

# BES II 物理分析中的多变量粒子鉴别方法\*

辛波<sup>1,2,1)</sup> 苑长征<sup>2</sup> 马凤才<sup>1</sup>

1 (辽宁大学物理系 沈阳 110036)

2 (中国科学院高能物理研究所 北京 100039)

**摘要** 以北京谱仪(BES II)上常用的粒子鉴别方法(联合 TOF 测得的飞行时间和 MDC 的  $dE/dx$ )为基础,尝试利用 TOF 的  $Q$  值和 BSC 的沉积能量,发展基于多变量的粒子鉴别. 将新的方法应用于  $p/\bar{p}$  的鉴别,对粒子鉴别的效率、误判率和引进的系统误差进行了仔细的研究,证明均比原来有所改善.

**关键词** 北京谱仪 粒子鉴别 效率 误判几率 系统误差

## 1 引言

北京谱仪(BES II)<sup>[1]</sup>是运行于北京正负电子对撞机(BEPC)上的一台通用型磁谱仪,工作于质心系能量 2—5 GeV 的能区,它采集的数据被广泛应用于  $\tau$  粲物理的实验研究. 在许多物理课题的研究中,如测量衰变道的分支比或截面,粒子鉴别起着举足轻重的作用. BES 可直接测量的  $e^+e^-$  对撞的末态粒子可分为带电粒子( $e, \mu, \pi, K, p$ )和中性粒子( $\gamma$  光子)两类.  $\gamma$  光子的鉴别是比较容易的,它只在簇射计数器(SC)中留下击中和能量沉积信息;而依靠  $\mu$  子计数器有无击中信息基本上就可将  $\mu$  子鉴别出来;电子在 SC 中基本上损失其全部能量,利用这一特性,可将电子与其他粒子区分开来.

关于  $\pi, K, p/\bar{p}$  的鉴别,目前普遍采用的方法是联合飞行时间计数器(TOF)测得的飞行时间和主漂移室(MDC)测得的  $dE/dx$ <sup>[2-4]</sup>. 当 TOF 和  $dE/dx$  单独用来鉴别粒子效果不好时,综合利用两种信息可以提高粒子鉴别的效率. 令  $i = \pi, K, p$  代表不同的粒子,经过修正后的  $\chi_{dE/dx}^i$  和  $\chi_{TOF}^i$  都服从标准正态分布,因而  $\chi^2 = \chi_{dE/dx}^2 + \chi_{TOF}^2$  是一个自由度为 2 的  $\chi^2$  分布. 显然  $\chi^2$  越小,径迹是第  $i$  种假定的置信度越

大. 定义径迹鉴别为第  $i$  种粒子的几率  $Prob_i = \int_{\chi^2}^{\infty} f(x, \nu) dx$ , 其中  $f(x, \nu)$  是自由度为  $\nu$  的  $\chi^2$  分布的概率密度函数. 在研究的径迹同时具有  $dE/dx$  和 TOF 信息的情况下,  $\nu = 2$ . 根据概率统计的分析,真正的第  $i$  种粒子的样本的  $Prob_i$  分布是 0 到 1 之间的均匀分布,通常将这作为检验粒子鉴别系统是否可靠的判据之一.

在实际的分析中,在一定的动量段,如果 TOF 和  $dE/dx$  中有一种信息不好,通常的做法是在此动量段将此信息丢弃不用. 例如当  $p/\bar{p}$  动量小于 0.5 GeV 时,由于  $p$  尤其是  $\bar{p}$  与探测器物质强烈的相互作用,通常的做法是将此时  $p/\bar{p}$  的 TOF 信息丢弃,而只用  $dE/dx$  的信息进行鉴别. 就是说,大部分动量小于 0.5 GeV 的质子无法击中 TOF,而如此低动量的反质子也只能是通过与探测器物质的作用产生的次级粒子在 TOF 中留下不正确的信息. 然而,这本身就是一个有用的信息,就是说,如果探测器探测到一条带正电径迹,它的动量小于 0.5 GeV 但没有到达 TOF,那么它最大的可能是质子,而不是  $\pi$  或者  $K$ . 从下文可以看到,反质子的次级粒子在 TOF 中的击中信息也是和  $\pi, K$  截然不同的. 对于桶部簇射计数器(BSC)也有同样的情况.

2003-11-10 收稿

\* 中国科学院百人计划经费(U-25),辽宁省教育厅科研基金(20021003)资助

1) E-mail: xinb@mail.ihep.ac.cn

本文利用 BES II 在 2001—2002 年积累的 1400 万  $\psi(2S)$  数据, 挑选出纯度较高的  $\pi, K, p/\bar{p}$  的样本, 对不同动量段 TOF 的  $Q$  值和 BSC 的沉积能量分布做了细致的研究. 相应的 Monte Carlo (MC) 数据则是用基于 GEANT3 的北京谱仪新的探测器模拟程序 SIMBES (V104) 产生的. 利用这些有用的信息, 对原有的粒子鉴别提出了改进的方法.

## 2 鉴别方法的改进和样本的选取

实际的改进中采用的是似然法. 对每一条带电径迹, 定义估计量

$$E^i = \frac{P^i}{\sum_j P^j}, \quad (1)$$

其中  $i = \pi, K, p/\bar{p}, P^i$  由下式给出:

$$P^i = \prod_j P^i_j(x_j), \quad (2)$$

$P^i_j$  是由变量  $x_j$  给出的这条径迹为相应假定  $i$  的概率. 根据定义, 对于纯净的  $\pi, K, p/\bar{p}$  样本, 相应的  $E^i$  应该很接近 1. 共用到了 4 个变量:

- 1) 桶部簇射计数器 (BSC) 的沉积能量 (BSC-E);
- 2) 飞行时间计数器 (TOF) 的  $Q$  值 (TOF-Q);
- 3) 飞行时间计数器 (TOF) 的飞行时间 (TOF-T);
- 4) 主漂移室 (MDC) 的电离能量损失 ( $dE/dx$ ).

其中前两个是由上文所述的参考分布得到的, 后两个则由经过修正后的 TOF 和  $dE/dx$  所服从的高斯分布读取.

粒子鉴别的改进是对真实数据 (以下简称为数据) 和 MC 数据 (以下简称为 MC) 分别做的. MC 样本的获取比较容易, 关键是要在真实的实验数据中挑选出较纯的  $\pi, K, p/\bar{p}$  的样本, 来获得相应的 TOF 的  $Q$  值和 BSC 的沉积能量的分布, 但是要避免在挑选样本时使用相应的信息. 用到的反应道见表 1.

表 1 选取样本用到的反应道

	样本来源
$\pi$	$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow \pi^+ \pi^- \pi^0$
$K$	$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow K^+ K^- \pi^0$ $\psi(2S) \rightarrow K^+ K^- \pi^+ \pi^-$
$p/\bar{p}$	$\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}\pi^0$ $J/\psi \rightarrow p\bar{p}\eta, J/\psi \rightarrow p\bar{p}$ $\psi(2S) \rightarrow \pi^+ \pi^- p\bar{p}$

从表 1 中可以看出, 用到的反应道都是四叉事例, 因为这样的事例比较容易挑选. 主要的选择条件如下: (a) 带电径迹数为 4, 径迹拟合质量为 2, -9 或 -19; (b) 极角范围  $|\cos\theta| < 0.84$ ; (c) 要求相应假定的四动量拟合的  $\chi^2$  概率  $Prob(\chi^2, 4) > 0.01$ ; (d) 对不作为样本的粒子作较严的粒子鉴别; (e) 对作为样本的粒子作的粒子鉴别不能用到与所研究的分布相关的信息.

经过以上的选择, 可以得到比较纯的  $\pi, K, p$  和

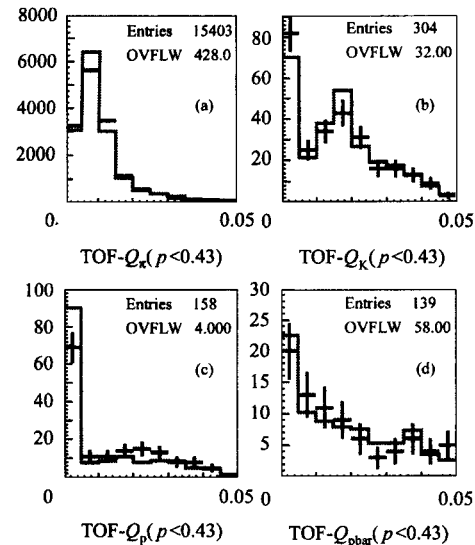


图 1 动量小于 0.43 GeV 时不同带电粒子 TOF 的  $Q$  值的分布

(a), (b), (c), (d) 分别为  $\pi, K, p$  和  $\bar{p}$ . 误差棒是数据, 直方图是 MC.

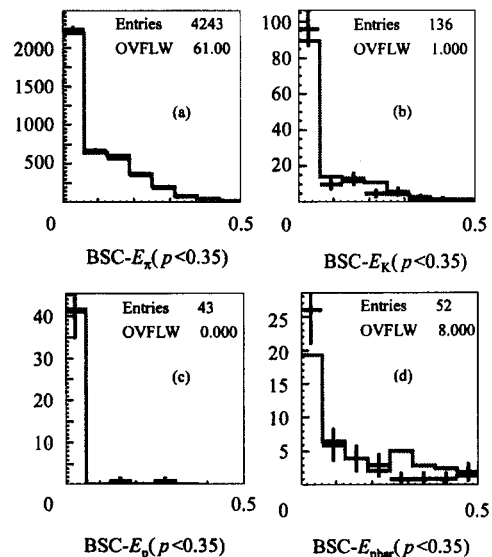


图 2 动量小于 0.35 GeV 时不同带电粒子 BSC 的沉积能量的分布

(a), (b), (c), (d) 分别为  $\pi, K, p$  和  $\bar{p}$ . 误差棒是数据, 直方图是 MC.

$\bar{p}$  的样本, 进而得到不同样本在不同动量段的 TOF 的  $Q$  值和 BSC 的沉积能量的参考分布. 图 1 是动量小于 0.43 GeV 时不同带电粒子 TOF 的  $Q$  值的分布. 图 2 是相应的 BSC 的沉积能量的分布. 从中可以看出,  $\pi, K, p$  和  $\bar{p}$  的分布的确有显著的不同, 这些信息的利用必将有助于它们更好的区分.

### 3 粒子鉴别的可靠性检验

#### 3.1 不同变量之间的关联

对于似然法, 必须保证不同变量之间是不相关的, 否则(2)式中应考虑关联. 利用从上文所述的反应道得到的 MC 样本, 对用到的 4 个变量进行各种可能的组合, 对它们之间的相关性进行了仔细的研究. 例如, 图 3 给出了 TOF 的  $Q$  值(TOF- $Q$ )对 TOF 的飞行时间(TOF- $T$ )的二维分布. 从中可以看出, 这 4 个变量基本上是不相关的. 惟一的例外是当 TOF 的  $Q$  值等于零时, 飞行时间也等于零. 这点很容易理解. 程序中已经考虑了这一点.

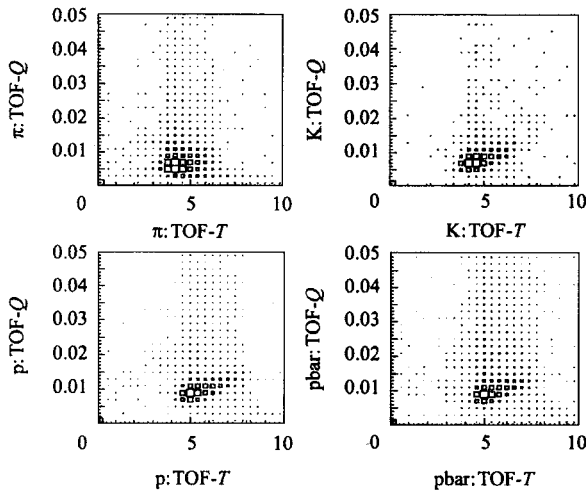


图 3 利用 MC 样本得到的 TOF 的  $Q$  值(TOF- $Q$ )对 TOF 的飞行时间(TOF- $T$ )的二维分布

表 2 利用  $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$  反应道对不同变量组合的  $p/\bar{p}$  鉴别效率的检验 ( $E^{p/\bar{p}} > 0.7$ )

数据		MC		TOF- $Q$	BSC- $E$	TOF- $T$	$dE/dx$
$p$	$\bar{p}$	$p$	$\bar{p}$				
$90.13 \pm 0.61$	$88.84 \pm 0.64$	$92.12 \pm 0.17$	$91.59 \pm 0.18$	✓	✓	✓	✓
$88.46 \pm 0.67$	$86.58 \pm 0.71$	$91.56 \pm 0.18$	$90.66 \pm 0.18$			✓	✓
$89.46 \pm 0.64$	$87.56 \pm 0.68$	$92.57 \pm 0.17$	$91.72 \pm 0.17$		✓	✓	✓
$89.54 \pm 0.63$	$88.13 \pm 0.66$	$91.64 \pm 0.18$	$90.91 \pm 0.18$	✓		✓	✓
$84.10 \pm 0.97$	$84.45 \pm 0.96$	$87.06 \pm 0.27$	$85.52 \pm 0.28$			✓	
$41.17 \pm 1.41$	$39.01 \pm 1.36$	$58.21 \pm 0.39$	$58.01 \pm 0.39$				✓

注: “✓”表示使用该信息.

#### 3.2 对带电粒子的鉴别情况

利用从上文所述的反应道得到的 MC 样本, 对新的粒子鉴别程序的鉴别能力分不同的动量段进行了检验. 所用到的选择条件与第二部分所述基本相同, 所不同的是( $e$ ), 这里对作为样本的粒子不能作任何粒子鉴别. 图 4 给出的是在不同的动量段, 新的方法对质子样本的鉴别情况. 由图可以看到, 在动量较低时, 质子鉴别的效率很高, 而在高动量区间, 质子与  $\pi, K$ , 尤其是  $K$ , 相混得比较严重.

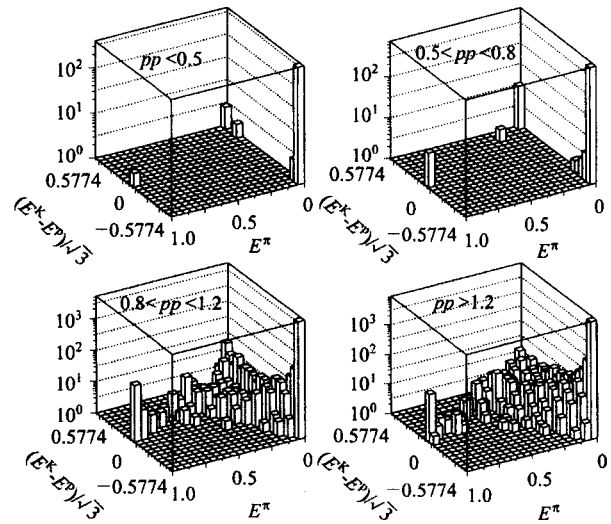


图 4 在不同的动量段, 新的程序对 MC 质子样本的鉴别情况

图 4 选择  $E^\pi$  和  $(E^K - E^p)/\sqrt{3}$  分别作为  $x$  轴和  $y$  轴. 根据定义,  $\pi, K, p/\bar{p}$  的坐标应分别为  $(1, 0), (0, 1/\sqrt{3})$  和  $(0, -1/\sqrt{3})$ , 即它们将分别位于正三角形的 3 个顶点上.

### 4 与原有鉴别方法的比较

#### 4.1 鉴别效率和误判几率的比较

判断粒子鉴别方法的好坏, 最基本的就是这两

表 3 对不同变量组合的  $K$  误判几率的检验 ( $K$  的动量大于  $1.0\text{GeV}$ )。第一、二列是利用  $\psi(2S) \rightarrow \pi^+ \pi^- K^+ K^-$  反应道得到的  $K$  样本;第三列是利用 MC Tester 产生子产生的  $K$  样本 ( $1.8\text{GeV} > P_K > 1.0\text{GeV}$ )。 ( $E^{p\bar{p}} > 0.7$ ) (“ $\checkmark$ ”表示使用该信息)

数据	MC	MC Tester	TOF-Q	BSC-E	TOF-T	dE/dx
$3.85 \pm 0.89$	$2.14 \pm 0.38$	$3.07 \pm 0.19$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
$5.54 \pm 1.08$	$3.96 \pm 0.55$	$3.69 \pm 0.21$			$\checkmark$	$\checkmark$
$4.31 \pm 0.94$	$2.38 \pm 0.41$	$3.22 \pm 0.20$	$\checkmark$		$\checkmark$	$\checkmark$
$5.04 \pm 1.02$	$3.06 \pm 0.48$	$3.74 \pm 0.21$		$\checkmark$	$\checkmark$	$\checkmark$
$6.71 \pm 1.45$	$5.22 \pm 0.79$	$5.17 \pm 0.25$			$\checkmark$	

个方面。为了定量地比较粒子鉴别的效率,利用  $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$  反应道对不同变量组合下的质子(反质子)鉴别效率进行了检验。利用运动学拟合和其余 3 条径迹的粒子鉴别,可以获得比较纯的第四条径迹的样本。对于  $E^{p\bar{p}} > 0.7$  的条件,所得的结果列于表 2。可以看到,和以往只用 TOF-T 和 dE/dx 的情况相比,新的鉴别方法对于 p 和  $\bar{p}$  的鉴别效率均有 1%—2% 的提高。从上文已经看到,在高动量区间,质子与  $\pi, K$ , 尤其是  $K$ , 相混得比较严重。对于误判几率的研究,用的是  $\psi(2S) \rightarrow \pi^+ \pi^- K^+ K^-$  反应道。对于  $E^{p\bar{p}} > 0.7$  的条件,所得的结果列于表 3。可以看到,新的方法下  $K$  的误差几率有了显著的降低,大约降为原来的 70%。

#### 4.2 系统误差的比较

将新的方法用于  $\chi_{c1} \rightarrow p\bar{p}$  的分析。分支比的系统误差的研究中所用的基本方法,就是利用干净的样本,分别研究数据和 MC 的鉴别效率,比较它们的差别<sup>[5]</sup>。由于  $\chi_c$  衰变中 p 和  $\bar{p}$  动量处于以下分析的两个样本动量之间,这里使用内推的方法。

##### 4.2.1 $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$

这个反应道的事例比较容易选择,样本比较干净。首先,用运动学拟合和其他 3 条带电径迹的粒子鉴别获得 p 和  $\bar{p}$  的样本。结果见图 5 和表 4,从中可以看到数据和 MC 之间的差别。因为选择  $p\bar{p}$  事例的时候是要求 p 和  $\bar{p}$  之一被鉴别出来即可。粒子鉴

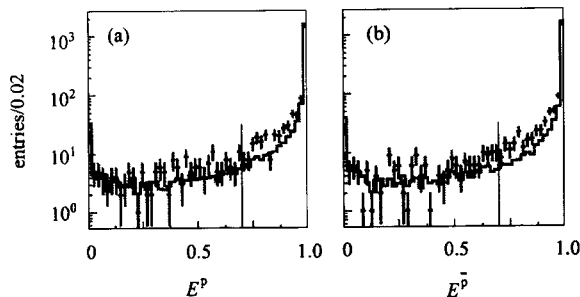


图 5 粒子鉴别中 p(a)和  $\bar{p}$ (b)的  $E^{p\bar{p}}$  的分布  
样本来自  $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$ 。误差棒是数据,直方图是 MC。

别的效率  $\epsilon_{or}$  由下式给出:

$$\epsilon_{or} = 1 - (1 - \epsilon_p)(1 - \epsilon_{\bar{p}}).$$

表 4 p 和  $\bar{p}$  的粒子鉴别效率 ( $E^{p\bar{p}} > 0.7$ )

粒子	数据 (%)	MC (%)
p	$90.13 \pm 0.61$	$92.12 \pm 0.17$
$\bar{p}$	$88.84 \pm 0.64$	$91.59 \pm 0.18$
p 或 $\bar{p}$	$98.90 \pm 0.09$	$99.34 \pm 0.02$

注:样本来自  $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$ 。

比较数据和 MC 的粒子鉴别的效率,发现对于  $J/\psi \rightarrow p\bar{p}$ ,修正因子为  $(99.558 \pm 0.095)\%$ 。其中的误差项来源于统计量的限制。可以看到,数据和 MC 的差别是很小的。我们不作修正,取 0.5% 作为粒子鉴别带来的系统误差。

##### 4.2.2 $\psi(2S) \rightarrow p\bar{p}$

为了内推的需要,用类似的方法研究了  $\psi(2S) \rightarrow p\bar{p}$  反应道。通过运动学拟合和对另一条带电径迹做较严的粒子鉴别,能够获得较纯净的 p 和  $\bar{p}$  的样本。结果见图 6。由于动量比较高,对单条带电径迹的鉴别效率比较低。但因为是要求 p 和  $\bar{p}$  之一被鉴别出来,因此仍然能够获得较高的总的鉴别效率 ( $\epsilon_{or}$ )。对于数据和 MC,其值分别为  $(92.28 \pm 0.58)\%$  和  $(92.30 \pm 0.28)\%$ 。从而得到效率修正因子应为  $(99.98 \pm 0.69)\%$ 。从这个反应道得到的粒子鉴别的系统误差约为 0.7%。因为它比从  $\psi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$  得到的结果 (0.5%) 稍大,我们采用

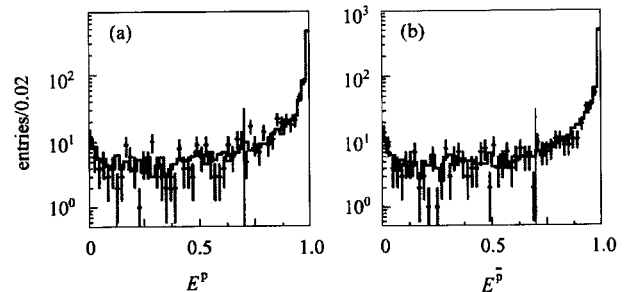


图 6 粒子鉴别中 p(a)和  $\bar{p}$ (b)的  $E^{p\bar{p}}$  的分布  
样本来自  $\psi(2S) \rightarrow p\bar{p}$ 。误差棒是数据,直方图是 MC。

0.7% 作为最终  $\chi_{\text{cl}} \rightarrow p\bar{p}$  的粒子鉴别的系统误差。

为了进行比较,在粒子鉴别中只利用 TOF 的飞行时间,应用于  $\chi_{\text{cl}} \rightarrow p\bar{p}$  的反应道。然后对系统误差进行了同样的研究。对于  $\phi(2S) \rightarrow \pi^+ \pi^- J/\psi, J/\psi \rightarrow p\bar{p}$  反应道,数据和 MC 之间的差别为  $(0.61 \pm 0.23)\%$ 。对于  $\phi(2S) \rightarrow p\bar{p}$  为  $(3.83 \pm 0.82)\%$ 。很明显,如果对  $E^{p\bar{p}}$  的 cut 采用相同的值,新的鉴别方法导致的系统误差要比原来小得多。

## 5 结果和讨论

以 BES II 物理分析中常用的粒子鉴别方法(联

合 TOF 的飞行时间和 MDC 的  $dE/dx$ )为基础,尝试增加利用 TOF 的  $Q$  值和 BSC 的沉积能量,发展了基于多变量的粒子鉴别。经检验,对  $p, \bar{p}$  的鉴别效率比原来提高了 1%—2%,K 本底的混入率大约降为原来的 70%,引入的系统误差也比原来有所降低。

此方法可以推广至更多变量,如包括径迹在 BSC 中簇射的形状<sup>[6]</sup>, $\mu$  计数器中的击中信息<sup>[7]</sup>等。此方法也可推广至其他实验。

作者感谢朱永生、薛生田研究员对本工作的建议和关注。

## 参考文献 (References)

- 1 BAI J Z et al (BES Collab.). Nucl. Instr. Meth., 2001, **A458**:627
- 2 ZHENG Zhi-Peng, ZHU Yong-Sheng. Electron-Positron Physics at Beijing Spectrometer, Guang Xi: Guang Xi Science and Technology Press, 1998. 36—42 (in Chinese)  
(郑志鹏,朱永生. 北京谱仪正负电子物理. 广西:广西科学技术出版社,1998. 36—42)
- 3 YUAN Chang-Zheng. Ph. D thesis, Beijing: Institute of High Energy Physics, 1997 (in Chinese)  
(苑长征. 博士学位论文,北京:高能物理研究所,1997)
- 4 LI Hui-Hong, LIU Feng, LI Jin et al. HEP & NP, 2003, **27**(8):659—665 (in Chinese)  
(李会红,刘峰,李金等. 高能物理与核物理,2003, **27**(8):659—665)
- 5 BAI J Z et al (BES Collab.). Phys. Rev., 2003, **D67**:112001
- 6 DU Shu-Xian, YUAN Chang-Zheng, LIU Jue-Ping et al. HEP & NP, 2003, **27**(6):521—525 (in Chinese)  
(杜书先,苑长征,刘觉平等. 高能物理与核物理,2003, **27**(6):521—525)
- 7 WANG Zheng, YUAN Chang-Zheng, QI Na-Ding. HEP & NP, 2003, **27**(8):716—721 (in Chinese)  
(王征,苑长征,漆纳丁. 高能物理与核物理,2003, **27**(8):716—721)

## Multi-Variable Particle ID for BES II Physics Analysis\*

XIN Bo<sup>1,2,1)</sup> YUAN Chang-Zheng<sup>2</sup> MA Feng-Cai<sup>1</sup>

1 (Department of Physics, Liaoning University, Shenyang 110036, China)

2 (Institute of High Energy Physics, CAS, Beijing 100039, China)

**Abstract** A multi-variable particle identification method is developed for the BES II physics analysis. Besides the traditionally used TOF- $T$  and  $dE/dx$  information, TOF- $Q$  and BSC- $E$  are also used to improve the particle discrimination ability. It is shown that the particle identification efficiency, misidentification probability and systematic error could be improved with this new method. The multi-variable method can be extended to even more variables, and can be used in other experiments.

**Key words** BES II, particle identification, efficiency, misidentification, systematic error

Received 10 November 2003

\* Supported by One Hundreds Person Project of The Chinese Academy of Sciences (U-25), Educational Research Foundation of Liaoning Province (20021003)

1) E-mail: xinb@mail.ihep.ac.cn