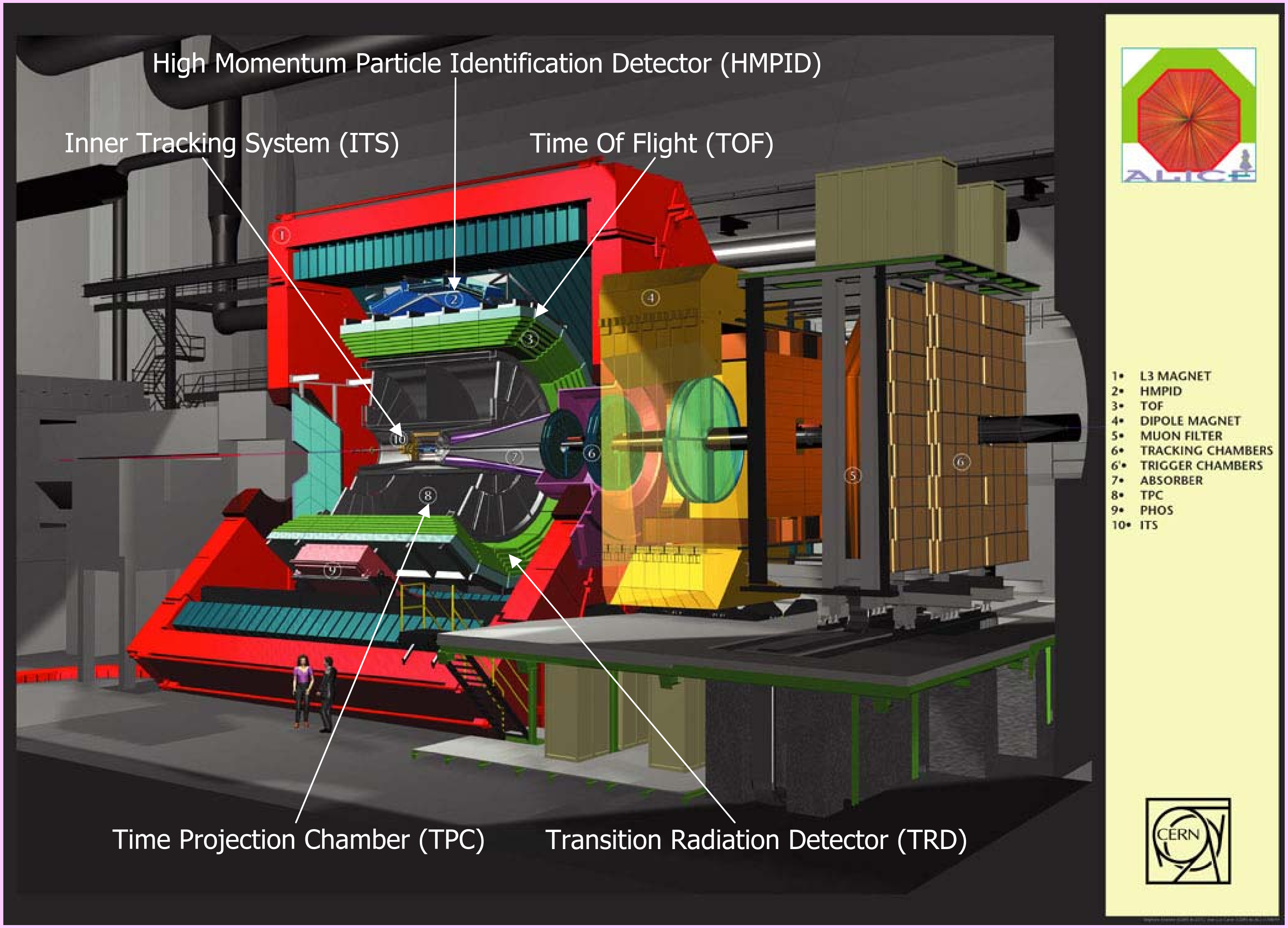
Bayesian Approach for Combined Particle Identification in ALICE Experiment at LHC.

I.Belikov, P.Hristov, M.Ivanov, T.Kuhr, K.Safarik
CERN, Geneva, Switzerland

Outline

- Layout of the ALICE detector.
- The problem of Particle Identification (PID) in ALICE.
- Bayesian PID with a single detector.
  - Obtaining the conditional probability density functions.
  - Obtaining the *a priori* probabilities.
- PID combined over several detectors.
- Conclusions.

Layout of the ALICE detector



The problem of PID in ALICE

- Very big data volume (10^7 events × 10^4 tracks).
  - The procedure must be as much as possible automatic.
- Very broad momentum range (0.1 – 10 GeV/c).
  - PID signals of different nature (*dE/dx*, TOF, Cherenkov, …).
  - Distributions of the PID signals for particles of different types may significantly overlap (each track is assigned a set of PID weight).
- Very rich physics program (particle spectra – jets/beauty).
  - The final PID depends on a particular analysis (event and track selection).

Bayesian PID with a single detector

Probability to be a particle of i-type ($i = e, \mu, \pi, K, p, \dots$),
if the PID signal in the detector is s:

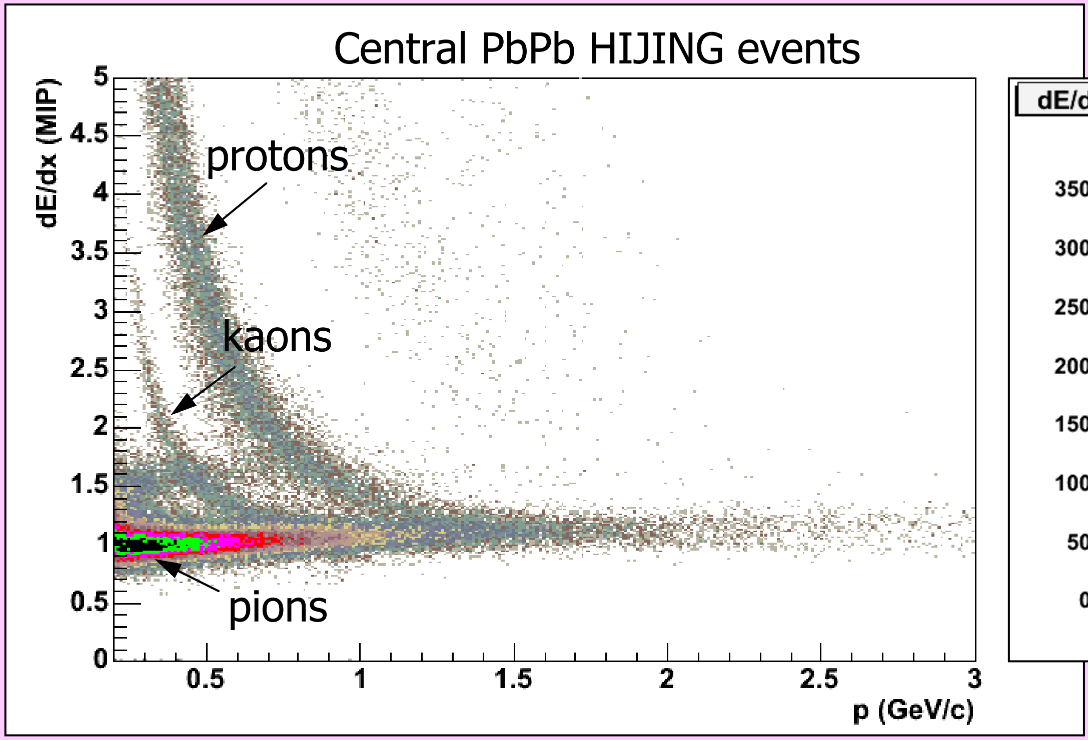
$$\boxed{w(i \mid s) = \frac{C_i r(s|i)}{\sum_{k=e,\mu,\pi,\dots} C_k r(s \mid k)}}$$

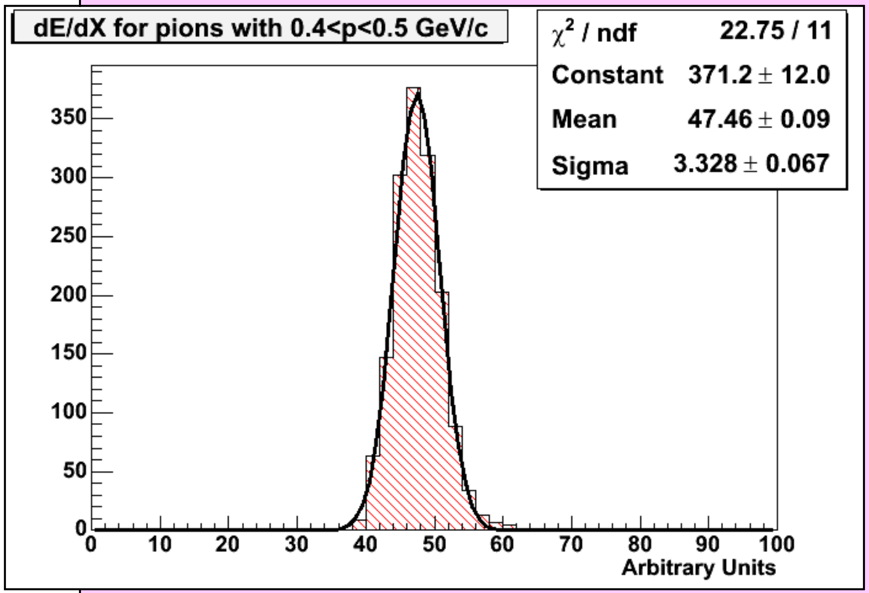
- C_i - *a priori* probabilities to be a particle of the i-type. "Particle concentrations", that depend on the track selection.
- $r(s|i)$ – conditional probability density functions to get the signal s, if a particle of i-type hits the detector.
  "Detector response functions", that depend on properties of the detector.

Both the "particle concentrations " and the "detector response functions" can be extracted from the data.

Bayesian PID with a single detector

Example of obtaining the conditional PDFs: “TPC response function”

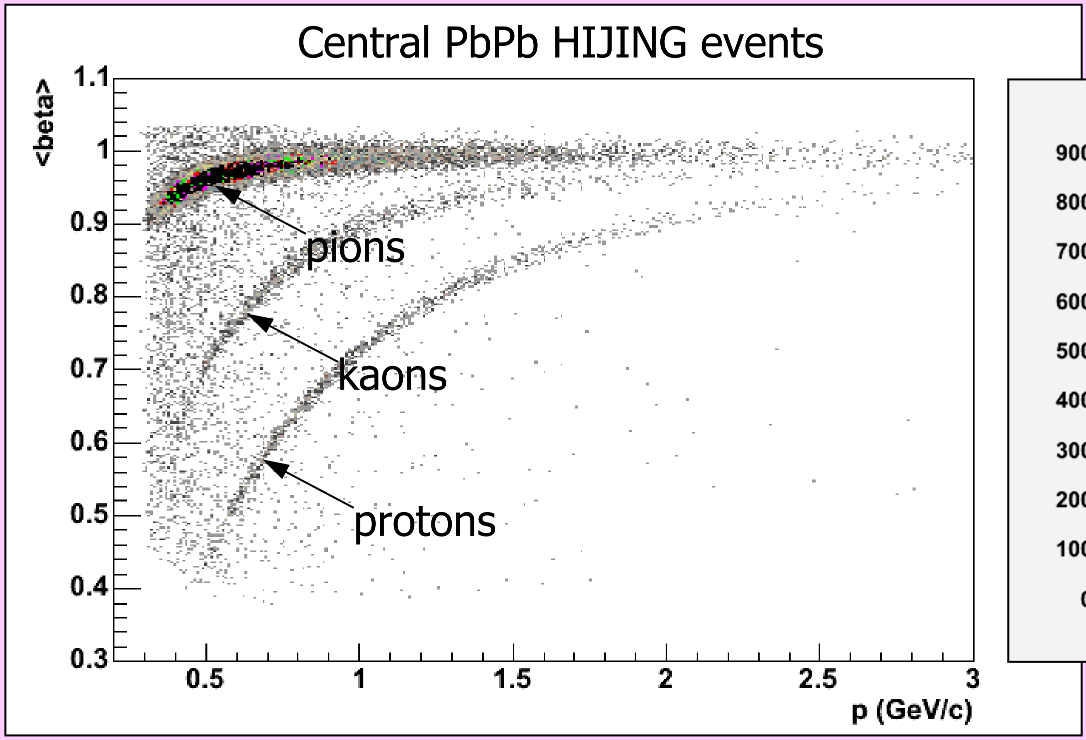


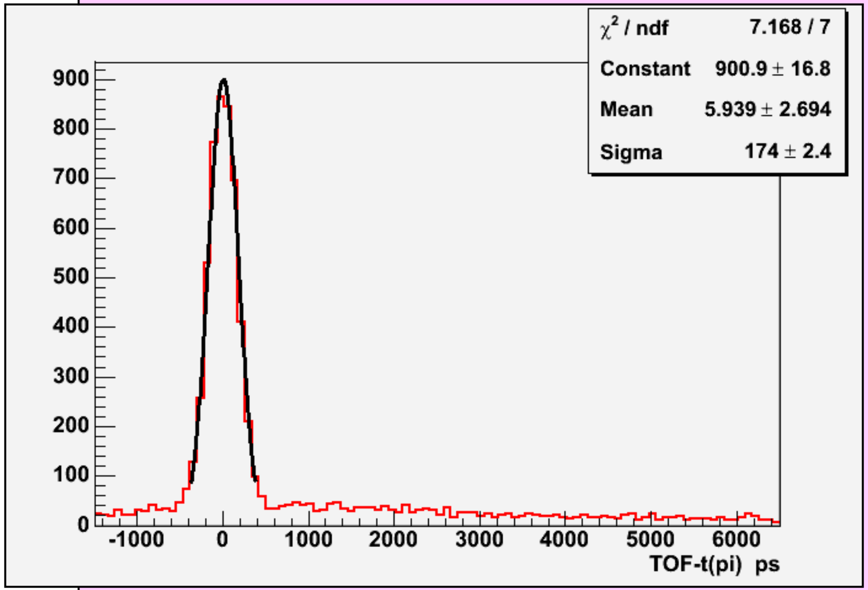


For each momentum p the function $r(s|i)$ is a Gaussian with centroid $<dE/dx>$ given by the Bethe-Bloch formula and sigma $\sigma = 0.08<dE/dx>$

Bayesian PID with a single detector

Example of obtaining the conditional PDFs: "TOF response function"

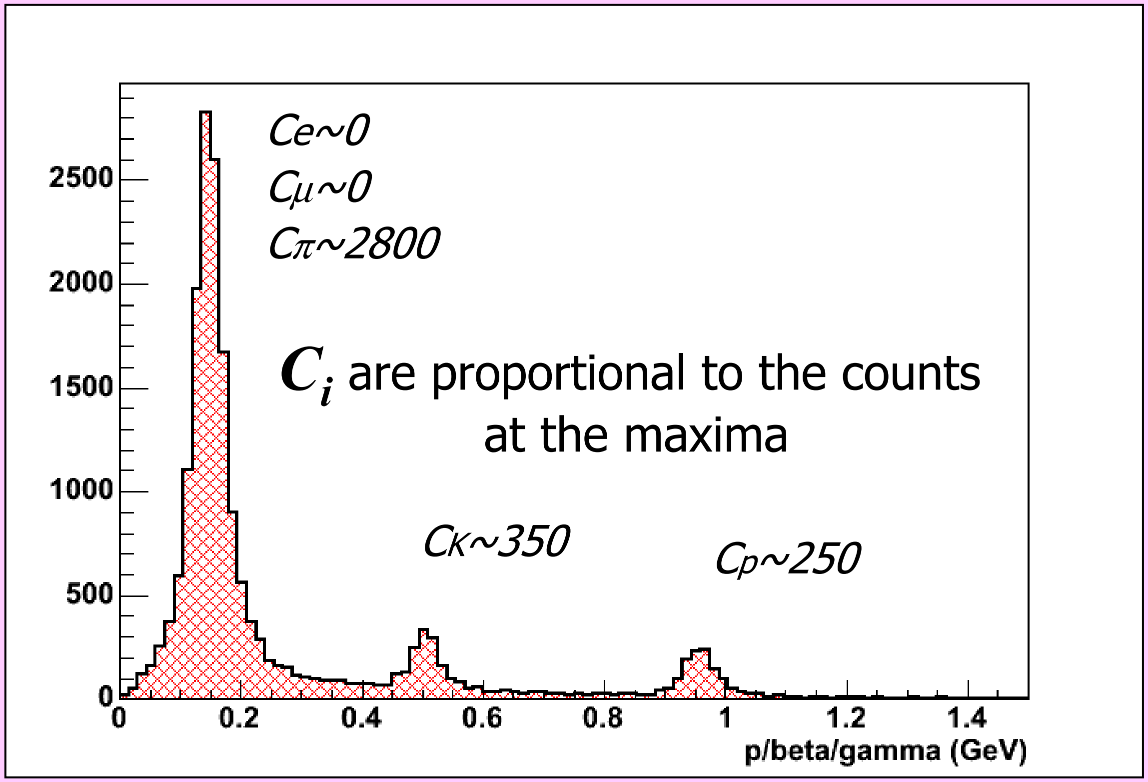




For each momentum $\boldsymbol{p}$ the function $\boldsymbol{r(s|i)}$ is a Gaussian with centroid at $\boldsymbol{0}$ and $\boldsymbol{\sigma}$ given by the distribution of $\boldsymbol{(TOF\text{-}t_\pi)}$, $\boldsymbol{t_\pi}$– time calculated by the tracking for the pion mass hypothesis.

Bayesian PID with a single detector

Example of obtaining the *a-priori* probabilities ("particle concentrations")



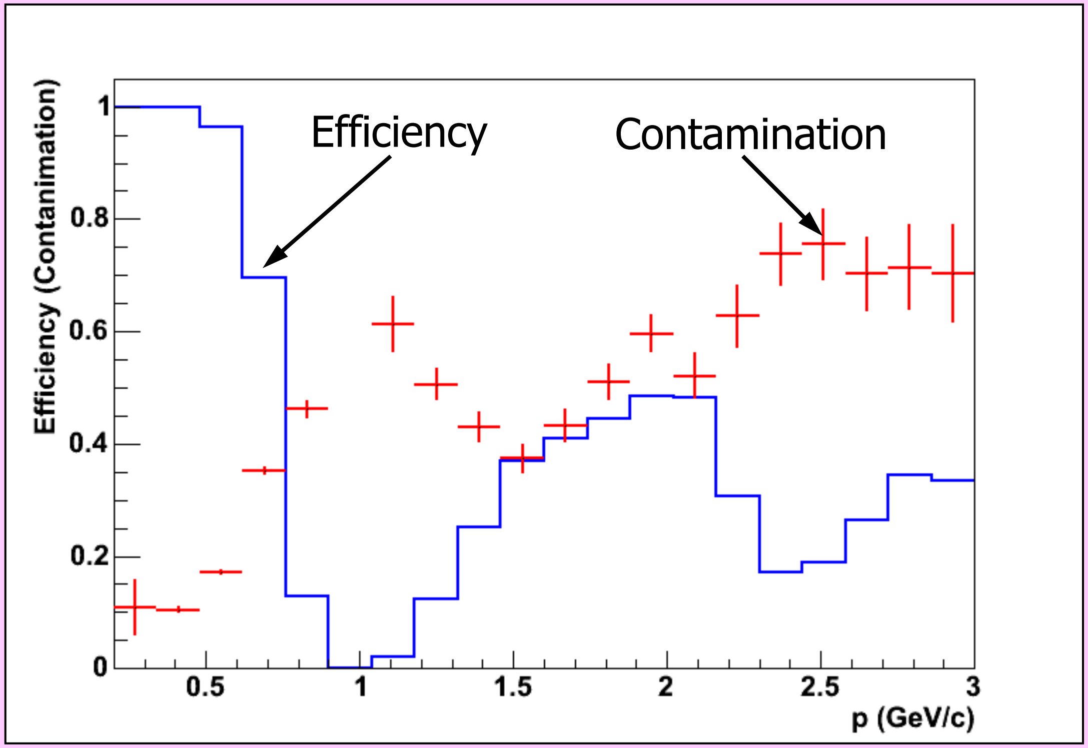
Selection ITS & TPC & TOF
Central PbPb HIJING events

- *p* – track momentum measured by the tracking
- *beta=L/TOF*
  L – track length measured by the tracking

The "particle concentrations" depend on the event and track selection !

PID with the TPC (Kaons only)

Selection : ITS & TPC & TOF (central PbPb HIJING events)



PID decision:

w(K|s) is the biggest.

Nt – number of true Kaon tracks;

Nc – number of correctly identified Kaon tracks;

Nw – number of wrongly identified Kaon tracks;

Efficiency = ***Nc/Nt*** Contamination = ***Nw/(Nc+Nw)***

PID combined over several detectors

Probability to be a particle of $\boldsymbol{i}$-type ($\boldsymbol{i = e, \mu, \pi, K, p, \ldots}$),
if we observe a vector $\boldsymbol{S = \{s_{ITS}, s_{TPC}, s_{TOF}, \ldots\}}$ of PID signals:

$$\boxed{W(i \mid S) = \frac{C_i R(S|i)}{\sum_{k=e,\mu,\pi,\ldots} C_k R(S \mid i)}}$$

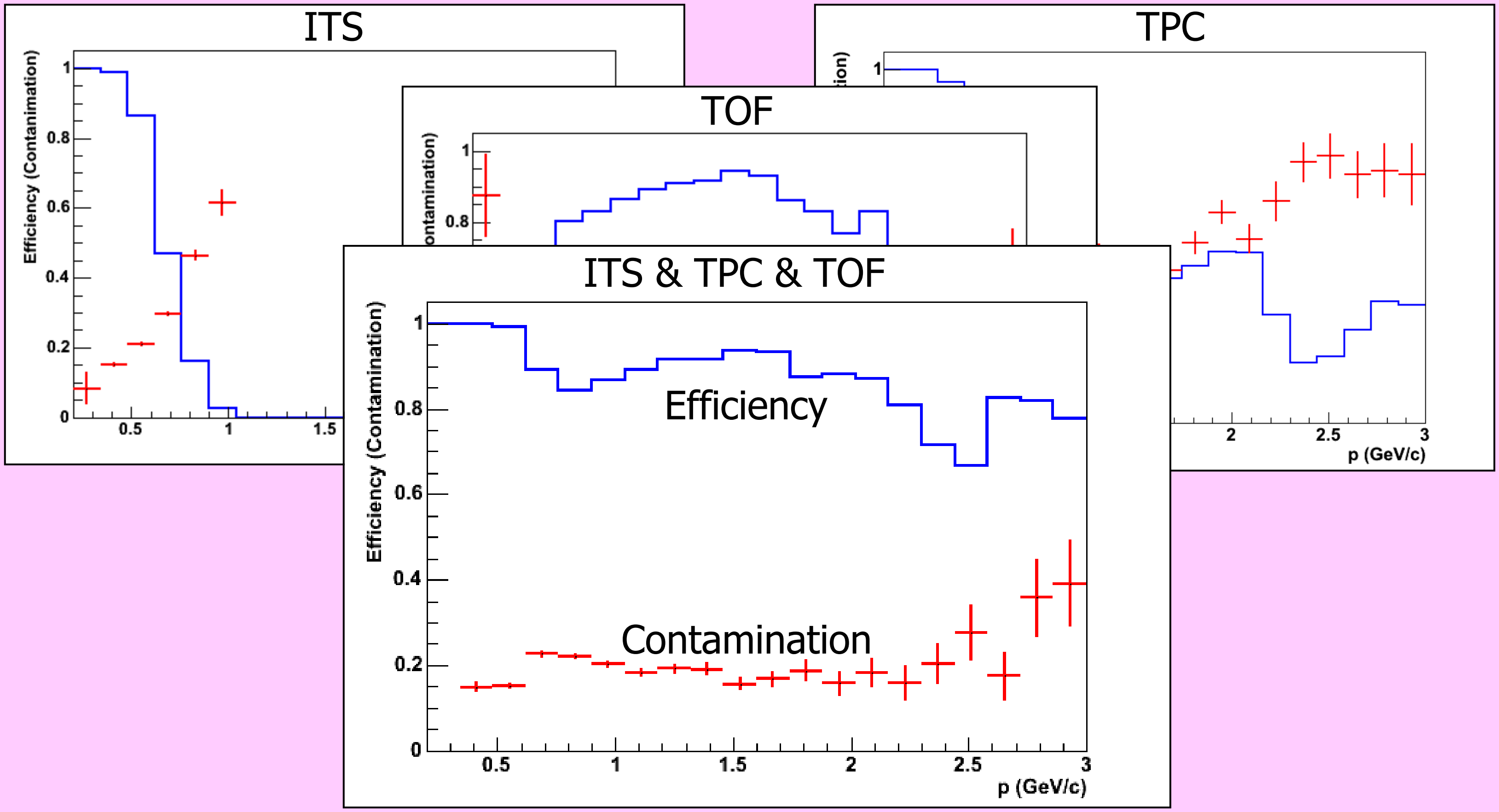
$R(S \mid i) \approx \prod_{d=ITS,TPC,\ldots} r_d(s_d \mid i)$ is the combined response function.

$\boldsymbol{C_i}$ are the same as in the single detector case (or even something reasonably arbitrary like $\boldsymbol{C_e \sim 0.1, C_\mu \sim 0.1, C_\pi \sim 7, C_K \sim 1}$, …)

The functions $\boldsymbol{R(S|i)}$ are not necessarily “formulas” (can be “procedures”).
Some other effects (like mis-measurements) can be accounted for.

PID combined over ITS, TPC and TOF (Kaons)

Selection : ITS & TPC & TOF (central PbPb HIJING events)



Efficiency of the combined PID is higher (or equal) and the contamination is lower (or equal) than the ones given by any of the detectors stand-alone.

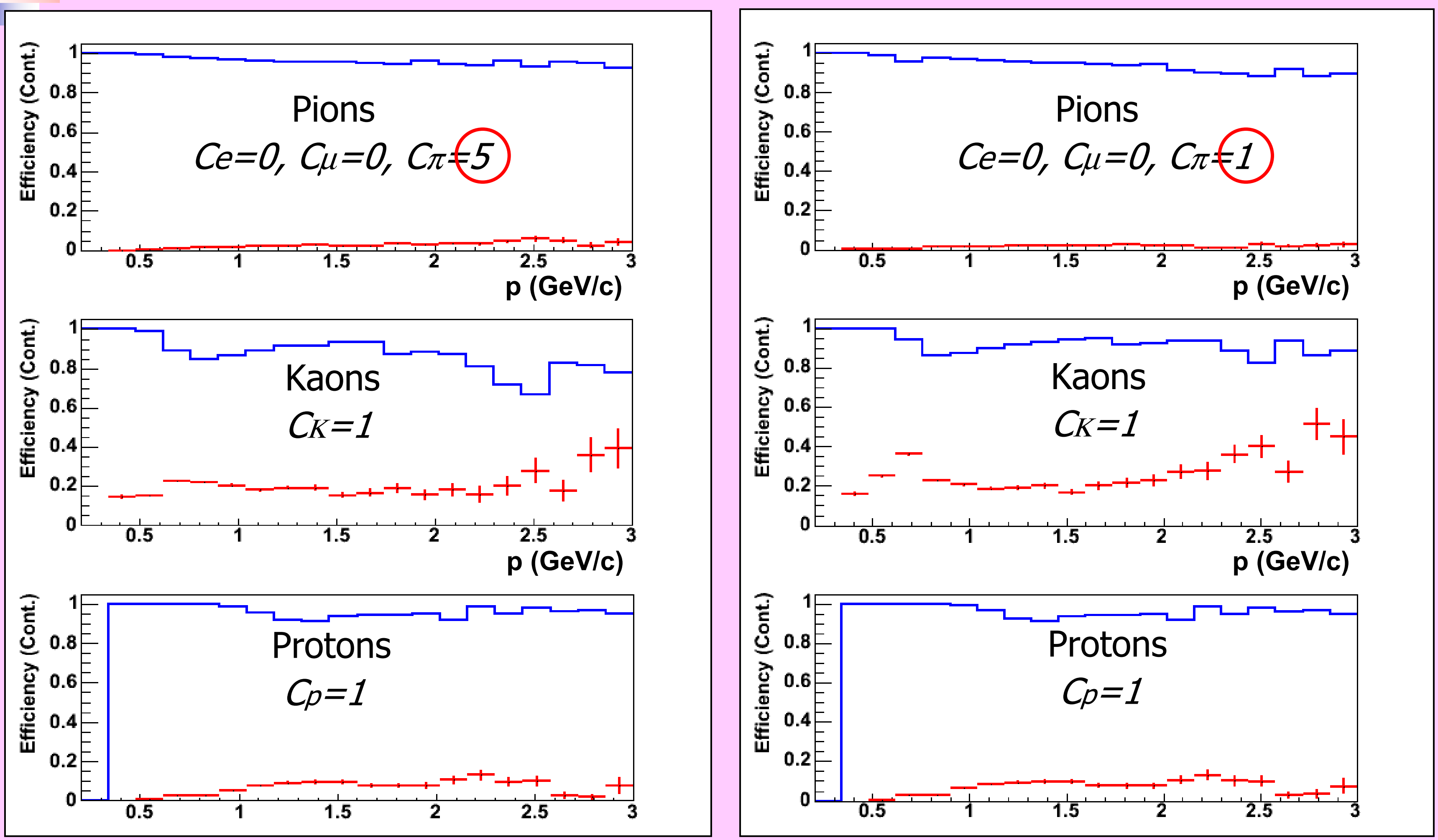
Conclusions

- Particle Identification in ALICE is done in a Bayesian way. It consists of three parts:
  - Calibration part, performed by the calibration software. Obtaining the single detector response functions.
  - "Constant part", performed by the reconstruction software. Calculating (for each track) the values of detector response functions, combining them and writing the result to the Event Summary Data (ESD).
  - "Variable part", performed by the analysis software. Estimating (for a subset of tracks selected for a particular analysis) the concentrations of particles of each type, calculating the final PID weights by means of Bayes' formula using these particle concentrations and the combined response stored in the ESD.

Conclusions

- The procedure allows to combine PID signals of quite different nature (***dE/dx***,TOF,Cherenkov,…) in a common way.
- It naturally takes into account the fact that the PID depends, due to different event and track selection, on a particular kind of performed physics analysis.
- The procedure is fully automatic.
No cuts (graphical/multidimensional) are involved.

Stability of the procedure



In this momentum region we are not so sensitive to choice of the *a priori* probabilities C_i !