Kaon Selection At BaBar

Chris Marker,
Royal Holloway, University of London

Introduction

- The Silicon Vertex Tracker (SVT), Drift Chamber (DCH) and Detector of Internally reflected Cherenkov light (DIRC) provide the main source of information for identifying kaons.

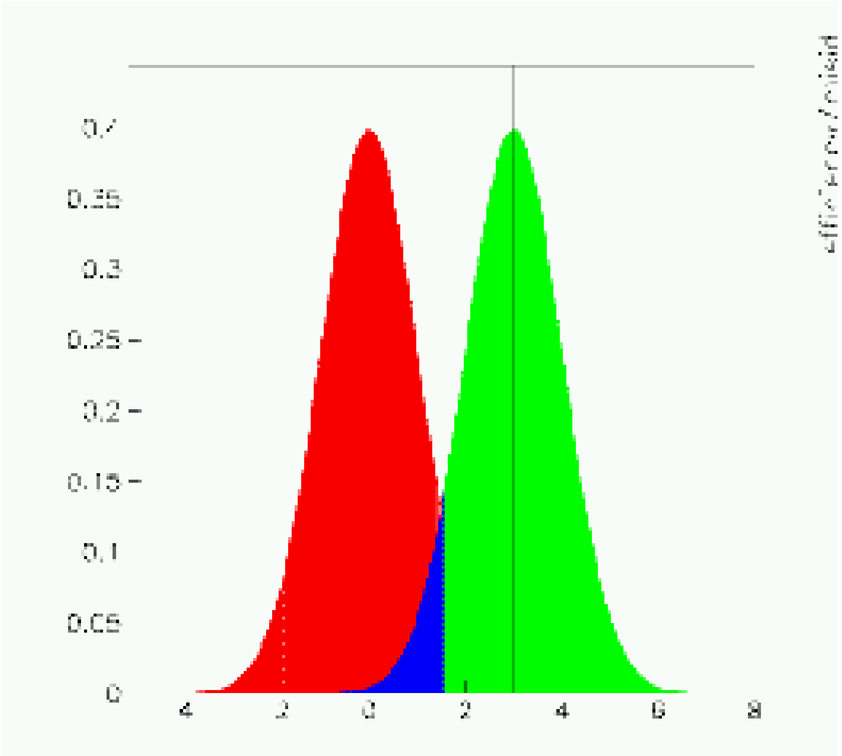
- Better than 2σ separation of pions and kaons is possible up to 0.6 GeV/c particle momentum in the SVT and 0.7 GeV/c in the DCH using dE/dX measurements. For higher momentum kaons the Cherenkov angle from the DIRC provides the main source of information for kaon identification.

Principle of Likelihood Selection

Consider the Cherenkov angle θ_c distribution for pions and kaons with respect to the expected angle θ_π θ_K. Assume distribution follows a gaussian and the the error σ_{θ_c} is known. The pull is defined as

$\Delta_i = \theta_c$- $\theta_i/\sigma_{\theta_c}$, i=,π K.

Plot Δ_K for some momentum interval.



Assuming a gaussian distribution of the Cherenkov angle around its mean(left) one can calculate the number of selected kaons and pions by a certain cut between the distributions of the two particle types.

The probability density function is a gaussian function g_i

$$g_i = 1/\sqrt{2\pi}\ \sigma_{\theta_c} e^{-0.5\Delta_i^2}$$

To separate pions from kaons choose a ratio criterion

$g_K/g_\pi >$ r.

Due to overlap between measured distributions for pions and kaons a certain fraction of pions will survive this selection.

A selection based on likelihood ratios tests the closeness to the expected curve.

The NN Based Selector

This uses sub-detector likelihoods as they are stored in the micro. Micro quantities can be corrected for miscalibration.

For the SVT and DCH the PDF is a gaussian distribution where the mean is given by the dE/dX calibration. The momentum dependent relative error is given by the calibration for both subsystems.

The DIRC likelihood is determined from a global (event) likelihood which is computed at the photomulitiplier tube hit level and takes into consideration all the tracks in an event as background for each other, as well as machine background.

The KNET Neural Network is used to combine all the information

Although a single continuous variable is the natural output five lists of kaon candidates are provided. These lists are :

- Very Tight which is optimized to keep the misid below 2% up to momenta of 4GeV/c using the DIRC at higher momentum only.

- Tight which is optimized to keep the misid below 5% up to momenta of 4GeV/c using the DIRC at higher momentum only.

- Loose , this is optimized to enhance the efficiency at higher momenta by incorporating the SVT and DCH. Keeps misid below 7% up to momenta of 4GeV.

- Very Loose , this uses the product of the likelihoods from all three detectors over the whole momentum range.

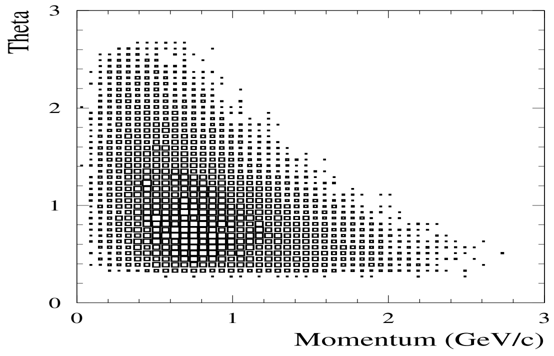
- Not A Pion , this is optimized with respect to kaon efficiency by rejecting pions.

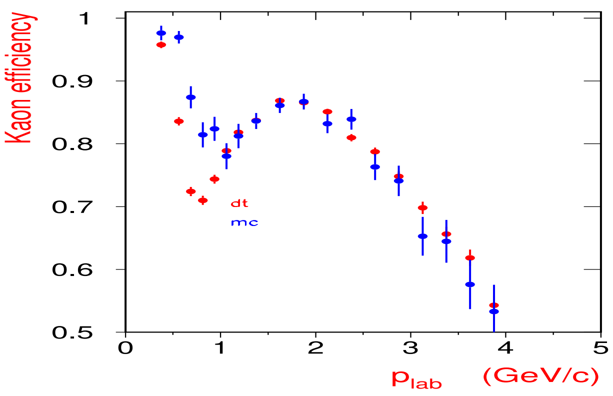
Kaon ID in the $B \rightarrow \eta_c K, \eta_c \rightarrow 4K mode$

In the decay mode being studied 3 of the kaons were required to be neural network KTight and one was required to be 1 NOT A PION.

Initial estimates for signal yield on MC proved to be a factor of four bigger than was found in data .

One of the possible reasons for this discrepancy was the difference between MC and data for the PID used, especially since most of the kaons in our sample are in the momentum region with the poorest agreement between MC and data.





All kaons from η_c

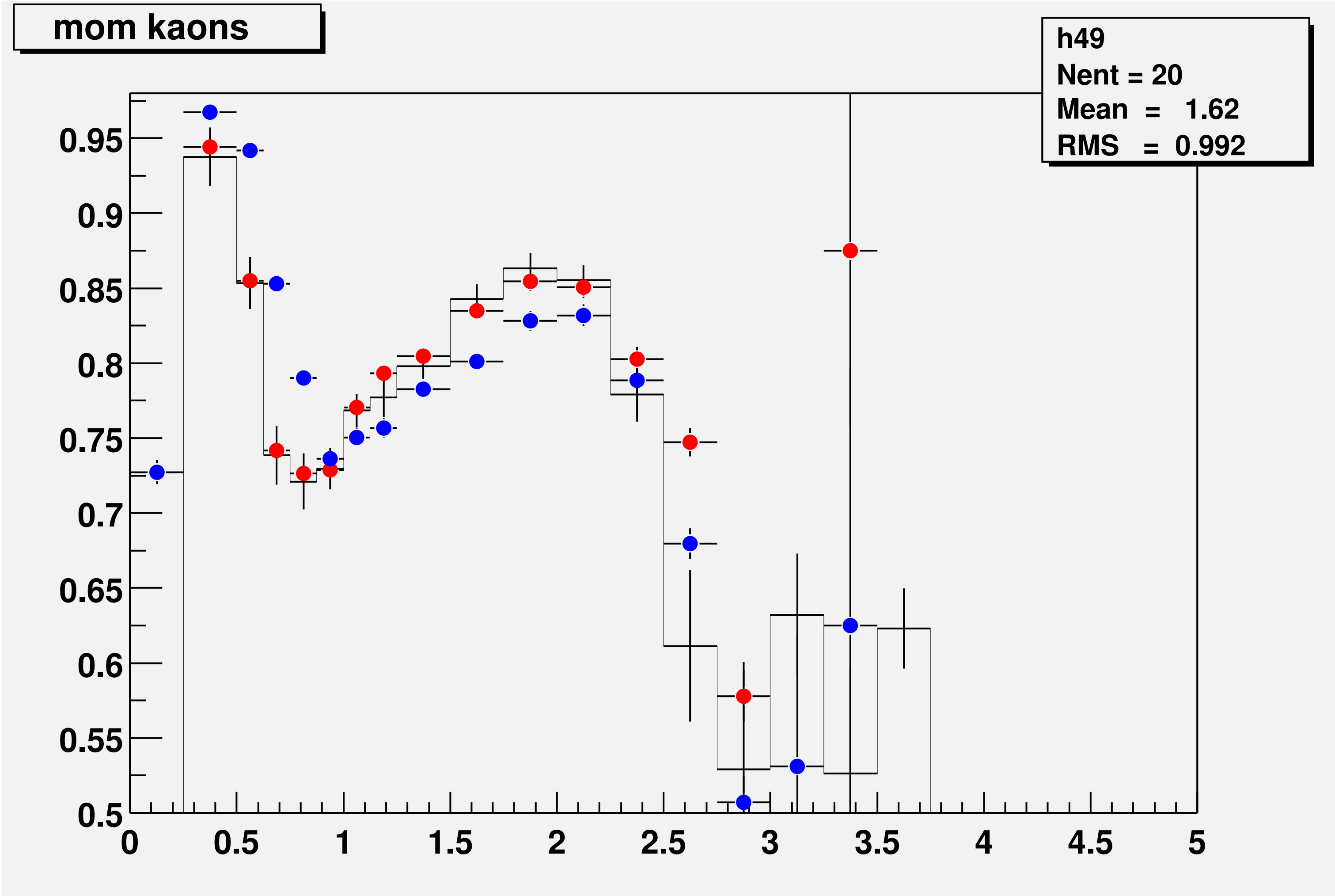
Efficiency for NN KTight PID.

(GIAMPI .MANCINELLI)

In order to make the MC look more like the data a PID killing procedure is applied to MC. This substitutes the usual PID lists (Loose, Tight ...) with ones created starting from efficiency tables computed from data control samples.

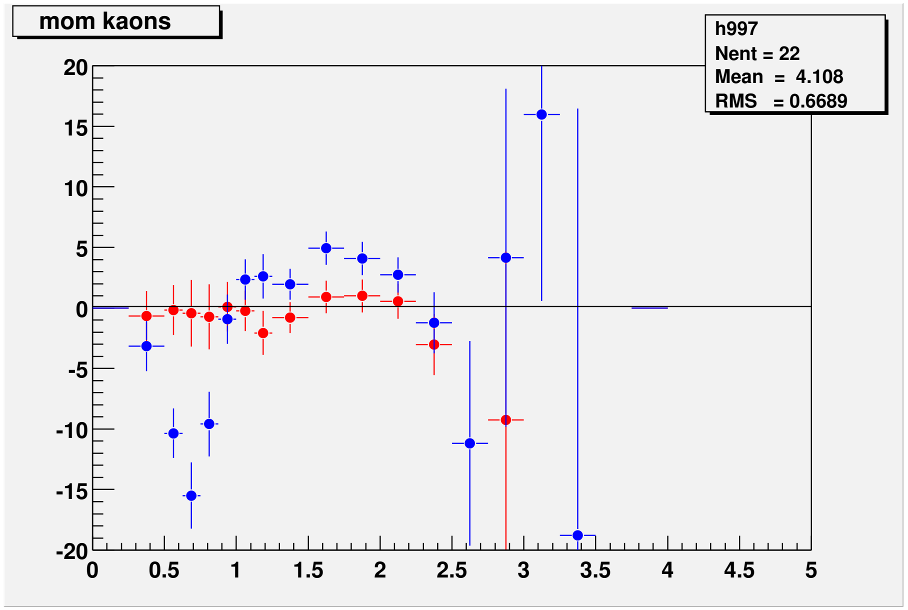
Validation of PID Killing Procedure

To make sure the PID killing was working properly plot the efficiency values from the PID efficiency tables, plot the efficiency for PID corrected (PID killed MC) and uncorrected MC.



Efficiency for KTight selector for ALL kaons, red points killed MC, blue points unkilled MC, histogram from PID tables 4K mode .

Difference between histogram and killed MC (red points) and histogram and unkilled MC and histogram (blue points) for all kaons 4K mode.



The fact that the killed MC reproduces the histogram from the PID tables well was a validation that the PID killing procedure was working correctly.

Statistical and Systematic Errors

STATISTICAL ERROR

The corrections to the PID efficiency are provided in a set of tables binned in momentum. There are different tables for positive and negative kaons , different tables for 1900 or 1960 volts data and the tables are split up into non DIRC and DIRC regions. Each bin has along with an efficiency a statistical error coming from the limited amount of statistics in the control sample used to make the PID tables.

So how do we calculate the statistical error ?

Have efficiency for a set of events that depends on the event type. The events are classified into type bins each with an efficiency ϵ_i

which corresponds to a weight $w_i = 1/\epsilon_i$. There are N_{bin} such type bins.
Each weight w_i has an error s_i dominated by MC statistics and the w_i in different bins are statistically independent.
In our case the fraction f$_i$ of events expected to be of type i is known.
Using an application of maximum likelihood an estimator for the number of events produced in our experiment N_{Prod} is

$$\ln L = \sum_{i=0}^{N_{bin}} (-f_i \epsilon_i N_{Prod} + n_i \ln(f_i \epsilon_i N_{Prod})$$

where N_{bin} is the number of event types and n_i is the number of events of type i. Take the derivative and solve for $N_{Prod} = 0$ to find the maximum of lnL.

$$0 = d \ln L / dN_{Prod} = N_{event}/N_{prod} - \sum_{i=0}^{N_{bin}} f_i \epsilon_i$$

Where $N_{event} = \sum n_i$ is the total number of events in our data samle. The best estmate of N_{prod} is then

$N_{prod} = N_{event} \; / \sum f_i \epsilon_i = N_{event} / \; \overline{\epsilon}$

To get the error on N_{prod} take the average of $(\delta N_{prod})^2$ over a hypothetical infinite ensemble of MC experiments similar to ours

$\delta N_{prod} =$ - $N_{event} / \overline{\epsilon}^2 \times \delta \overline{\epsilon}$

with $\delta\ \bar{\epsilon} = \sum f_i \delta\epsilon_i$

Putting all together, squaring and averging yields

$$\sigma_{N_{prod}} = \mathrm{N}_{event}/\bar{\epsilon}^2 \times \sqrt{\sum(f_i \epsilon_i^2 s_i)^2} = \mathrm{N}_{event}/\bar{\epsilon}^2 \sqrt{\sum(f_i \sigma_{\epsilon,i})^2}$$

Where $\sigma_{\epsilon,i}$ are the errors on the efficiencies.

NOT A PION kaons 1900 volts relative statistical error = 0.00220

KTight kaons 1900 relative statistical error = 0.00478
We have 3 KTight koans
Multiply relative error by 3 for quadratic sum = 0.00828
Multiply relative error by 3 for linear sum = 0.01434

In order to be conservative take 0.01434 as error , also because we have not taken into account correlations between kaons

Fast kaon (KVeryLoose) 1900 volts relative statistical error = 0.00281

.

SYSTEMATIC ERROR

There are two components, one that has to be added linearly if there is more than one kaon in the event and has to do with background subtraction in determining the efficiencies from the control sample used to make the PID tables (whose value is 1.36%) and one (not perfectly evaluated yet but should be smaller than 0.4%) that has to be added linearly that is the statistical error on the background subtraction itself.

So if we have n charged kaons identified by a NN or SMS selector in the event the relative systematical error coming from PID is:

4K K+ (5 kaons) 5.67% , 4KKs (4 kaons) 7.09%