theta$  partitions to obtain the corresponding resolutions and offsets. Figure 7(e) shows the variations of kaon identification efficiency and pion contamination rate as functions of momentum. In the real world, there are often tails on distributions due to track confusion, nonlinearities in detector response, and many other experimental sources which are imperfectly described in PDFs. Herein, it is helpful to apply NN technique in hadron separation.

For hadron separation, the network is trained with two PID variables: the  $O_{\text{TOF}}$  and the  $O_{dE/dx}$ . The PID ability is studied with the cuts on the output of sequential network  $O_{\text{seq}}$ . The results are shown in Figure 7(f). Both the likelihood method and the neural network method give the consistent results. Below 1 GeV/c, one can see that the kaon-ID efficiency is greater than 95% while the pion contamination rate is less than 10%. The  $K/\pi$  separation is extremely good for low momentum tracks (less than 0.8 GeV/c).

## 5. The future PID algorithms at BESIII

The shapes of PID variables in EMC and MUC systems are complicated and there may exist non-linear correlations between PID variables. It is difficult for the likelihood method to construct the PDFs analytically and handle the correlations properly. From our studies, good electron-ID and muon-ID can be easily achieved from the neural network at BESIII with the full detector information. In a simple application, e.g., for hadron separation, we get the consistent results from the neural network and the likelihood method, where the PID variables in TOF and  $dE/dx$  systems are quasi-Gaussian, and the correlation between two-layer TOF measurements is approximately linear. The flexible configuration of PID networks can be employed in both the simple and the complicated applications.

There are still a lot of factors need to be taken into account while applying the artificial neural network technique in the particle identification at BESIII. For example, one or several input variables have to be removed due to the imperfect consistency between data and Monte Carlo simulations. The impurity of the training sample may introduce the additional systematical uncertainties. More detailed studies are needed in the future. Currently, both the likelihood and the artificial neural network PID algorithms are studied in parallel at BESIII. The final PID algorithm will definitely be a combination of all the methods, which have been created and applied in high energy physics experiment, one way might be by using the likelihood method to combine the neural network outputs from sub-detectors. Of course, to make the final decision of the BESIII PID algorithm, there is still a long way to go.

## References

- [1] BESIII Design Report, Interior Document in Institute of High Energy Physics, 2004.
- [2] T. Carli and B. Koblitz, *Nucl. Instrum. Methods* **A501** (576–588) 2003; L. Holmström, R. Sain and H. E. Miettinen, *Comput. Phys. Commun.* **88** (195) 1995.
- [3] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, *Annals of Eugenics* **7** (179-188) 1936.

- [4] P. C. Bhat (DØ Collaboration), arXiv:hep-ex/9507007.
- [5] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon, Oxford 1998; R. Beale and T. Jackson, *Neural Computing: An Introduction*, Adam Hilger, New York 1991.
- [6] B. P. Roe, H. J. YANG et al. *Nucl. Instrum. Methods* **A543** (577–584) 2005.
- [7] F. A. Harris (BES Collab.), arXiv:physics/0606059.
- [8] W. D. Li, H. M. Liu et al. , *The Offline Software for the BESIII Experiment*, in proceeding of *CHEP06*, Mumbai 2006.
- [9] Z. Y. Deng et al. , *HEP & NP* **30(5)** (371–377) 2006 (in Chinese).
- [10] S. Agostinelli et al. (Geant4 Collab.), *Nucl. Instrum. Methods* **A506** (250) 2003.
- [11] D. Y. Wang, K. L. He, W. G. Li et al. , *Software Developments and Studies of BESIII  $dE/dx$  Reconstruction*, the Publications for the 5th Seminar of Nuclear Instrument and its applications, China 2005 (in Chinese).
- [12] J. F. Hu, K. L. He, Z. P. Zhang et al. , will appear in *HEP & NP* **31(10)**.
- [13] C. Edwards et al. , *Phys. Rev.* **D25** (3065) 1982.
- [14] S. S. Sun, K. L. He et al. , *HEP & NP* **29(2)** (162–167) 2005 (in Chinese).
- [15] R. Brun, F. Rademakers et al. , *ROOT Users Guide 5.12*. (73–76) 2006.
- [16] I. Narsky, *StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data*, arXiv:physics/0507143.