Event Reconstruction and Particle Identification with ALICE

I. Belikov for the ALICE detector performance group (PWG1)

- Introduction
- Offline tracking with the central ALICE detectors
- Charged-particle ID with the central ALICE detectors
- Appendix: A few unclear questions…

Workshop ALICE-France, IPHC Strasbourg, 19 June 2009

Reconstruction and PID: What is this ?

- (Offline) reconstruction and PID is software.
- Reconstruction is everything between the raw data and ESD.
  - The raw data are files containing encoded "positions & ionization" (in pad/wire number, ADC/TDC counts etc).
  - The Event Summary Data are files containing fitted particle momenta and vertex positions, probabilities related to PID (in GeV/c, cm, etc).
- Particle Identification is everything that provides some information about the masses of the registered particles.
  - PID is tightly connected with the reconstruction (tracking needs masses, fitted momenta are needed for PID). PID is based on quite similar algorithms as many parts of reconstruction. PID contributes to ESD. This is a part reconstruction.
  - However, PID extends beyond the recontruction towards the physics analysis. This is quite fundamental…

Reconstruction and PID: What is this ?

- (Offline) reconstruction and PID is software.
- Reconstruction is everything between the raw data and ESD.
  - The raw data are files containing encoded "positions & ionization" (in pad/wire number, ADC/TDC counts etc).
  - The Event Summary Data are files containing fitted particle momenta and vertex positions, probabilities related to PID (in GeV/c, cm, etc).
- Particle Identification is everything that provides some information about the masses of the registered particles.
  - PID is tightly connected with the reconstruction (tracking needs masses, fitted momenta are needed for PID). PID is based on quite similar algorithms as many parts of reconstruction. PID contributes to ESD. This is a part reconstruction.
  - However, PID extends beyond the recontruction towards the physics analysis. This is quite fundamental…

Reconstruction | PID | Physics analysis

A view on Reconstruction and PID

Mathematics	Statistics
Programming	Physics

A view on Reconstruction and PID

Mathematics

Statistics

Reconstruction & PID

Programming

Physics

A view on Reconstruction and PID

Mathematics

Statistics

Analyt. geometry
Diff. equations
...

Bayesian methods
Kalman filter
...

Reconstruction & PID

...
Num. algorithms
Design/implem.

...
Particle prop.
Material prop.

Programming

Physics

A view on the statistical methods (used in ALICE reconstruction and PID)

Bayesian methods

A view on the statistical methods (used in ALICE reconstruction and PID)

Bayesian methods

Likelihood

(Bayesian, with equal priors)

A view on the statistical methods (used in ALICE reconstruction and PID)

Bayesian methods

Likelihood
(Bayesian, with equal priors)

Kalman filter
(Likelihood, with Gaussian PDFs)

A view on the statistical methods (used in ALICE reconstruction and PID)

Bayesian methods

Likelihood
(Bayesian, with equal priors)

Kalman filter
(Likelihood, with Gaussian PDFs)

We are quite used to the Kalman-filter-based tracking, to likelihood-based fitting of histograms, to the Bayesian approach in PID…

A view on the statistical methods (used in ALICE reconstruction and PID)

Bayesian methods

Likelihood
(Bayesian, with equal priors)

Kalman filter
(Likelihood, with Gaussian PDFs)

We are quite used to the Kalman-filter-based tracking, to likelihood-based fitting of histograms, to the Bayesian approach in PID…

However, there is nothing wrong about fitting histograms with Kalman filter, or doing a Bayesian tracking ! ☺

What is the Kalman filter ?

Example: measuring the "length of a table"

$l(i),\ i=1,\ldots,N$ -- the results of N successive independent measurements

σ – the precision of the measurements

The classical estimate:

- $l(1,\ldots,N) = [l(1)+l(2)+\ldots+l(N)]\ /N\ ;\qquad \sigma^2(1,\ldots,N) = \sigma^2/N$

What is the Kalman filter ?

Example: measuring the "length of a table"

$l(i),\ i=1,\ldots,N$ -- the results of N successive independent measurements

σ – the precision of the measurements

The Kalman filter estimate:

1. $l(1),\ \sigma$ (the seed)

2. $l(1,2) = \frac{1}{2}\, l(1) + \frac{1}{2}\, l(2) = \frac{1}{2}\, l(1) + \frac{1}{2}\, m_2$; ($l(i) \equiv m_i,\ i>1$, the measurements)

   $\sigma^2(1,2) = \sigma_{12}^2 = \frac{1}{4}\,\sigma^2(1) + \frac{1}{4}\,\sigma^2(2) = [1/\sigma^2 + 1/\sigma^2]^{-1} = \frac{1}{2}\,\sigma^2$

   …

k+1. $l(1,\ldots,k+1) = w_{1,\ldots,k}\, l(1,\ldots,k) + w_k m_{k+1} = \dfrac{1/\sigma^2_{1,\ldots,k}}{1/\sigma^2_{1,\ldots,k} + 1/\sigma^2}\, l(1,\ldots,k) + \dfrac{1/\sigma^2}{1/\sigma^2_{1,\ldots,k} + 1/\sigma^2}\, m_{k+1}$

   $\sigma^2_{1,\ldots,k+1} = \left[1/\sigma^2_{1,\ldots,k} + 1/\sigma^2 \right]^{-1}$

   …

What is the Kalman filter ?

Example: measuring the "length of a table"

$l(i),\ i=1,\ldots,N$ -- the results of N successive independent measurements

σ – the precision of the measurements

The Kalman filter estimate:

1. $l(1),\ \sigma$ (the seed)

2. $l(1,2) = \frac{1}{2}\, l(1) + \frac{1}{2}\, l(2) = \frac{1}{2}\, l(1) + \frac{1}{2}\, m_2$; $(l(i) \equiv m_i,\ i>1$, the measurements)

   $\sigma^2(1,2) = \sigma_{12}^2 = \frac{1}{4}\,\sigma^2(1) + \frac{1}{4}\,\sigma^2(2) = [1/\sigma^2 + 1/\sigma^2]^{-1} = \frac{1}{2}\,\sigma^2$

   …

k+1. $l(1,\ldots,k+1) = w_{1,\ldots,k}\, l(1,\ldots,k) + w_k m_{k+1} = \dfrac{1/\sigma^2_{1,\ldots,k}}{1/\sigma^2_{1,\ldots,k} + 1/\sigma^2}\, l(1,\ldots,k) + \dfrac{1/\sigma^2}{1/\sigma^2_{1,\ldots,k} + 1/\sigma^2}\, m_{k+1}$

   $\sigma^2_{1,\ldots,k+1} = \left[1/\sigma^2_{1,\ldots,k} + 1/\sigma^2 \right]^{-1}$

…

The final result is the same as the classical !

What is so special about the Kalman filter ?

What is so special about the Kalman filter ?

- The nature (or/and the hardware) may work sequentially (weather forecasts, radar applications…)

What is so special about the Kalman filter ?

- The nature (or/and the hardware) may work sequentially (weather forecasts, radar applications…)
- The "length-of-the-table" may change between the measurement. Dynamically (daily variations of temperature) or/and stochastically (weather variations of temperature)

What is so special about the Kalman filter ?

- The nature (or/and the hardware) may work sequentially (weather forecasts, radar applications…)
- The "length-of-the-table" may change between the measurement. Dynamically (daily variations of temperature) or/and stochastically (weather variations of temperature)

All this is quite natural in the Kalman approach

What is so special about the Kalman filter ?

- The nature (or/and the hardware) may work sequentially (weather forecasts, radar applications…)
- The "length-of-the-table" may change between the measurement. Dynamically (daily variations of temperature) or/and stochastically (weather variations of temperature)

  All this is quite natural in the Kalman approach

- Can all this be addressed also in the classical approach ?

What is so special about the Kalman filter ?

- The nature (or/and the hardware) may work sequentially (weather forecasts, radar applications…)
- The "length-of-the-table" may change between the measurement. Dynamically (daily variations of temperature) or/and stochastically (weather variations of temperature)

All this is quite natural in the Kalman approach

- Can all this be addressed also in the classical approach ?

Surely YES !

So, why have we chosen the Kalman filter in ALICE ?

So, why have we chosen the Kalman filter in ALICE ?

- Computational convenience
(with the conventional, “sequential” hardware)

So, why have we chosen the Kalman filter in ALICE ?

- Computational convenience
  (with the conventional, “sequential” hardware)
- Inertia of thinking (after the great success of the famous article by P.Billoir in NIM in 1984)

What is the Bayesian approach ?

Example: a supervisor choosing a summer student

What is the Bayesian approach ?

Example: a supervisor choosing a summer student

- What is the probability of choosing a girl or a boy ?

What is the Bayesian approach ?

Example: a supervisor choosing a summer student

- What is the probability of choosing a girl or a boy ?
- The answer depends on:
  - Probability with which this supervisor chooses a girl $p(g)$, or a boy $p(b)$
  - The number of application submitted by girls Ng, and by boys Nb (the priors)

What is the Bayesian approach ?

Example: a supervisor choosing a summer student

- What is the probability of choosing a girl or a boy ?
- The answer depends on:
  - Probability with which this supervisor chooses a girl $p(g)$, or a boy $p(b)$
  - The number of application submitted by girls Ng, and by boys Nb (the priors)
- The final probability is given by Bayes' formula:

What is the Bayesian approach ?

Example: a supervisor choosing a summer student

- What is the probability of choosing a girl or a boy ?
- The answer depends on:
  - Probability with which this supervisor chooses a girl $p(g)$, or a boy $p(b)$
  - The number of application submitted by girls Ng, and by boys Nb (the priors)
- The final probability is given by Bayes' formula:

$$P_g = \frac{p(g)Ng}{p(g)Ng + p(b)Nb}$$

$$P_b = \frac{p(b)Nb}{p(g)Ng + p(b)Nb}$$

What is so special about the Bayesian approach ?

What is so special about the Bayesian approach ?

- It puts together two quite different sources of information
  - The supervisor's preferences are property of the supervisor. They, probably, do not change with time, and so can be precalculated.
  - The numbers of applications by girls/boys is an example of the conditions that are completely external to the supervisor. These may change year-by-year, and so, fundamentally, cannot be calculated once and forever.

What is so special about the Bayesian approach ?

- It puts together two quite different sources of information
  - The supervisor's preferences are property of the supervisor. They, probably, do not change with time, and so can be precalculated.
  - The numbers of applications by girls/boys is an example of the conditions that are completely external to the supervisor. These may change year-by-year, and so, fundamentally, cannot be calculated once and forever.

- Can all this be addressed also in the classical approach ?

What is so special about the Bayesian approach ?

- It puts together two quite different sources of information
  - The supervisor's preferences are property of the supervisor. They, probably, do not change with time, and so can be precalculated.
  - The numbers of applications by girls/boys is an example of the conditions that are completely external to the supervisor. These may change year-by-year, and so, fundamentally, cannot be calculated once and forever.

- Can all this be addressed also in the classical approach ?

Yes… Probably…

So, why are we doing the Bayesian approach in ALICE ?

So, why are we doing the Bayesian approach in ALICE ?

- The explicit factorization of what can be done in reconstruction ("supervisor's preferences"), and what has to be done in physics analysis ("number of applications"). Already at the level of a single detector.

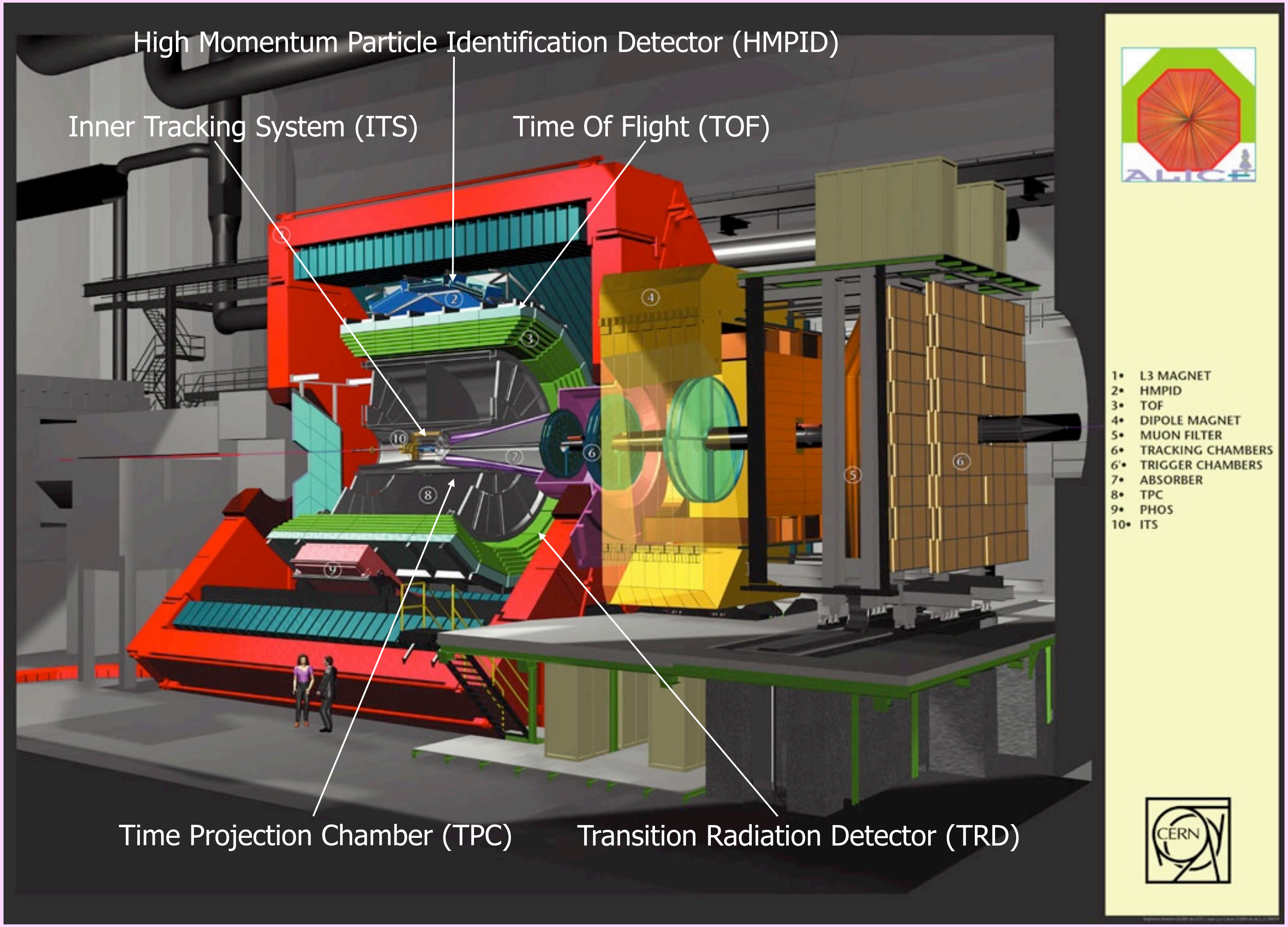
So, why are we doing the Bayesian approach in ALICE ?

- The explicit factorization of what can be done in reconstruction (“supervisor’s preferences”), and what has to be done in physics analysis (“number of applications”). Already at the level of a single detector.
- Computational convenience, when it comes to combining the PID information over the contributing detectors.

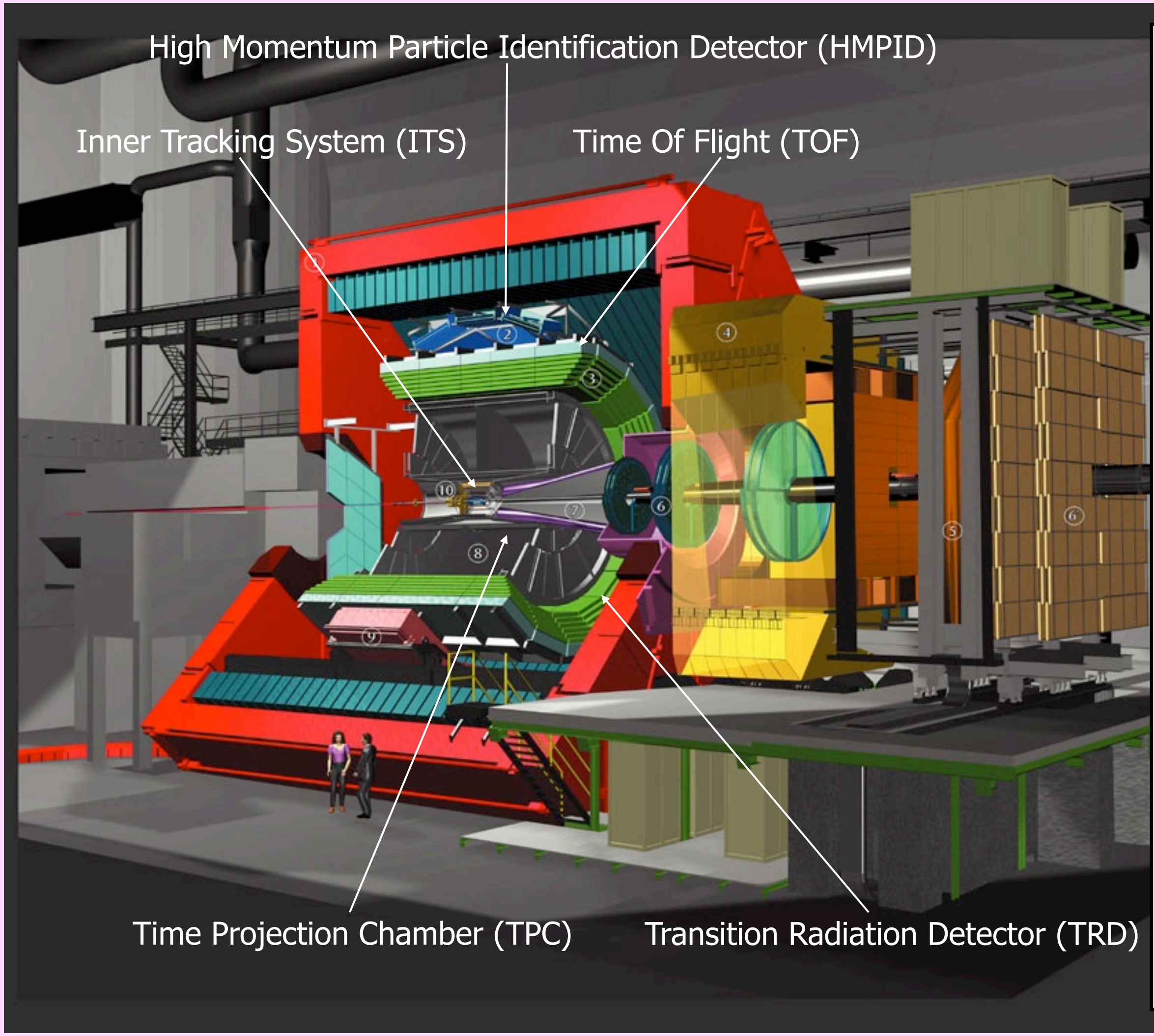
So, why are we doing the Bayesian approach in ALICE ?

- The explicit factorization of what can be done in reconstruction (“supervisor’s preferences”), and what has to be done in physics analysis (“number of applications”). Already at the level of a single detector.
- Computational convenience, when it comes to combining the PID information over the contributing detectors.
- Gain in the disk space at the level of AOD.

ALICE experiment at LHC



ALICE experiment at LHC



TPC ($-0.9<\eta<0.9$) tracking efficiency:
~80% for Pt<0.2 GeV/c
(limited by decays),
~90% for Pt>1 GeV/c
(limited by dead zones),
for > 10000 tracks in the TPC.

Momentum resolution (B=0.5 T):
~1% at Pt=1 GeV/c,
~5% at Pt=100 GeV/c (ITS+TPC+…).

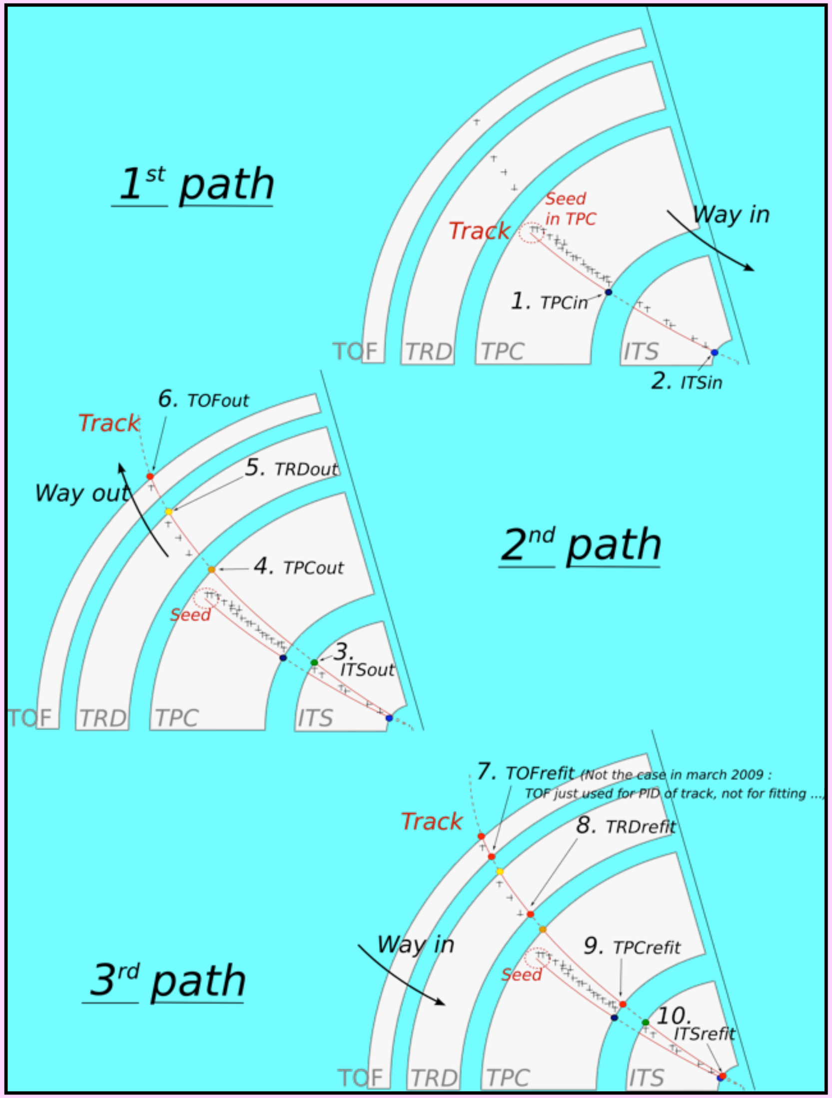
Precise secondary vertexing
better than 100 μm (ITS).

Excellent charged PID capability:
from P~0.1 GeV/c upto a few GeV/c,
(upto a few tens GeV/c, TPC rel. rise),
electrons in TRD, P>1GeV/c
(ITS+TPC+TRD+TOF+HMPID+…).

The general reconstruction strategy

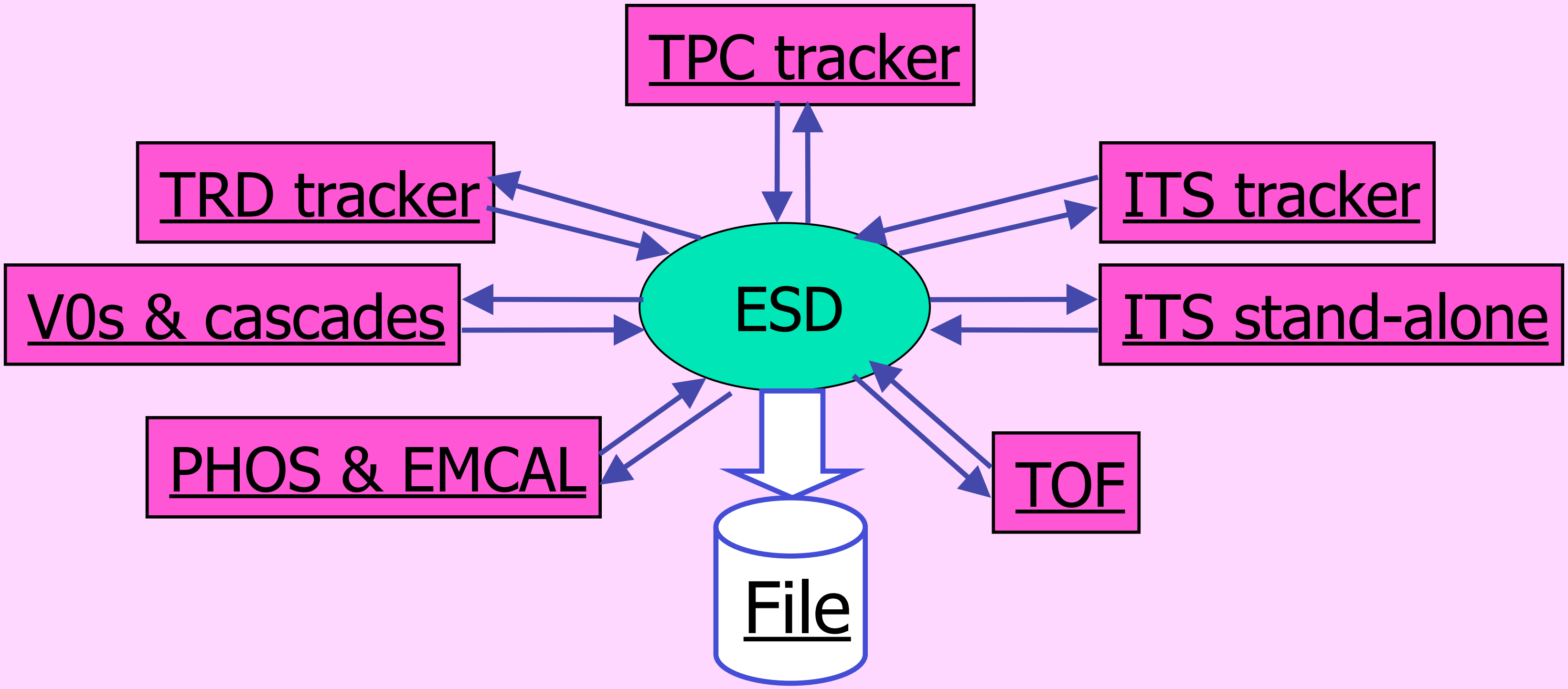
- Cluster finding in the detectors (centre of gravity)
  - Unfolding of overlapped clusters (optional)
- Primary vertex reconstruction using the ITS (SPD)
  - Pileup detection (optional)
- “Seeding” in TPC (with/out the vertex contraint)
  - Later, also the “seeding” in ITS and TRD (optional)
- Combined tracking
  - On-the-fly kink and V0 reconstruction (optional)
- Primary vertex using the tracks
- Secondary vertices using the tracks (V0s, cascades)

The three passes of the combined tracking



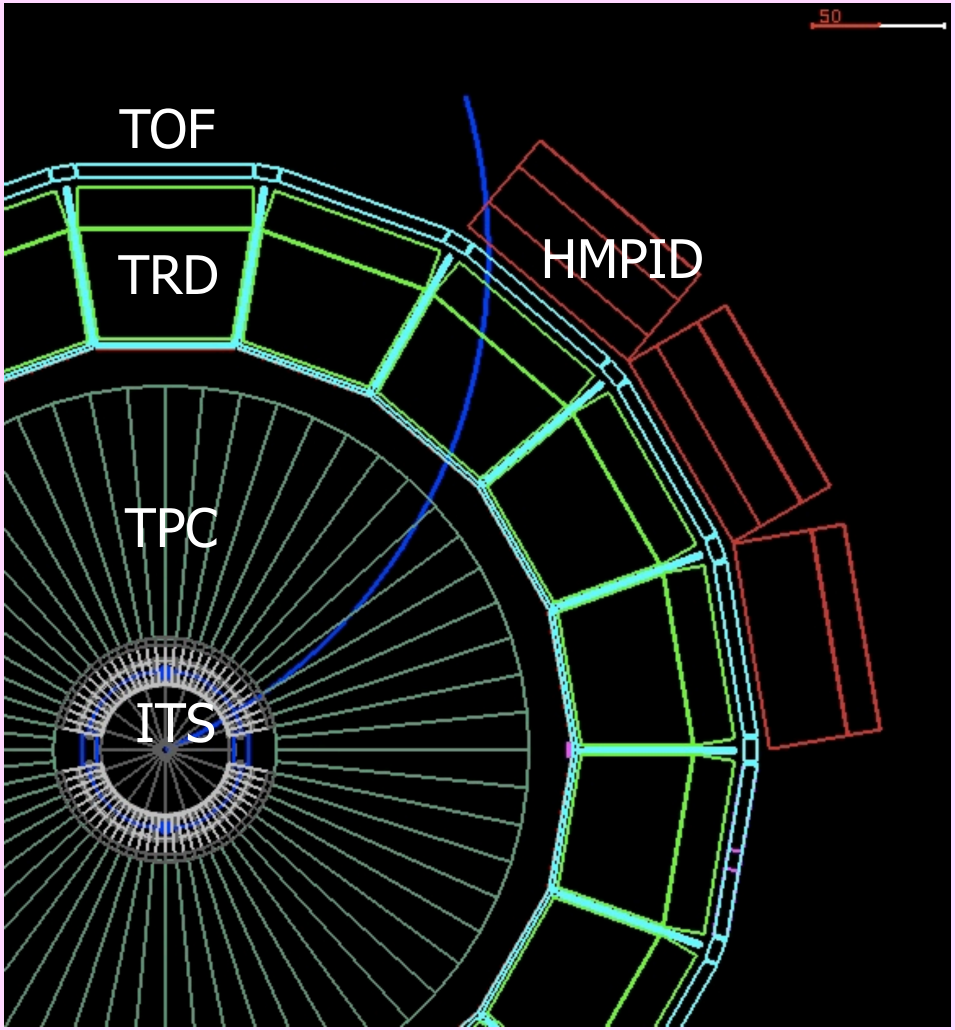
1. "Seeds" in TPC. Tracking from the outer to the inner wall of TPC. The same in ITS.
   - Track parameters are OK
   - PID is not yet OK
2. Tracking from the inner to outer layer of ITS. The same in TPC. The same in TRD. Matching with TOF, HMPID, PHOS/EMCAL
   - PID is OK
   - Track parameters are not OK
3. Tracking from the outer to inner TRD wall. The same in TPC. The same in ITS.
   - PID is OK
   - Track parameters are also OK

A bit of the software design…



- The general initialization and the processing sequence is defined by AliReconstruction.
- The process is configured/triggered by a special macro rec.C

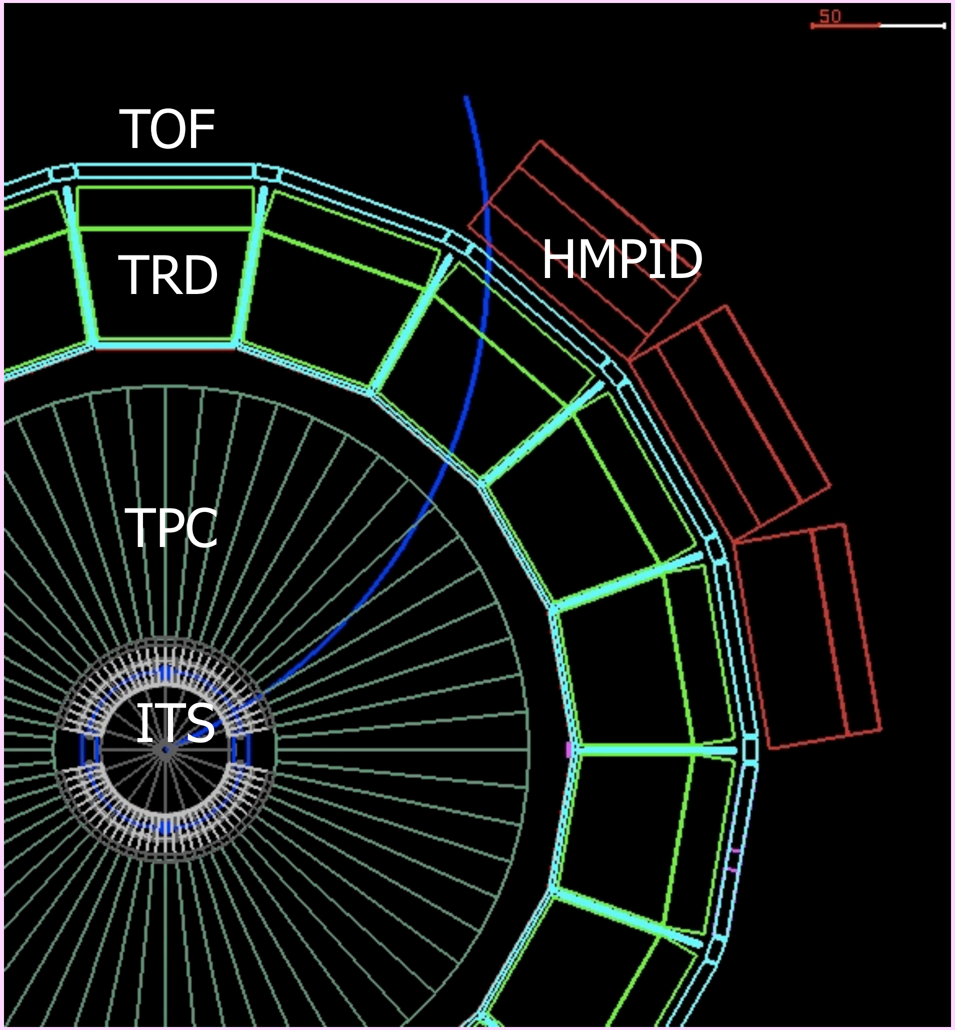
Charged PID with central ALICE



The same track can simultaneously be registered by 5 detectors that

- have quite different PID response,
- are efficient in complementary momentum ranges.

Charged PID with central ALICE



The same track can simultaneously be registered by 5 detectors that

- have quite different PID response,
- are efficient in complementary momentum ranges.

- The PID procedure cannot be Kalman-like (PDFs are not Gaussian).
- It cannot be Likelihood-like either (the priors cannot be fixed once and forever)…

Bayesian approach in PID

Probability to be a particle of **i**-type (**i = e,** μ, π, **K, p,** …),
if the PID signal in the detector is **s**:

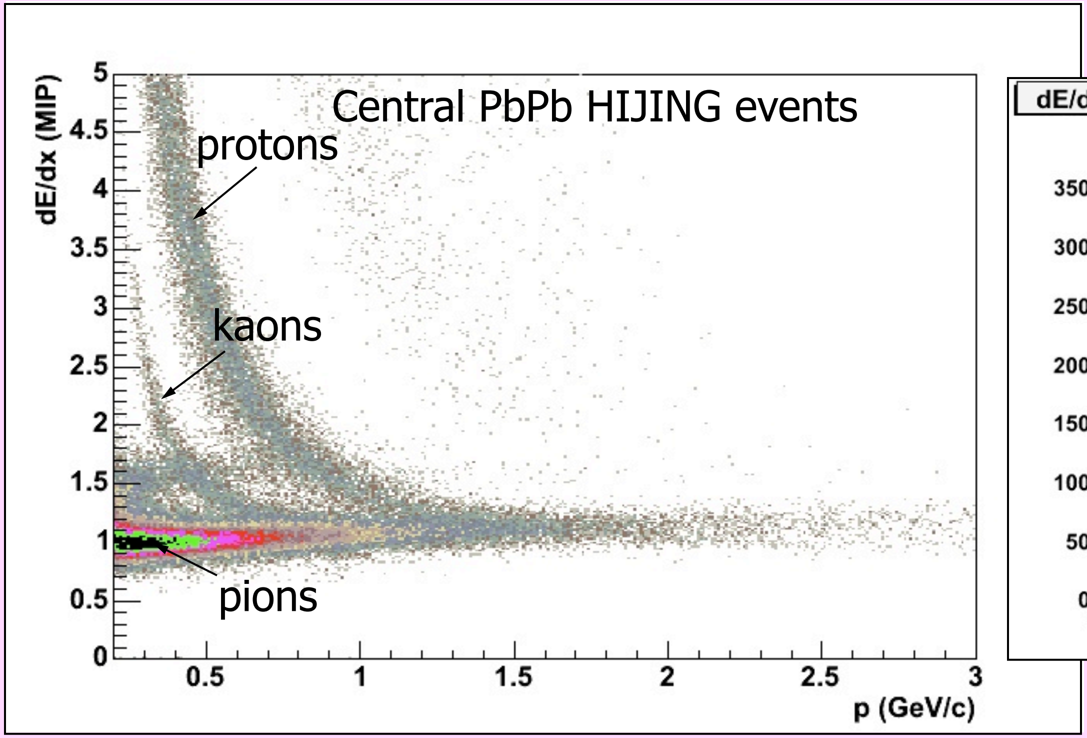
$$\boxed{w(i\,|\,s) = \frac{C_i r(s|i)}{\sum\limits_{k=e,\mu,\pi,\ldots} C_k r(s\,|\,k)}}$$

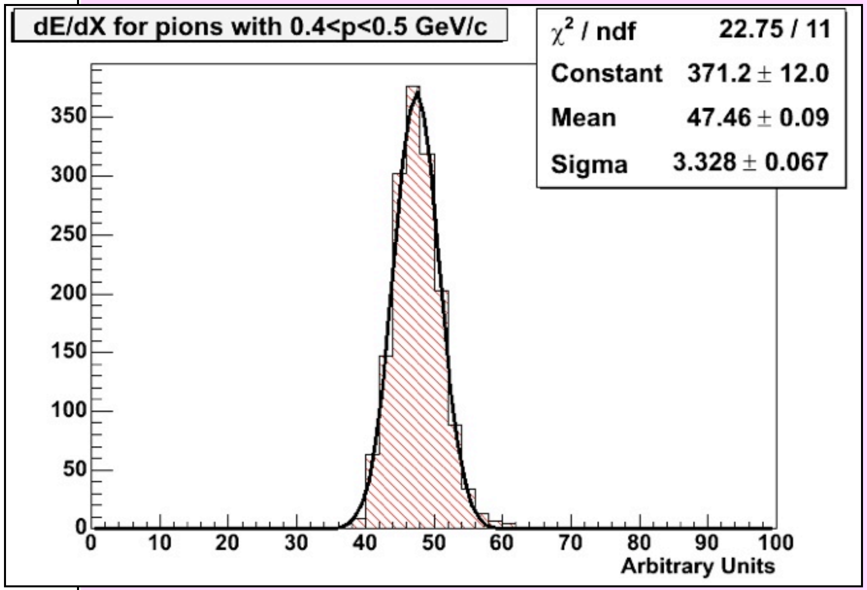
- **C_i** - a priori probabilities to be a particle of the **i**-type.
  "Particle concentrations", that depend on the event and track selection.
- **r(s|i)** – conditional probability density functions to get the signal **s**, if a particle of **i**-type hits the detector.
  "Detector response functions",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

In the case of N contributing detectors: $r(s_1,\ldots,s_N\,|\,i) \sim \prod_{d=1}^{N} r(s_d\,|\,i)$

Obtaining the conditional PDFs

Example: "TPC response function"





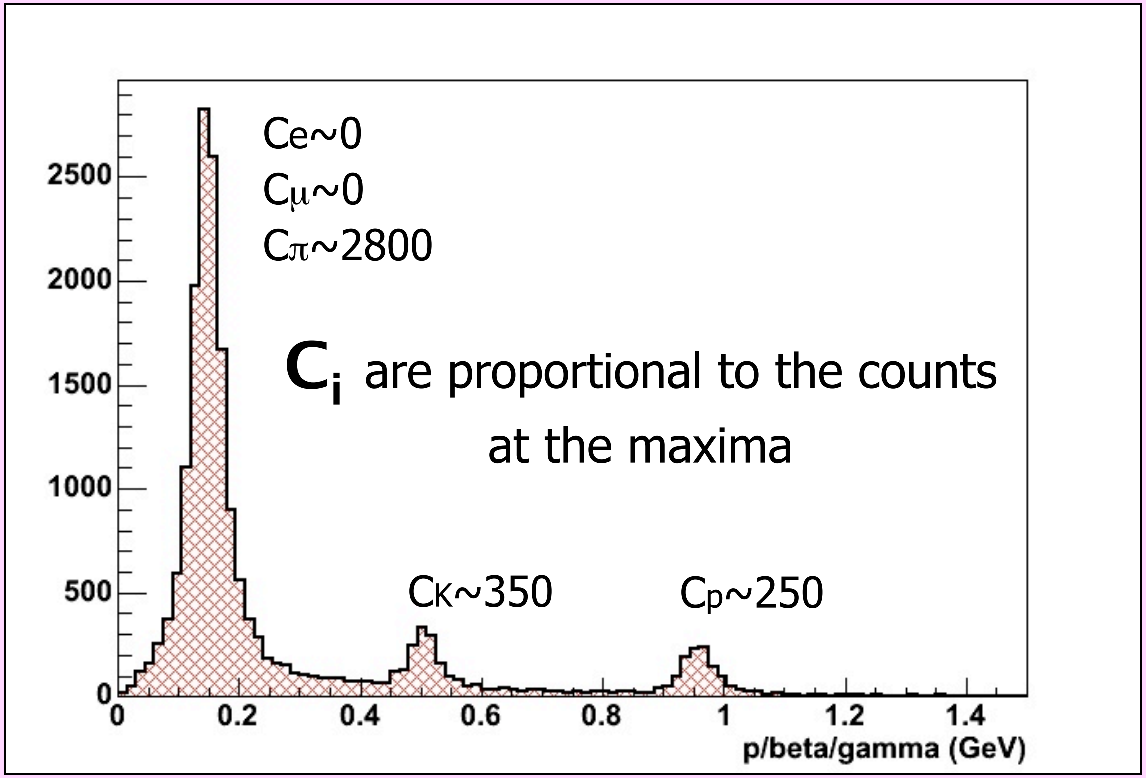
For each momentum **p** the function **r(s|i)** is a Gaussian with

- centroid **<dE/dx>** given by the Bethe-Bloch formula and
- sigma σ = **0.08<dE/dx>**

This is a property of the detector (TPC). Can be prepared in advance !

Obtaining the a priori probabilities

("particle concentrations")

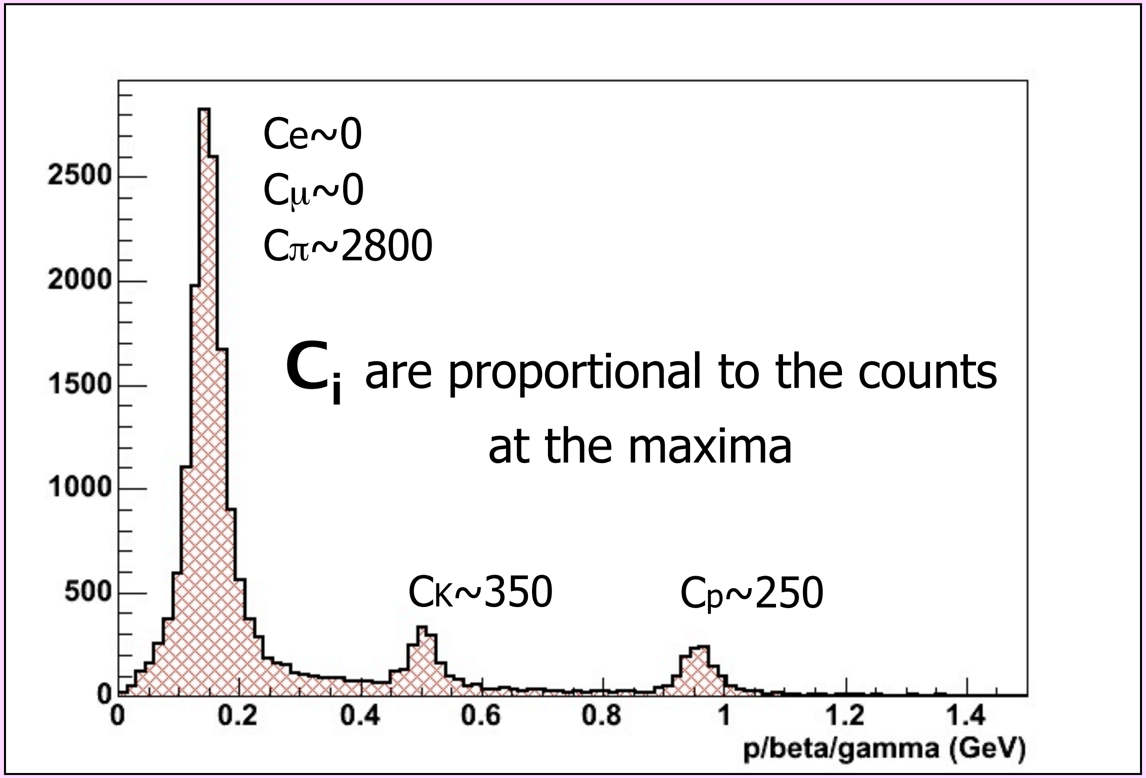


β by TOF, p by TPC

1. Sometimes, we may know the priors (V0s, cascades)
2. Sometimes, we can get the priors by iterating over the data
3. Anytime, we can use the raw PID signals
   - Simple histograming
   - Complicated fits…

Obtaining the a priori probabilities

("particle concentrations")



β by TOF, p by TPC

1. Sometimes, we may know the priors (V0s, cascades)
2. Sometimes, we can get the priors by iterating over the data
3. Anytime, we can use the raw PID signals
   - Simple histograming
   - Complicated fits…

The "particle concentrations" depend on the event and track selection.
They cannot be prepared once and for all kinds of analysis !

Fine, but what is the advantage ?

Fine, but what is the advantage ?

- The raw PID signals do not have to be stored for all the tracks (big gain in disk space for AOD)

Fine, but what is the advantage ?

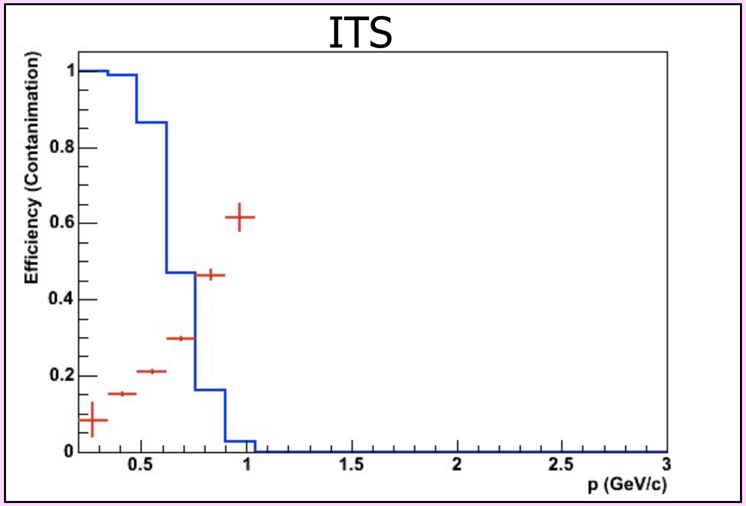
- The raw PID signals do not have to be stored for all the tracks (big gain in disk space for AOD)
- Much less parameters to fit at the analysis level
  - In fact, the “amplitudes” only. Because the “centroids and sigmas” are already given by the response functions (precalcualted by reconstruction).

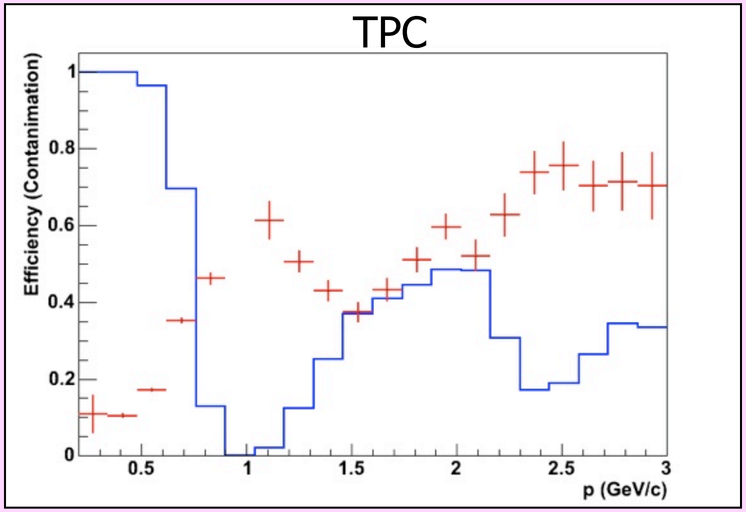
The three parts of the PID procedure

- Calibration part, belongs to the calibration software.
  Obtaining the single detector response functions.
  Done by detector experts.
- "Constant part", belongs to the reconstruction software.
  Calculating (for each track) the values of detector response functions, combining them and writing the result to the Event Summary Data.
  Done automatically, in massive reconstruction runs on the Grid.
- "Variable part", belongs to 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.
  Done by physicists involved in this particular analysis.

Example: Kaon identification with ITS,TPC and TOF

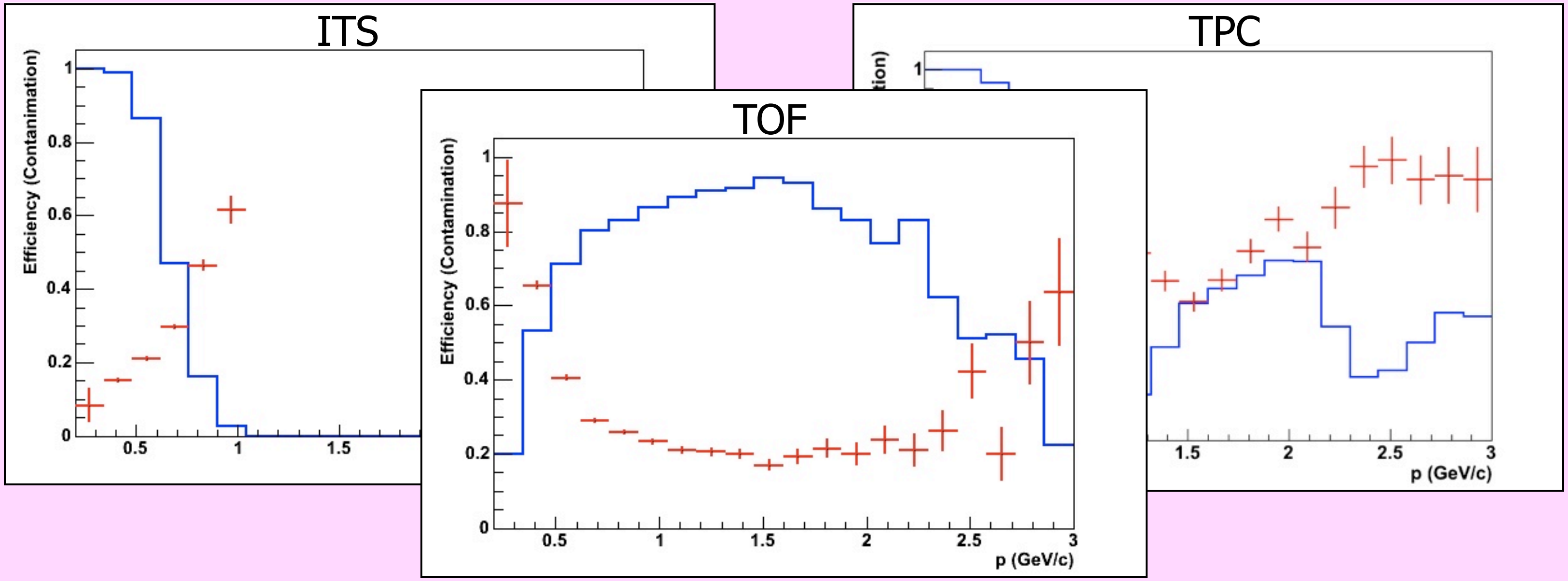
(central PbPb HIJING events)





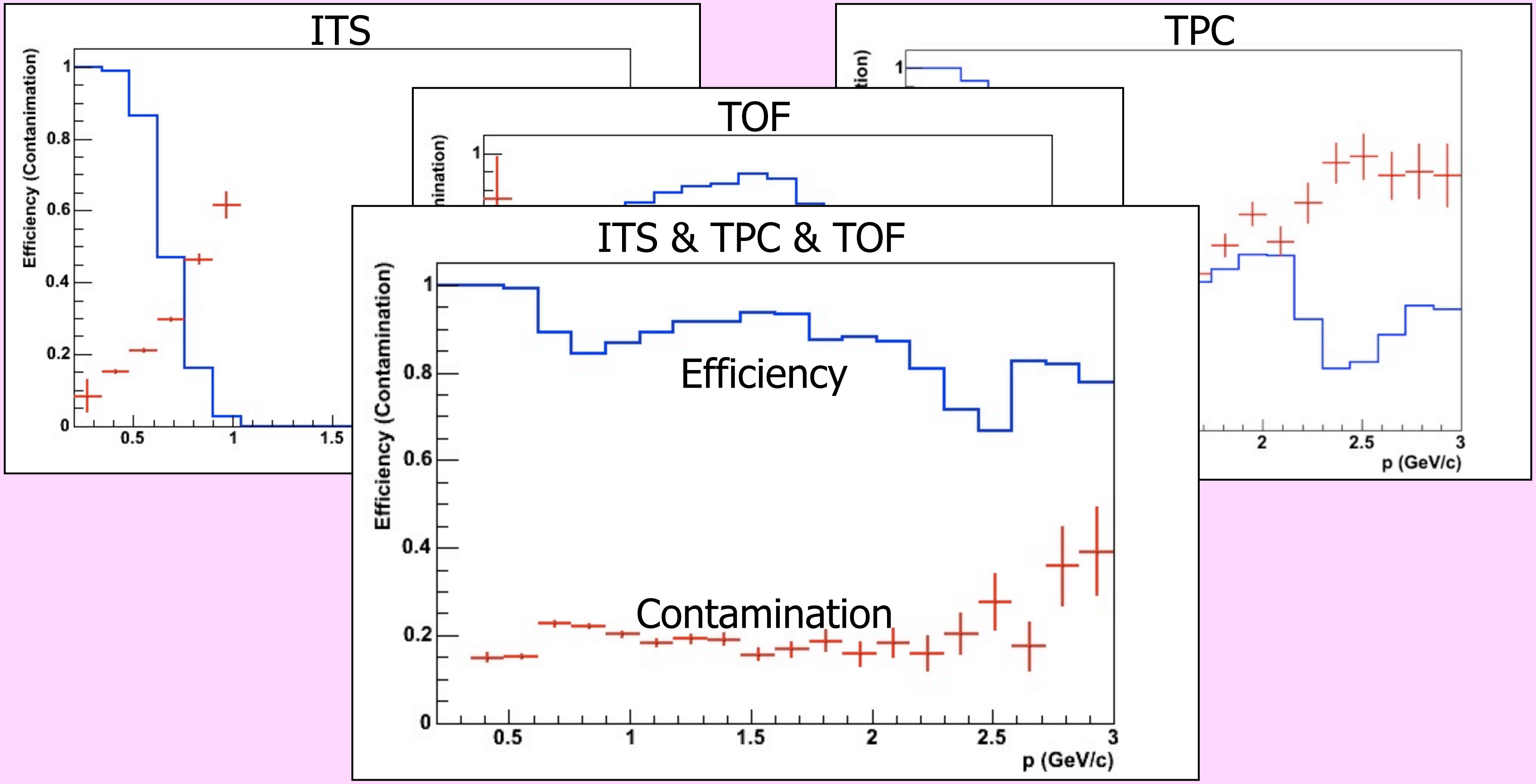
Example: Kaon identification with ITS,TPC and TOF

(central PbPb HIJING events)



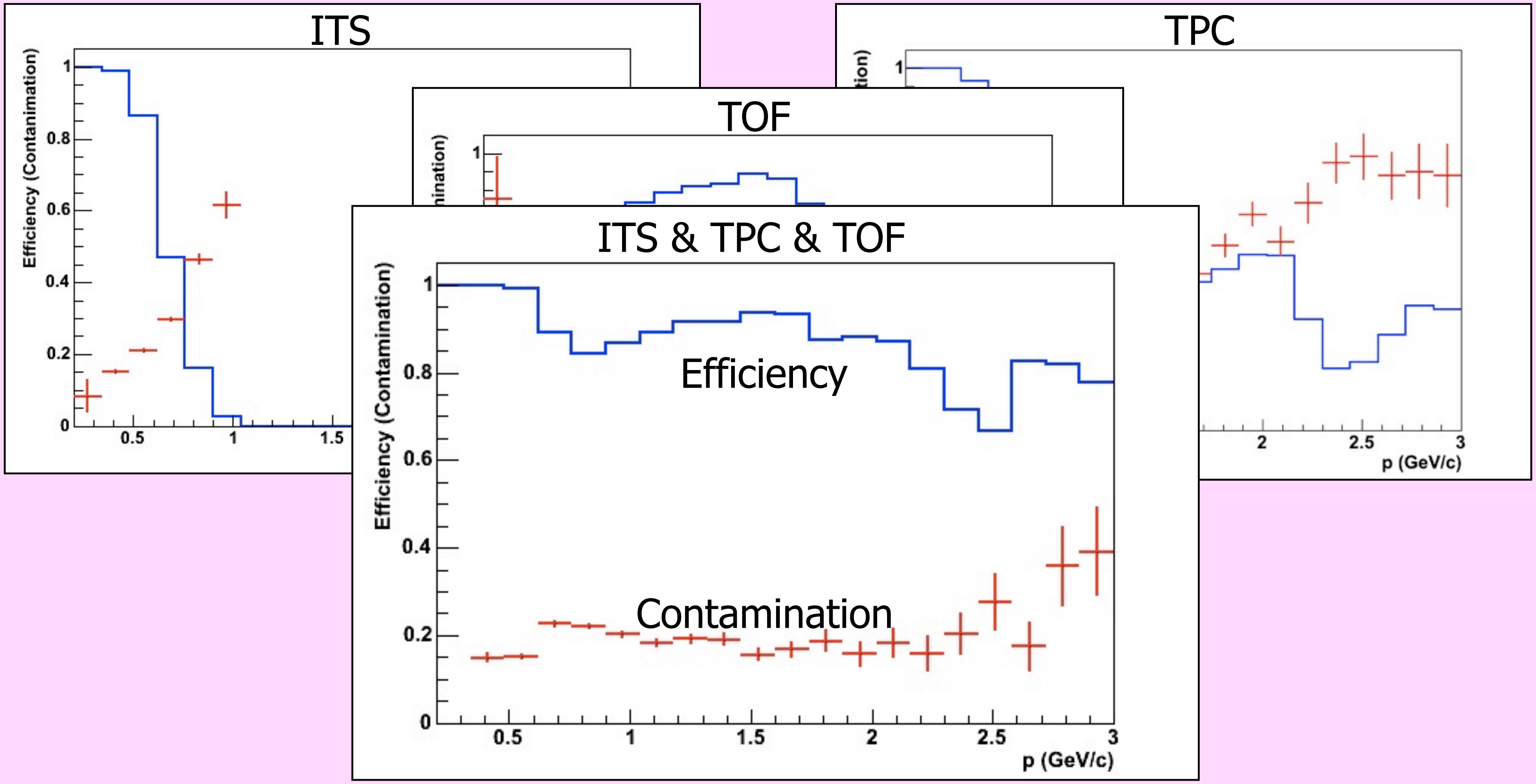
Example: Kaon identification with ITS,TPC and TOF

(central PbPb HIJING events)



Example: Kaon identification with ITS,TPC and 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.

A complementary approach: n-sigma cuts

A complementary approach: n-sigma cuts

- Guarantees a definite and constant over the momentum efficiency. Does not deal with the priors !
  - Does not tell anything about the contamination
  - Does not maximize the significance

A complementary approach: n-sigma cuts

- Guarantees a definite and constant over the momentum efficiency. Does not deal with the priors !
  - Does not tell anything about the contamination
  - Does not maximize the significance
- Everything that concerns the response functions is the same as for the Bayesian
  - A big piece of software can (and must) be shared by the two approaches.

Conclusions

Conclusions

- Reconstruction and PID in ALICE are challenging and quite interesting themselves.

Conclusions

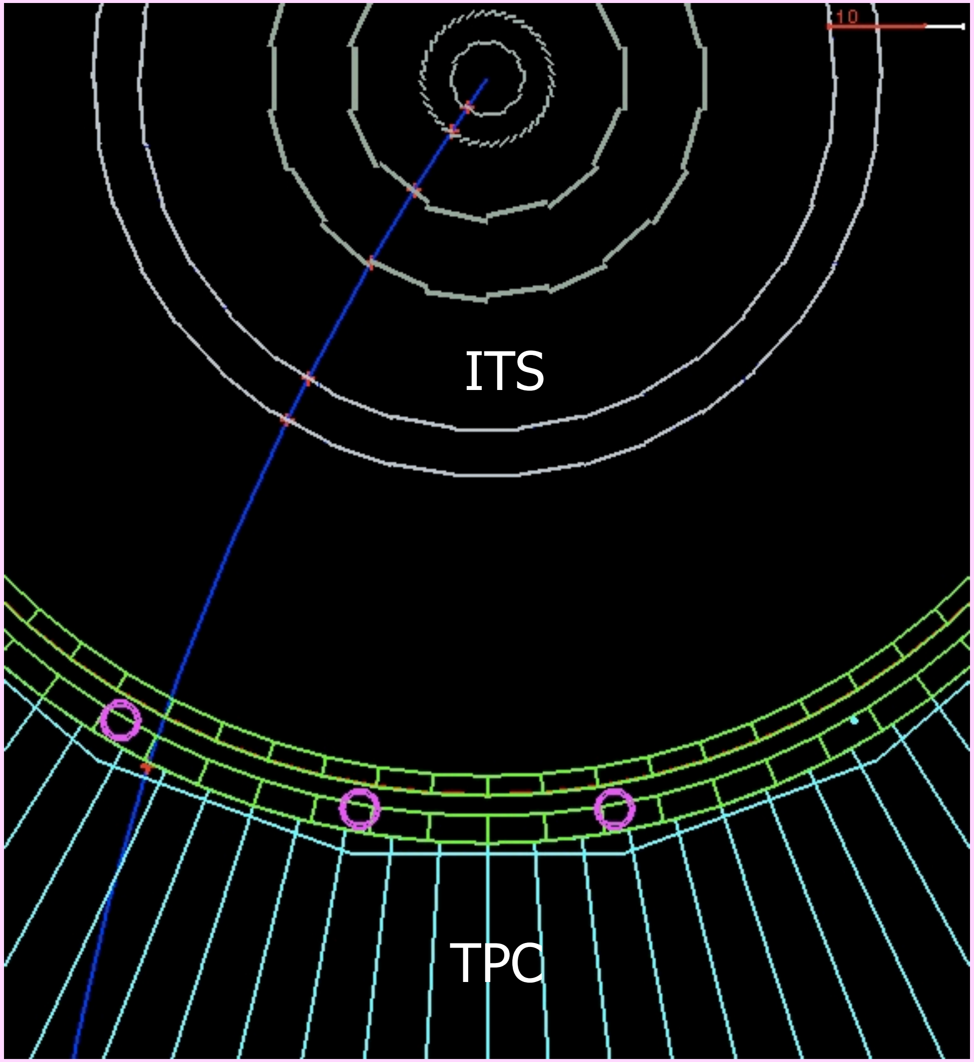
- Reconstruction and PID in ALICE are challenging and quite interesting themselves.
- However, ALICE physics is even more interesting.

Conclusions

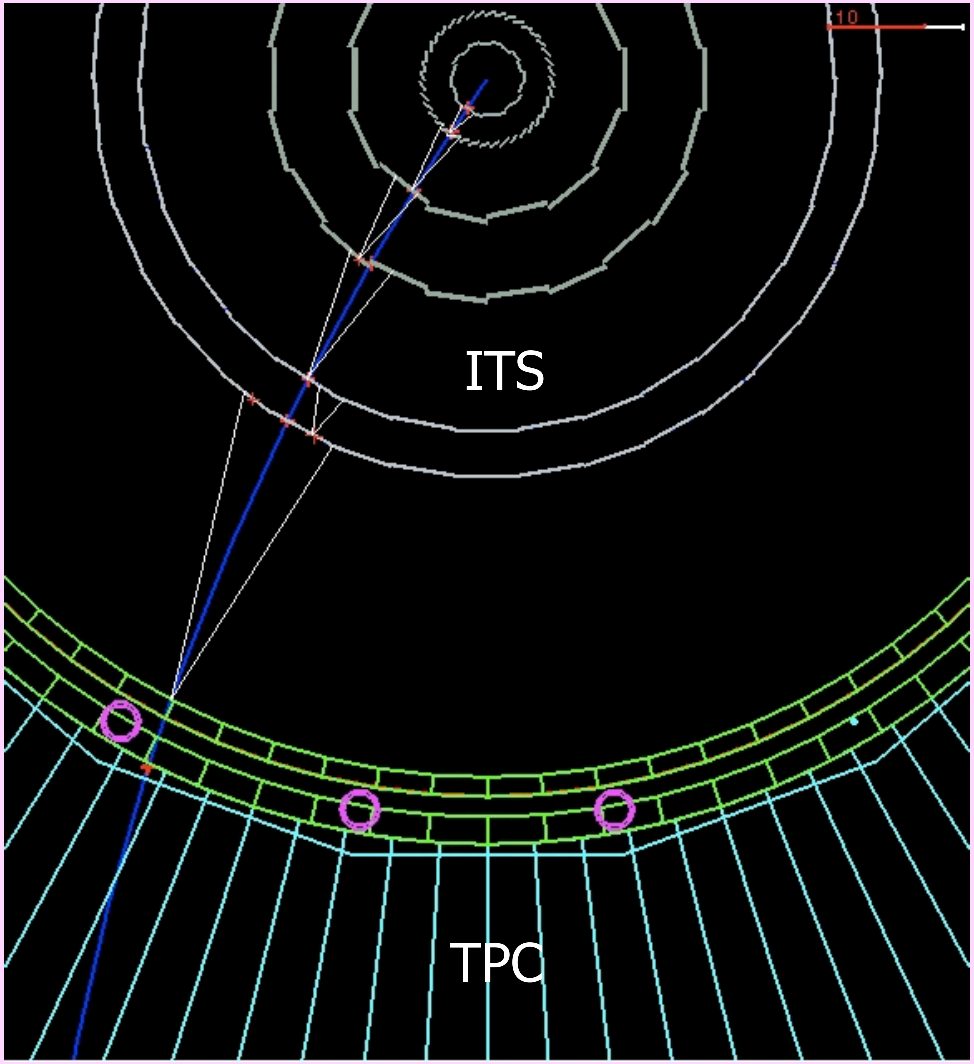
- Reconstruction and PID in ALICE are challenging and quite interesting themselves.
- However, ALICE physics is even more interesting.
- <u>It's about the time to start doing physics ! ☺</u>

Appendix: A few unclear questions…

Statistical problems with track finding in ITS

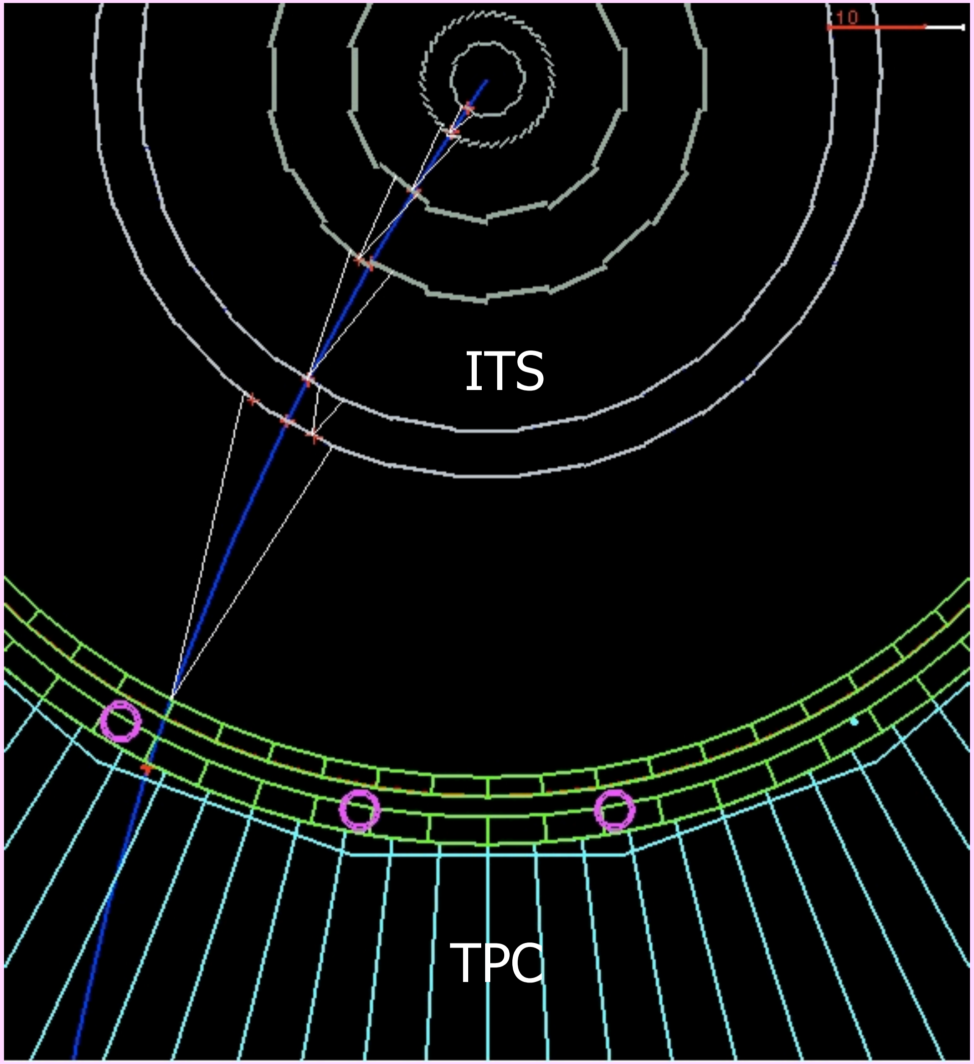


Statistical problems with track finding in ITS



Several clusters within the "road" defined by multiple scattering...

Statistical problems with track finding in ITS

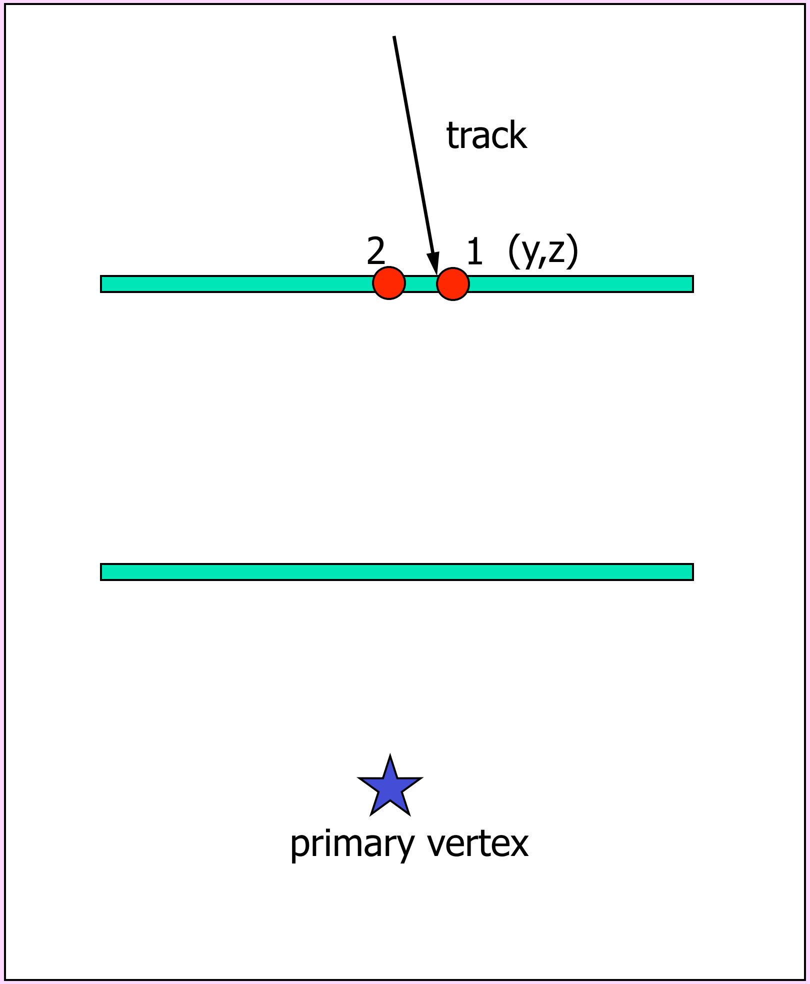


Several clusters within the "road" defined by multiple scattering…

Suggested solution:

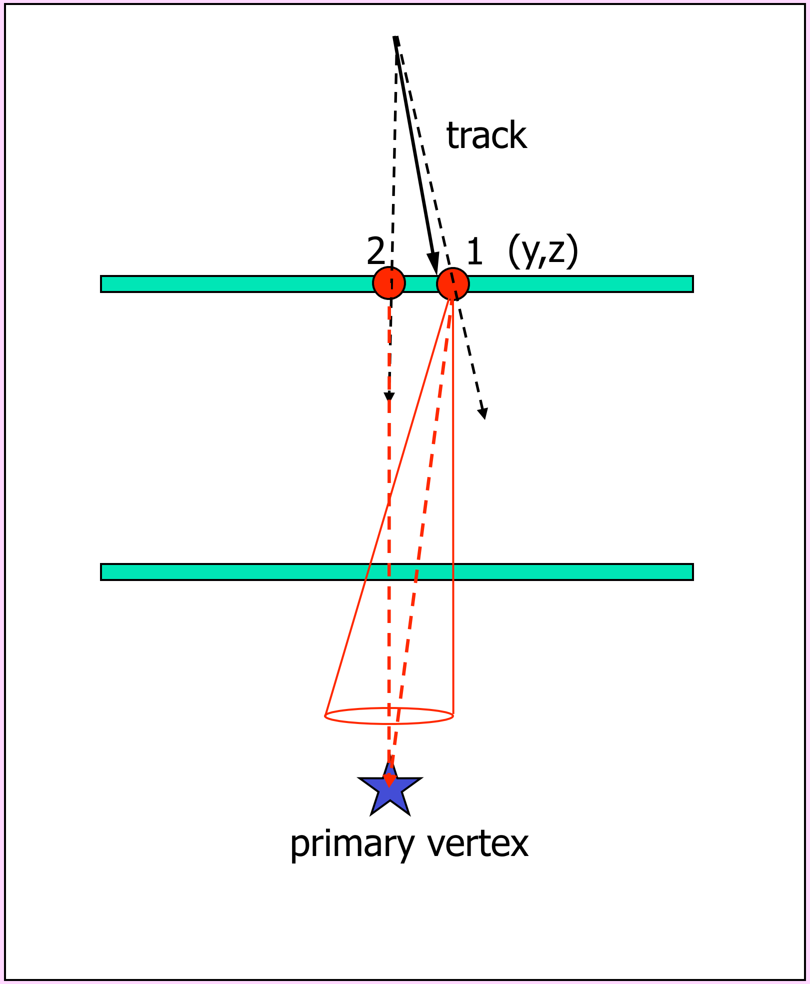
- Investigation of the whole tree of possible prolongations.
- Applying a "vertex constraint" (1^{st} pass).

Ad-hoc “vertex constraint”



Looking at the cluster position only, the cluster #1 is “better”.

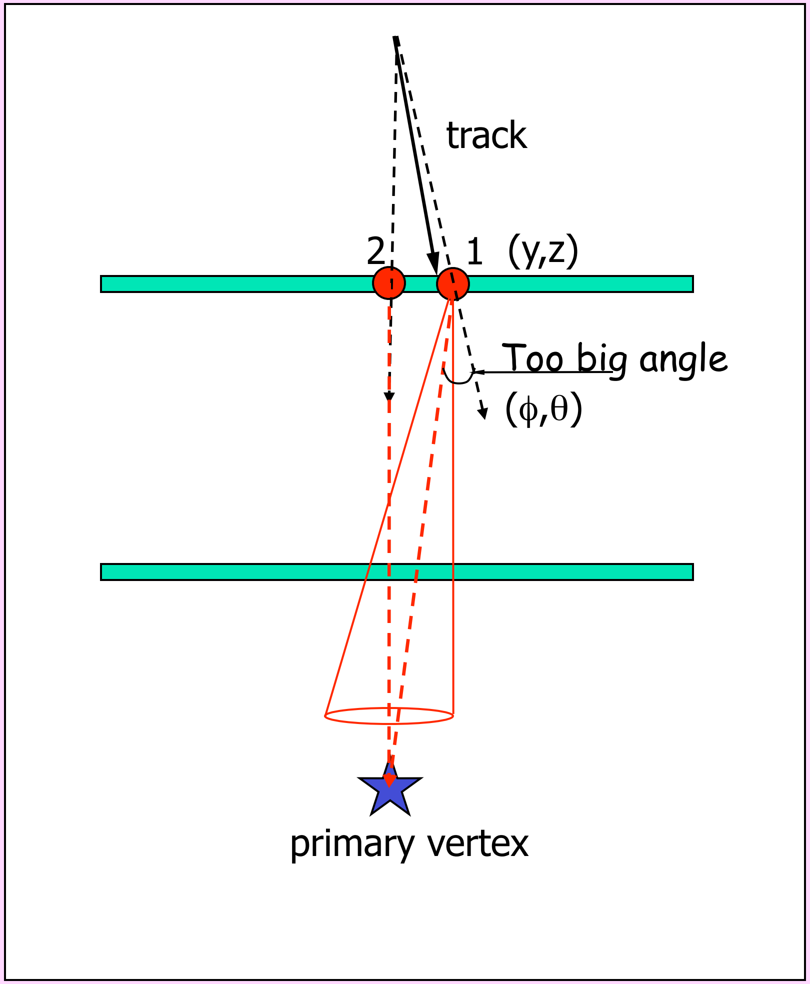
Ad-hoc “vertex constraint”



Looking at the cluster position only, the cluster #1 is “better”.

But, if also taking into account the direction towards the primary vertex, the cluster #2 becomes more preferable…

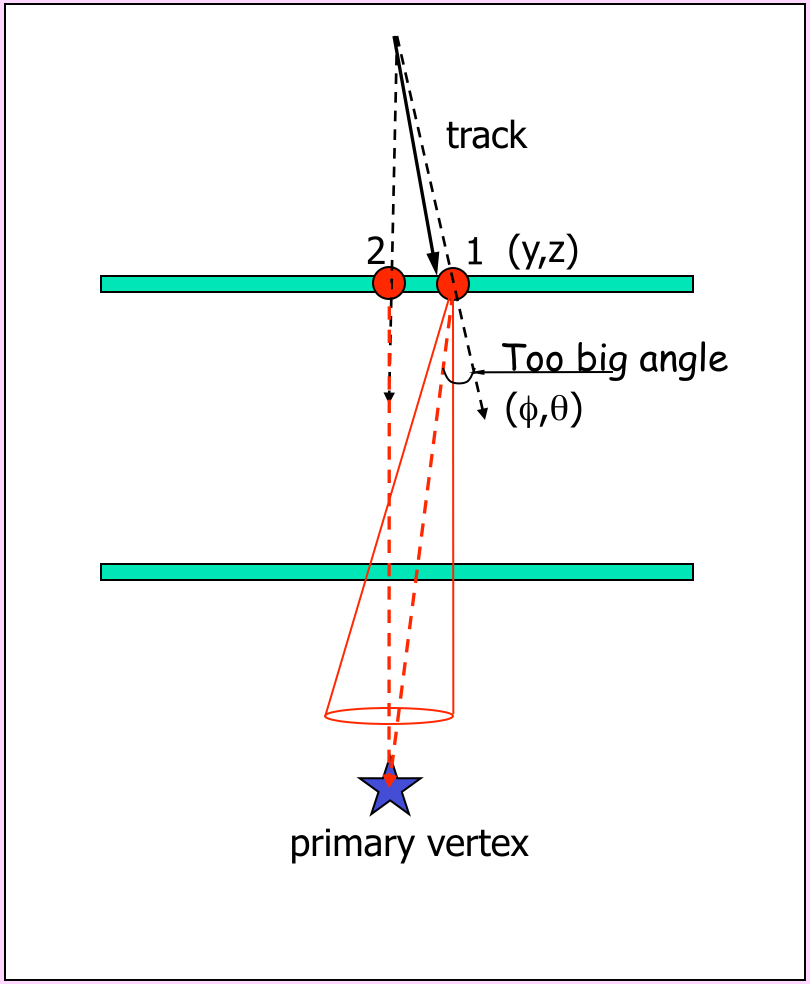
Ad-hoc “vertex constraint”



Looking at the cluster position only, the cluster #1 is “better”.

But, if also taking into account the direction towards the primary vertex, the cluster #2 becomes more preferable…

Ad-hoc “vertex constraint”

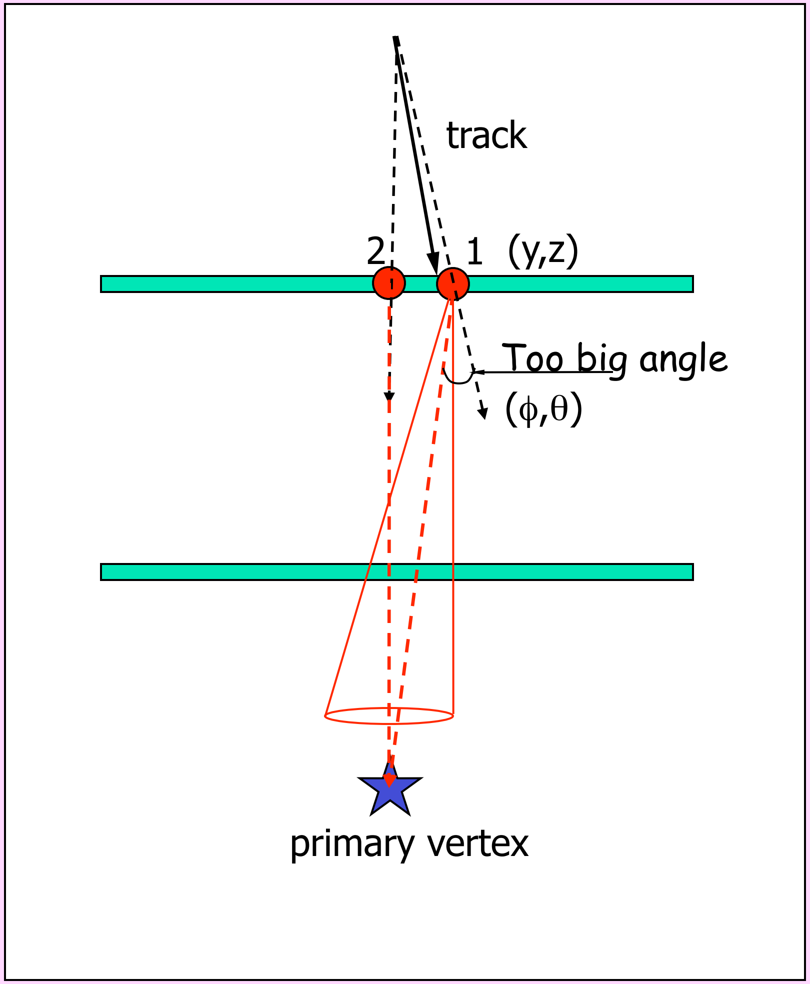


Looking at the cluster position only, the cluster #1 is “better”.

But, if also taking into account the direction towards the primary vertex, the cluster #2 becomes more preferable…

Technically, this is done by extending the “measurement” (y,z) -> (y,z,ϕ,θ)

Ad-hoc "vertex constraint"



Looking at the cluster position only, the cluster #1 is "better".

But, if also taking into account the direction towards the primary vertex, the cluster #2 becomes more preferable…

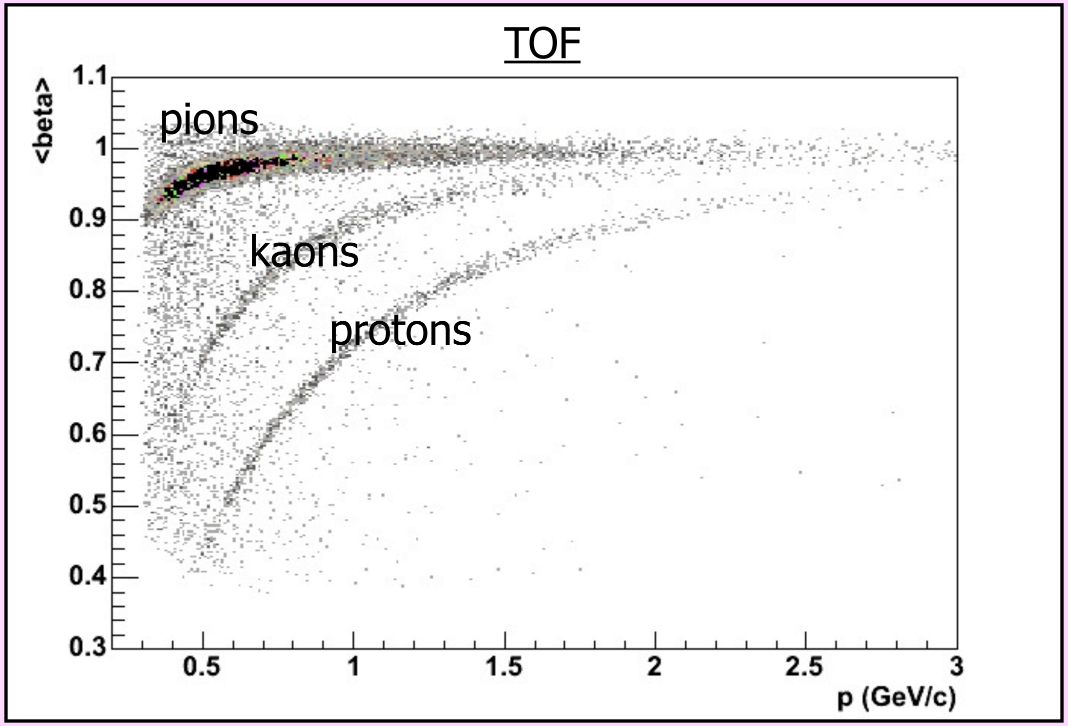
Technically, this is done by extending the "measurement" (y,z) -> (y,z,ϕ,θ)

A question:

Can all this be justified ? Improved ?

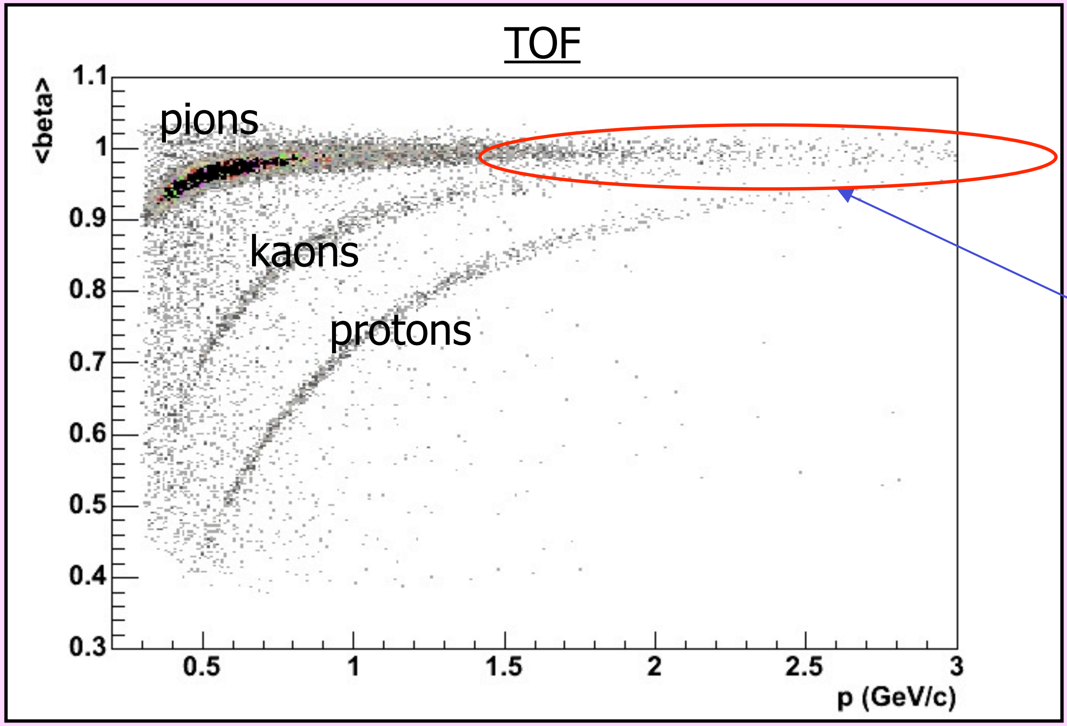
(especially, if applied the same track repeatedly ?)

The high-momentum limit for PID



The Bayesian calculations nicely glue together the momentum sub-ranges, but, as the momentum goes up, the “separation power” vanishes, and…

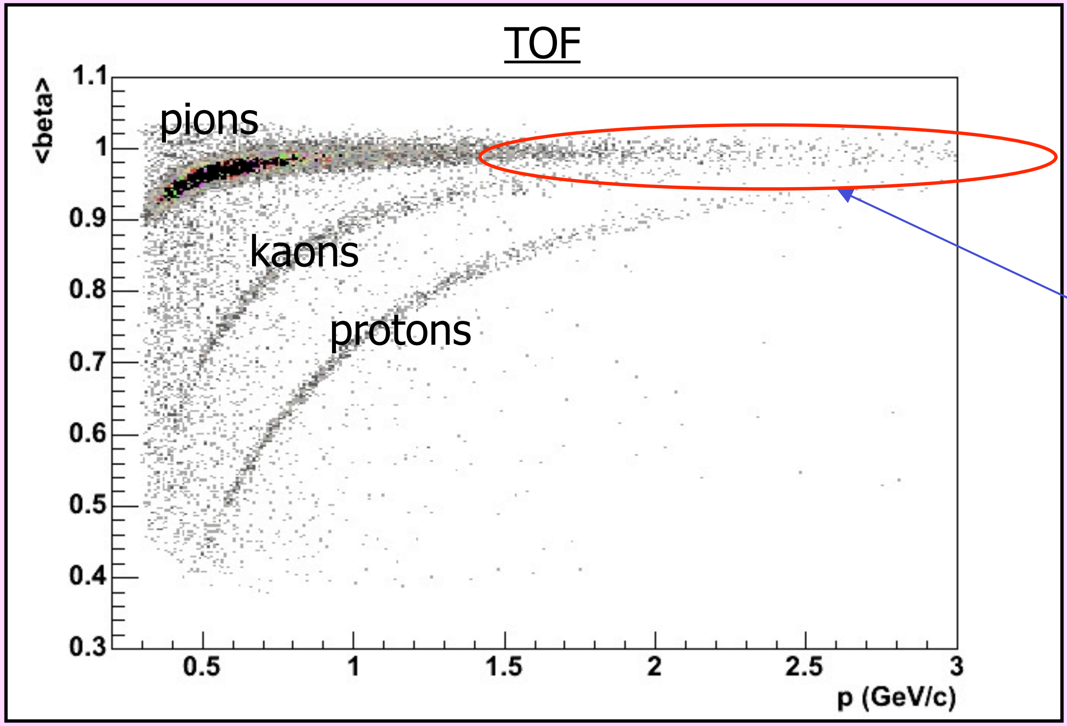
The high-momentum limit for PID



The Bayesian calculations nicely glue together the momentum sub-ranges, but, as the momentum goes up, the "separation power" vanishes, and…

We are left with the bare priors ☹

The high-momentum limit for PID



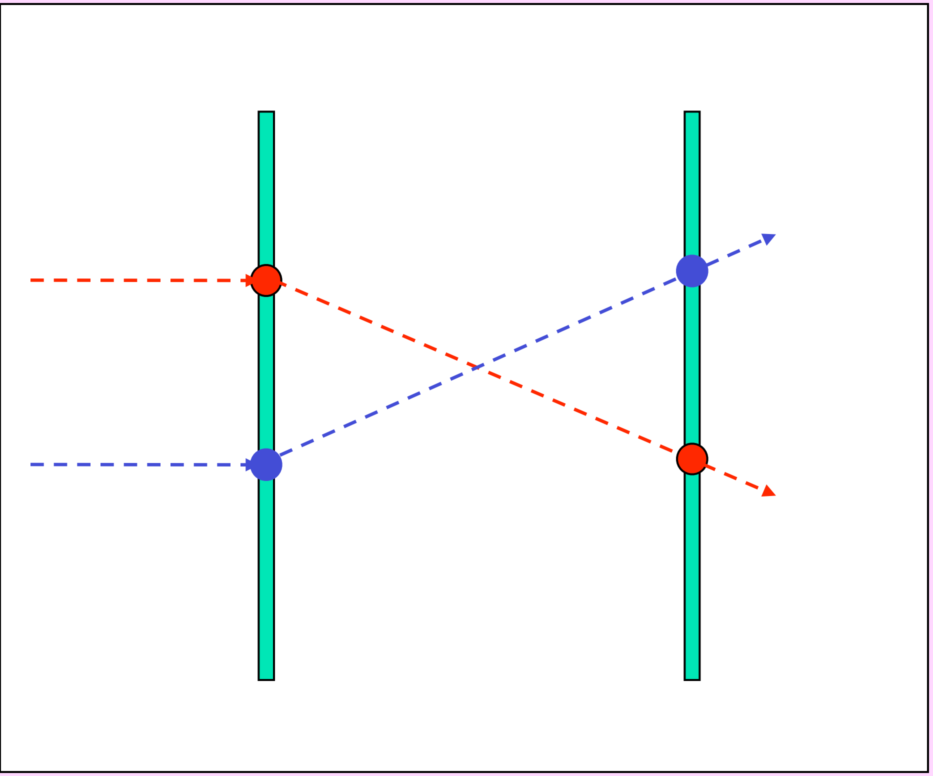
The Bayesian calculations nicely glue together the momentum sub-ranges, but, as the momentum goes up, the “separation power” vanishes, and…

We are left with the bare priors ☹

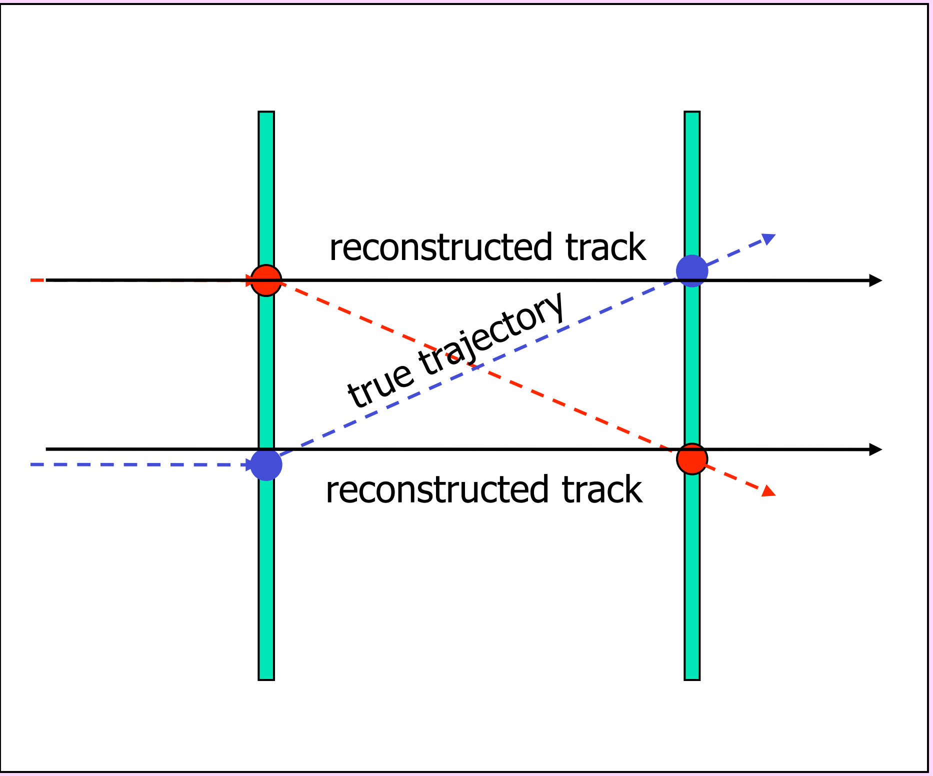
Questions:

- The influence of the priors on the final result: Can it be somehow quantified ?
- In any approach: at what momentum should we stop doing the PID ?

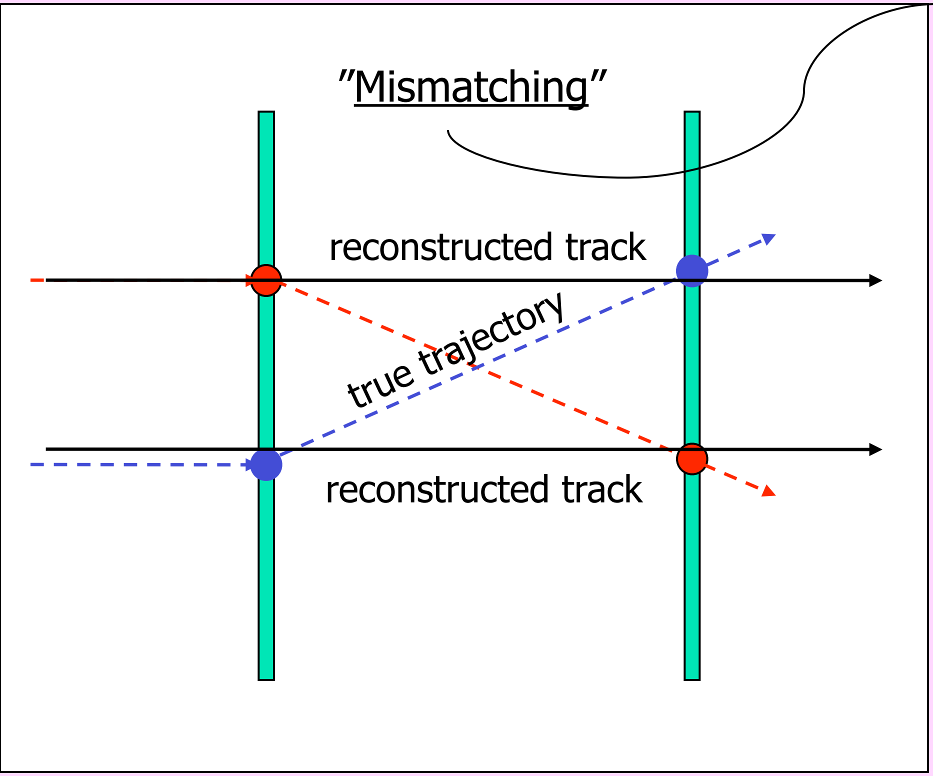
The problem of track mismatching

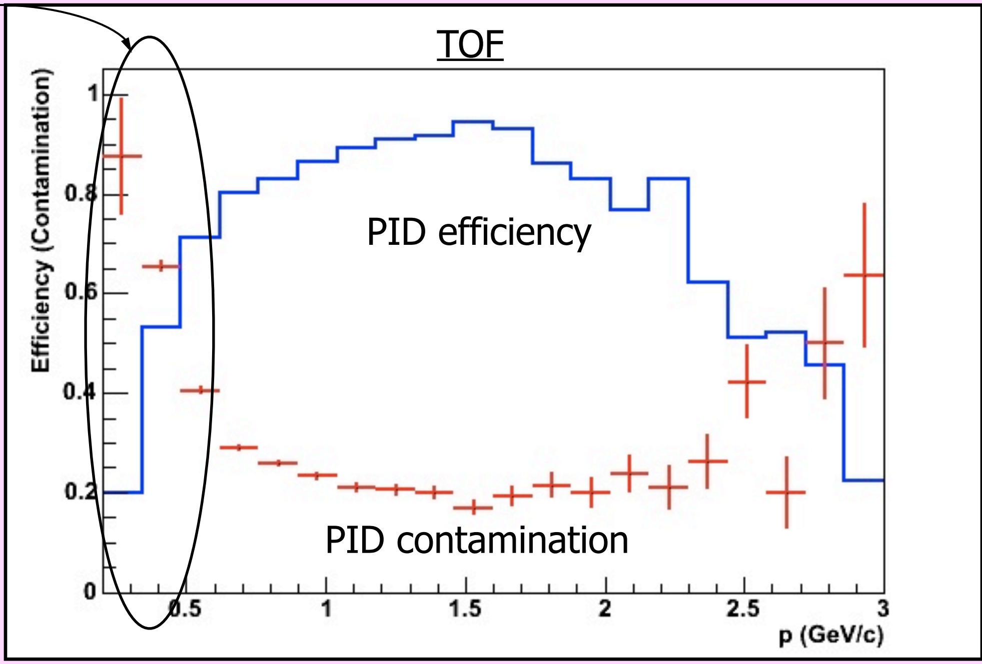


The problem of track mismatching



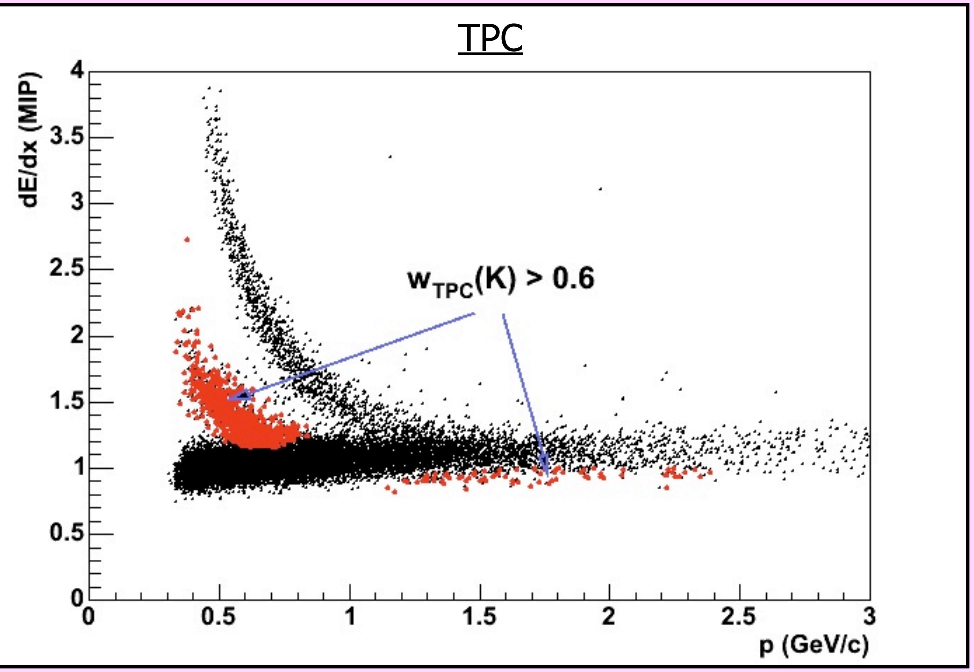
The problem of track mismatching

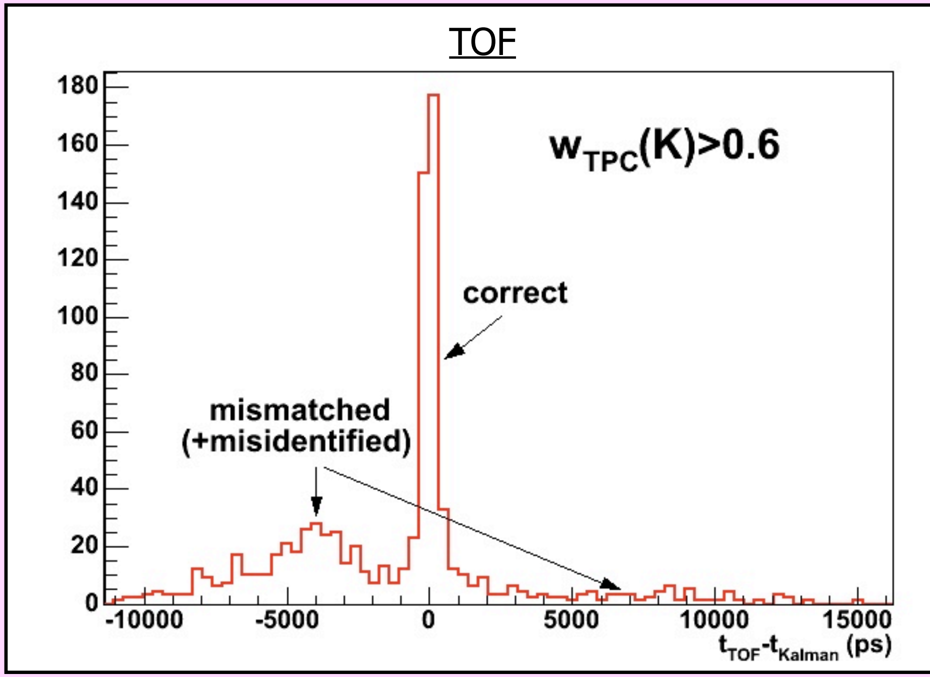




The track mismatching biases the combining (any kind of !) the PID information, because the main assumption that all the detectors register the same particle, is not satisfied…

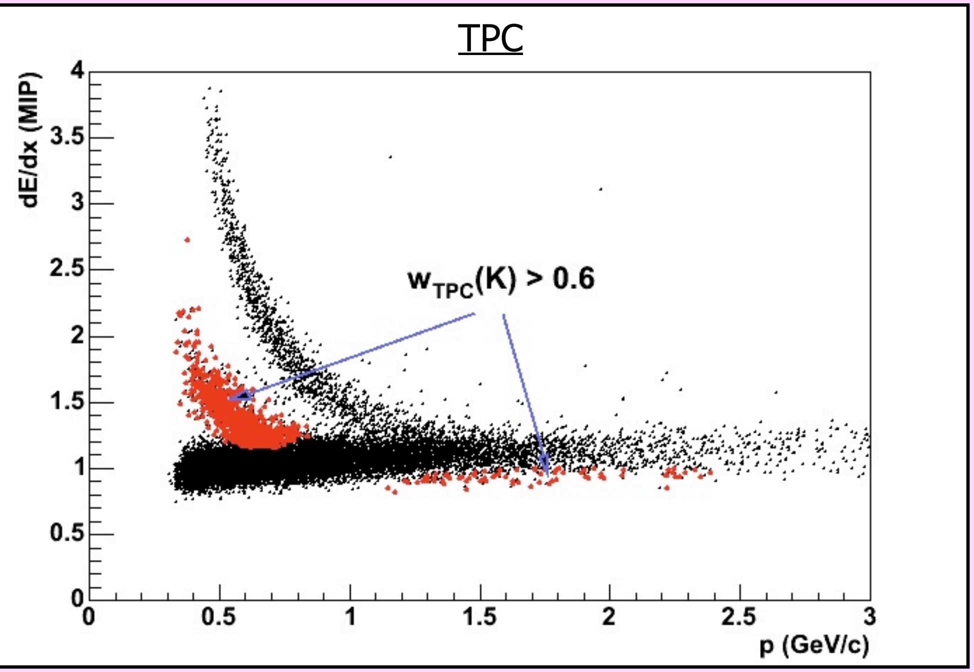
Ad-hoc treatment for the mismatching in TOF

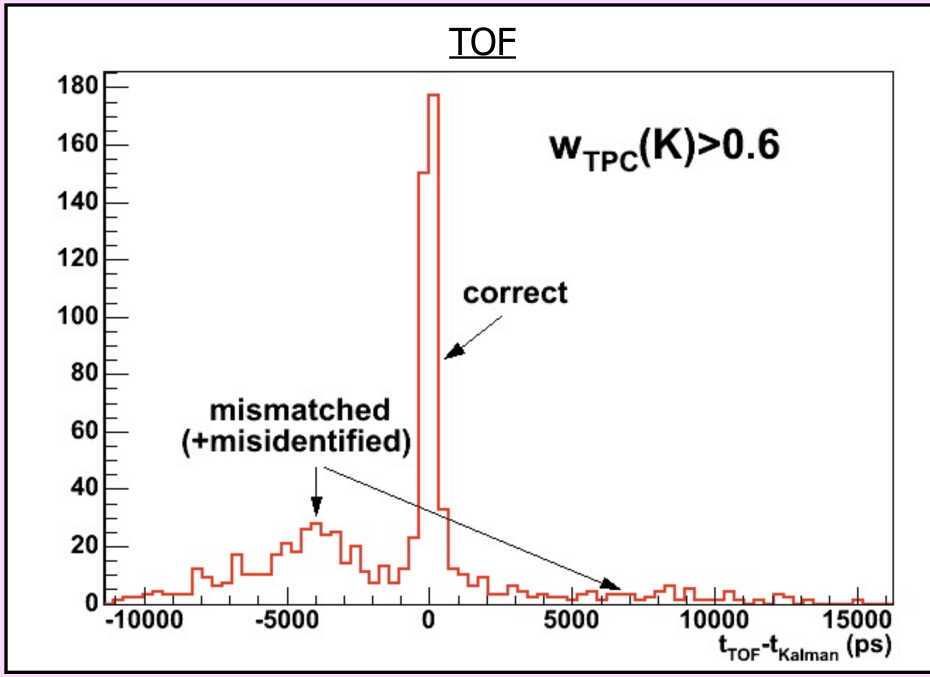




Observing in one of the detectors the distribution of signals for a clean sample of particles pre-selected in other detectors, we can get the range of signals, where the probability of mismatching is "high" → Veto in the combining…

Ad-hoc treatment for the mismatching in TOF

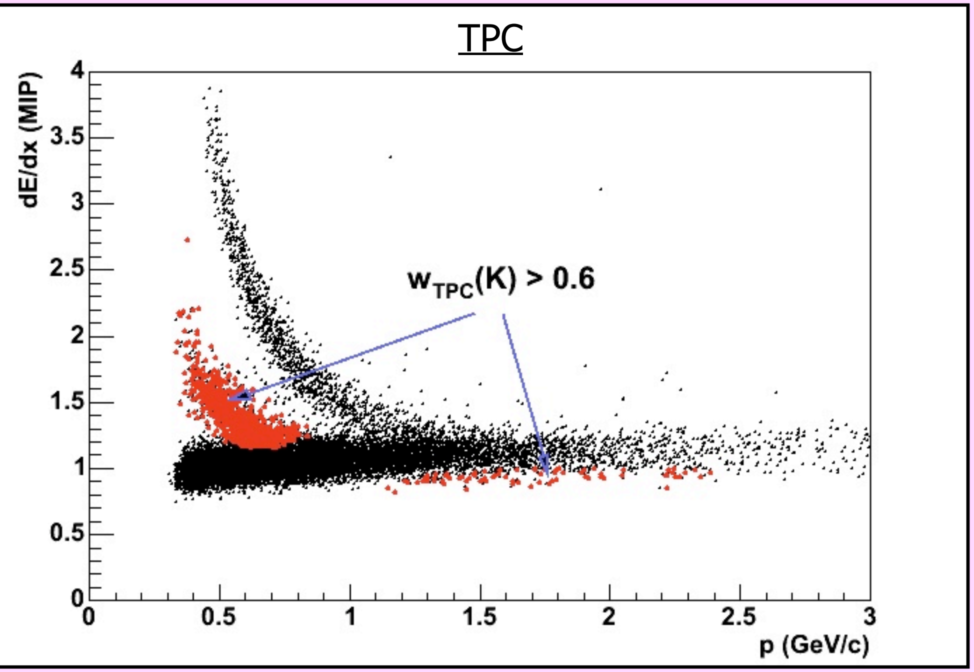


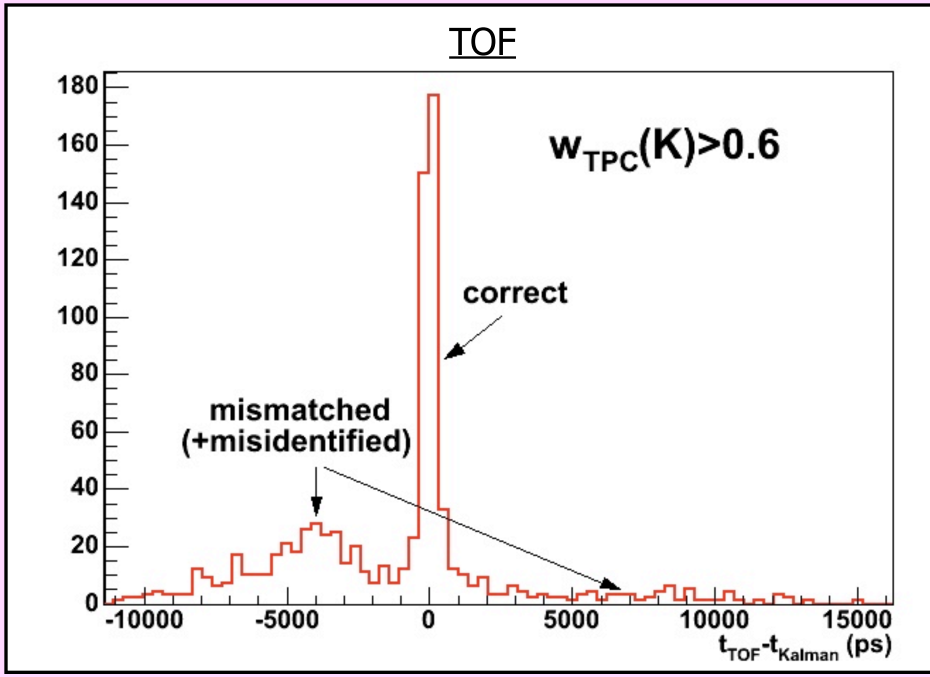


Observing in one of the detectors the distribution of signals for a clean sample of particles pre-selected in other detectors, we can get the range of signals, where the probability of mismatching is "high" → Veto in the combining…

A question: Can it be somehow generalized ? Made "smooth" ? Optimized ?

Ad-hoc treatment for the mismatching in TOF





Observing in one of the detectors the distribution of signals for a clean sample of particles pre-selected in other detectors, we can get the range of signals, where the probability of mismatching is "high" → Veto in the combining…

A question: Can it be somehow generalized ? Made "smooth" ? Optimized ?

Something like $w = (1-p_{12})w_1 + p_{12}w_{12}$ (p_{12} - prob. of a correct matching) ?