Istituto Nazionale di Fisica Nucleare
SEZIONE DI TORINO

From PID detectors to PID variables

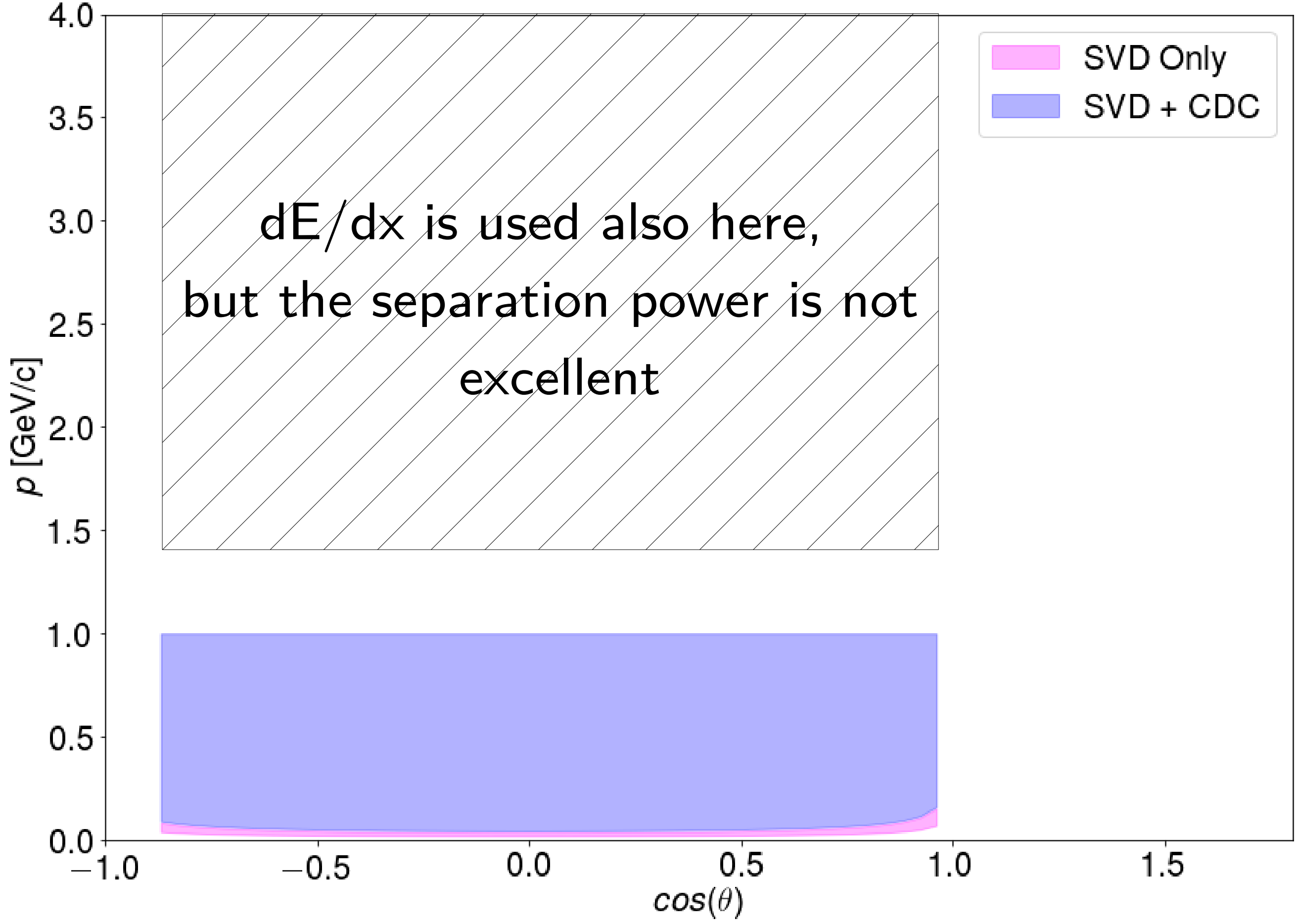
Umberto Tamponi
tamponi@to.infn.it

INFN – Sezione di Torino

Belle II starter kit
KEK, February 1st 2020

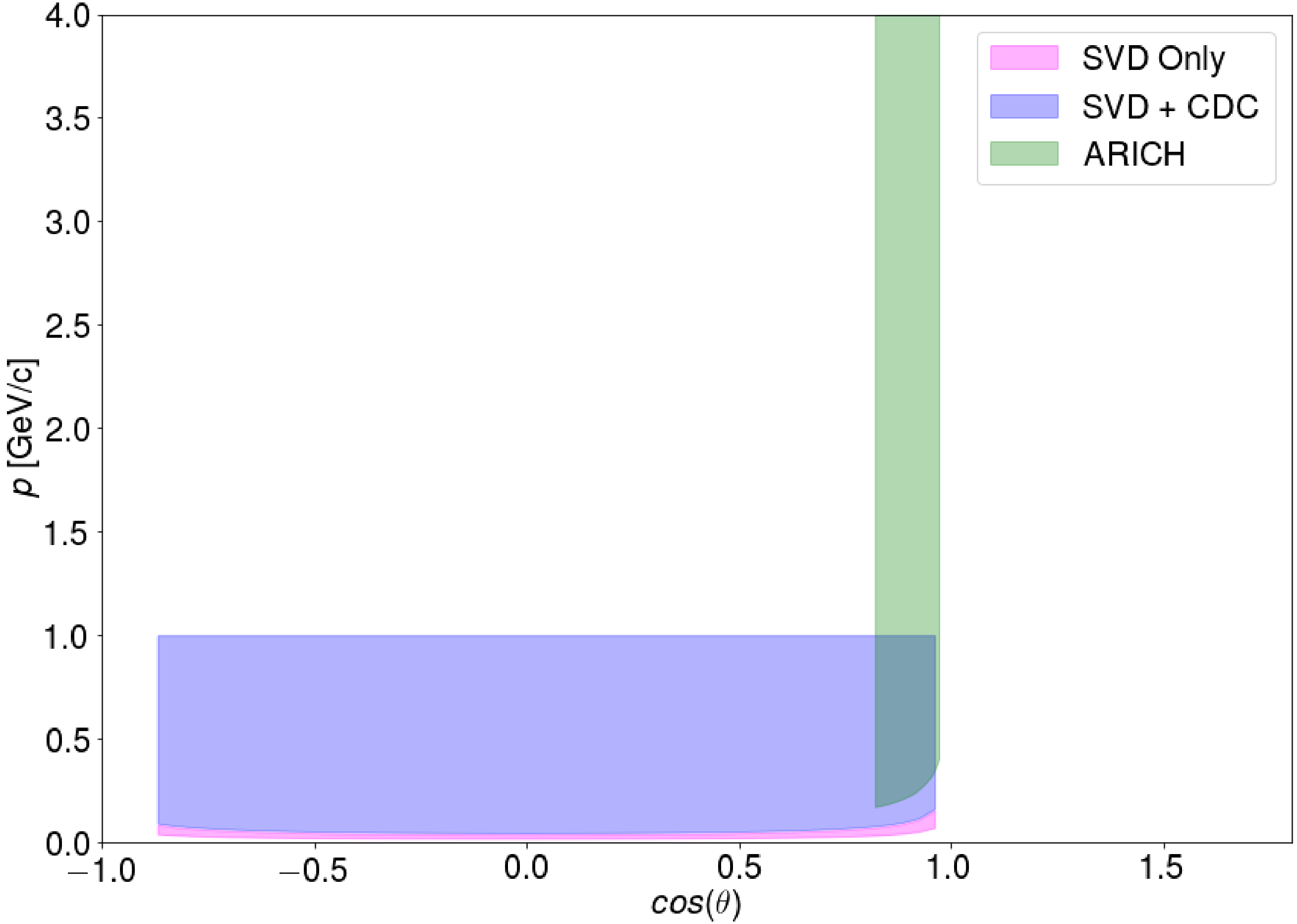
Side A: PID likelihoods

dE/dx: where does it matter the most



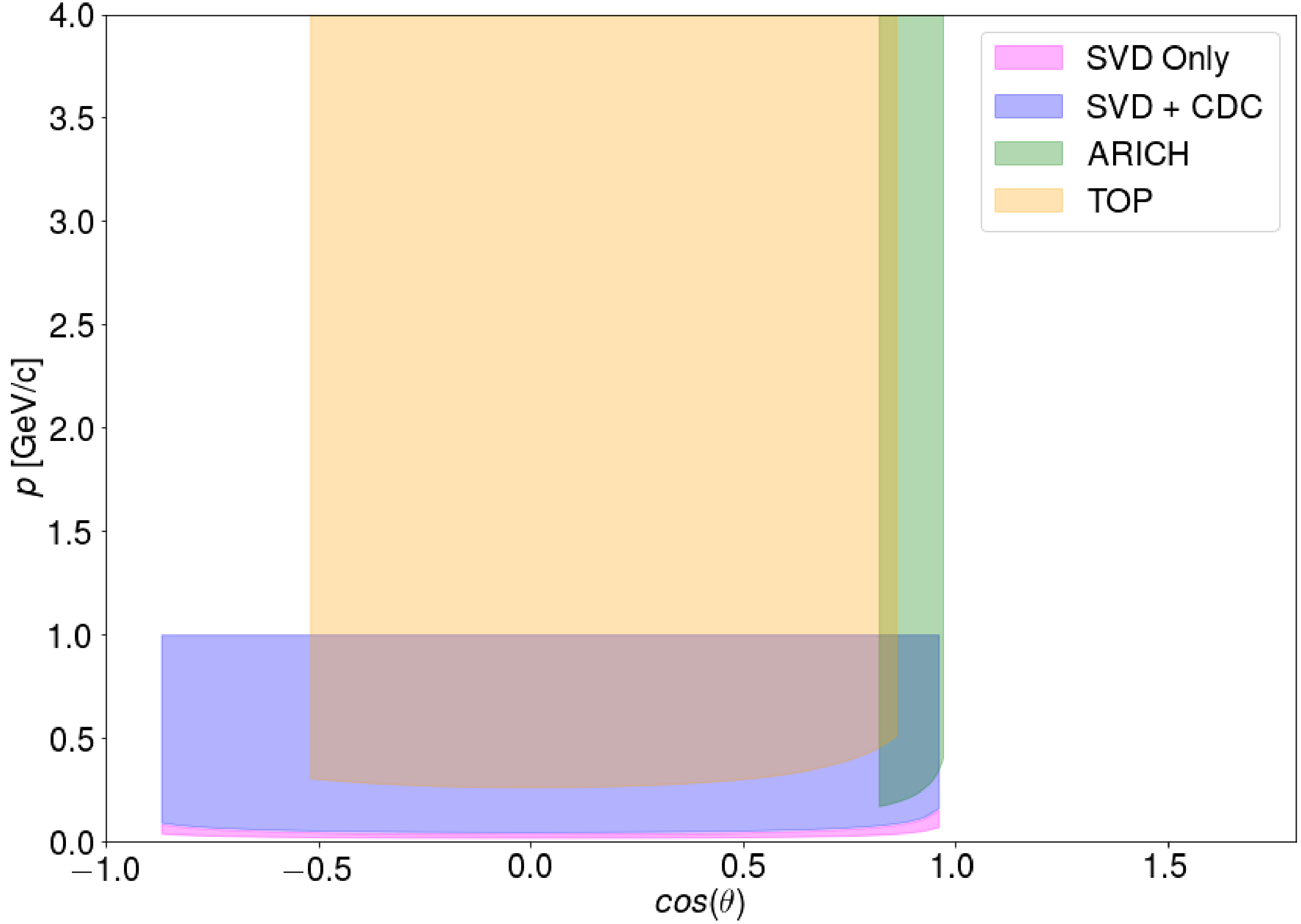
For illustration only. Do not use this plot to get some serious number

ARICH



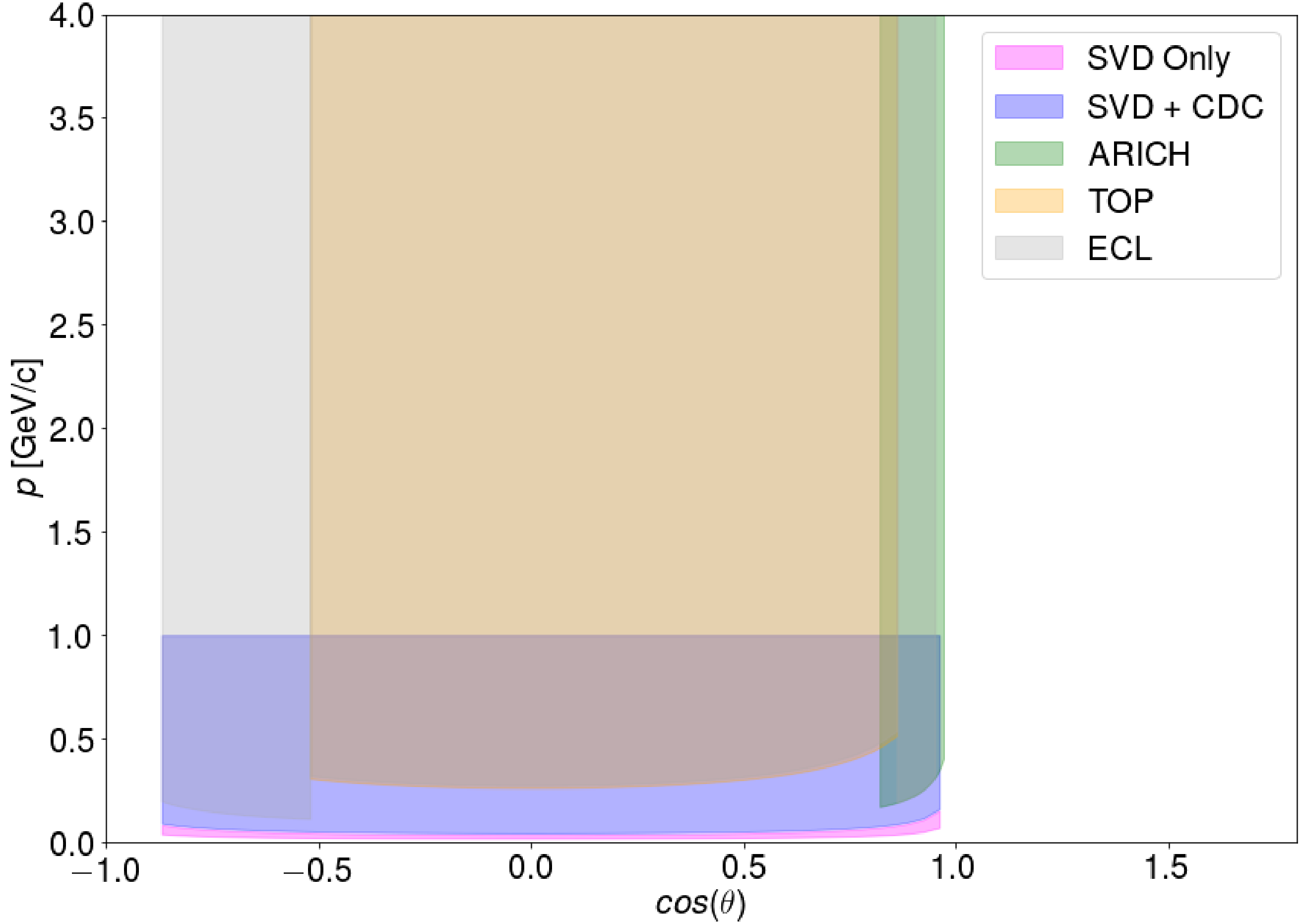
For illustration only. Do not use this plot to get some serious number

TOP



For illustration only. Do not use this plot to get some serious number

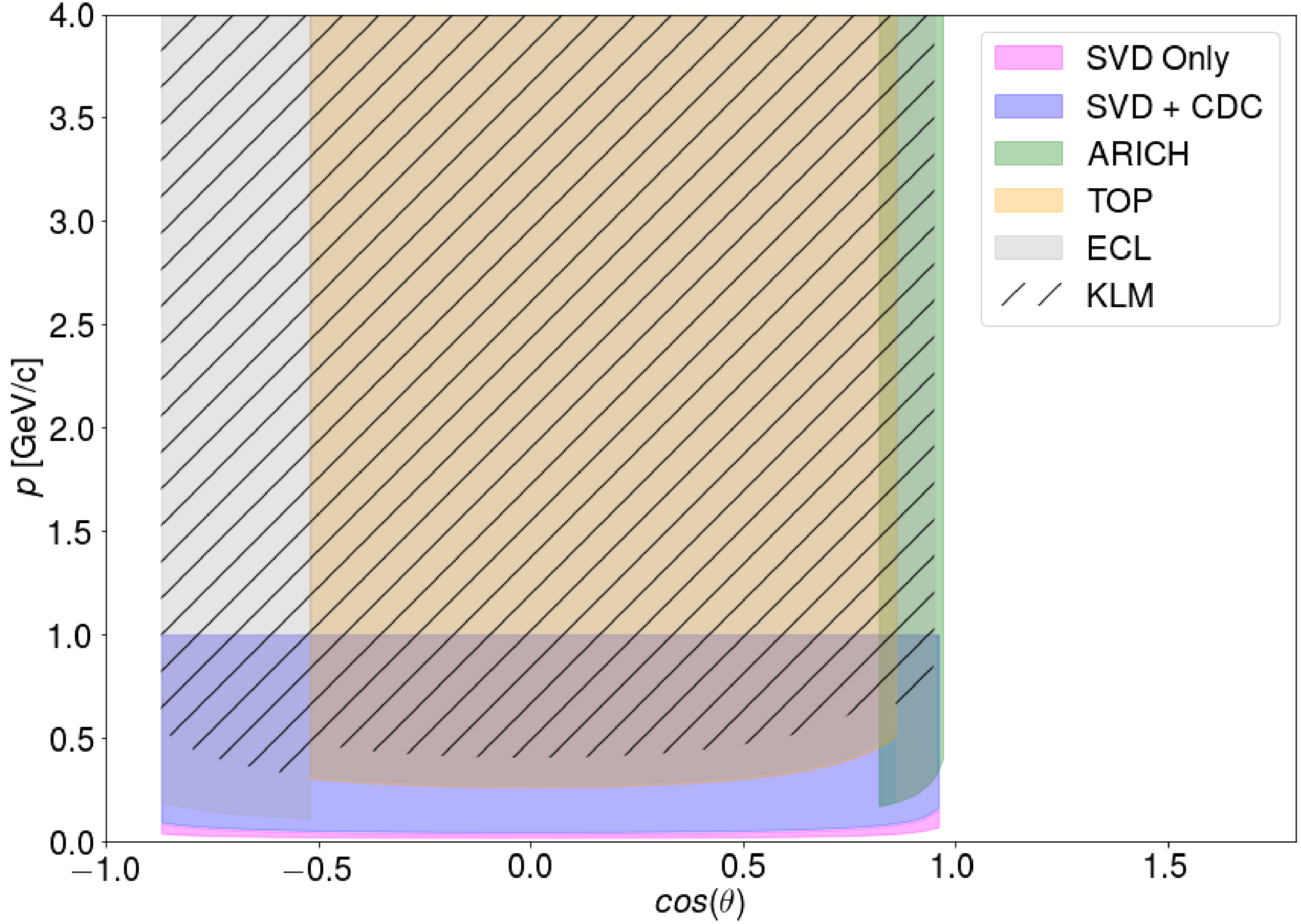
ECL



For illustration only. Do not use this plot to get some serious number

The ECL dominates the <u>electron</u> identification

KLM



For illustration only. Do not use this plot to get some serious number

The KLM dominates the <u>muon</u> identification

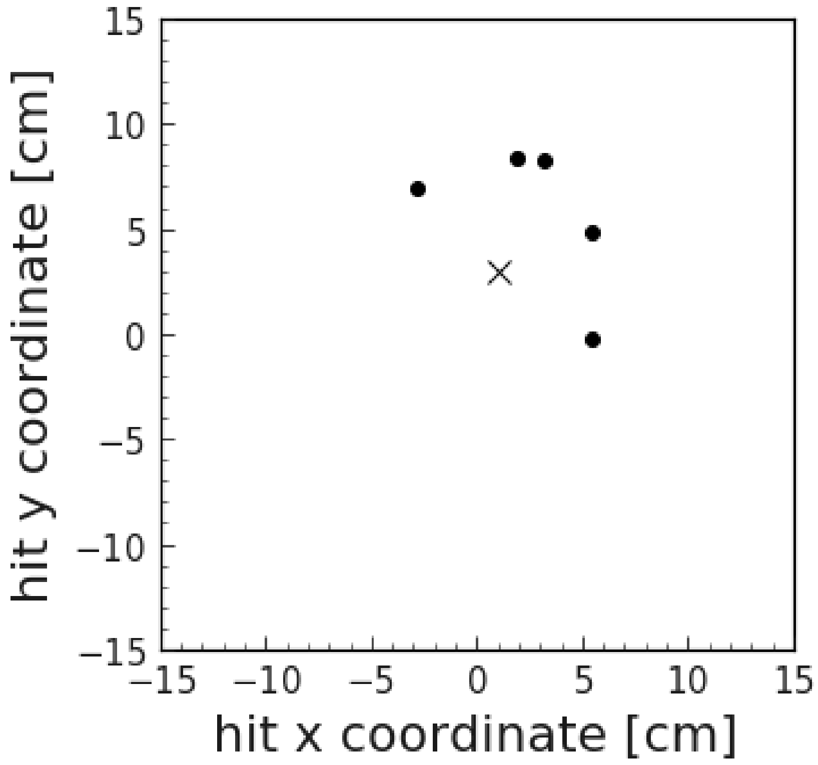
Combining all the information

How can we combine in a coherent way all the signals from the sub-detectors?

1) Each detector fits the distribution of its hits with six PDFs (one per species)

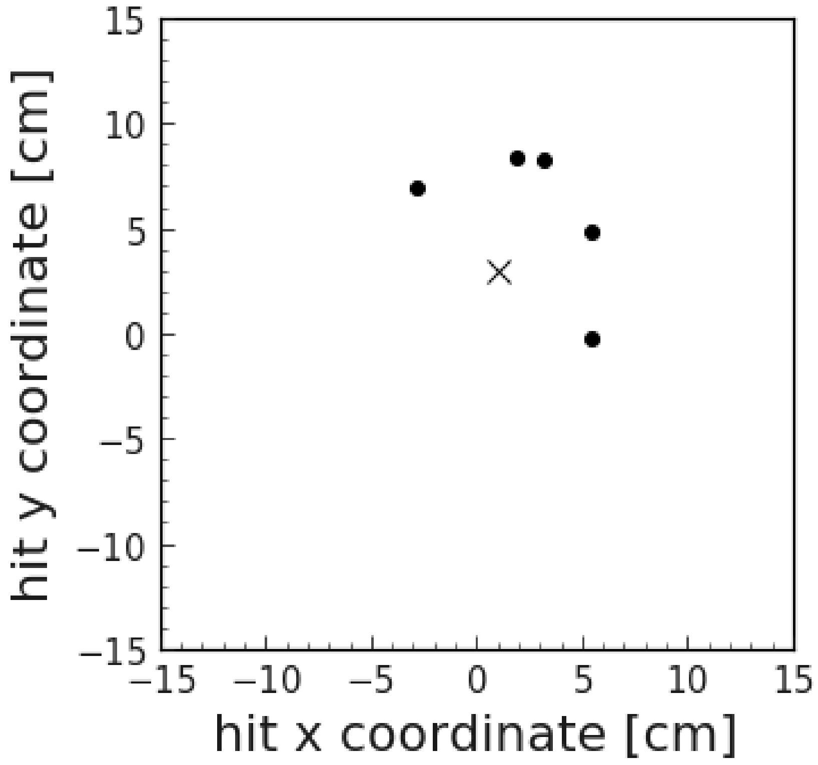
Toy Likelihood with an ARICH-like device

Detector level

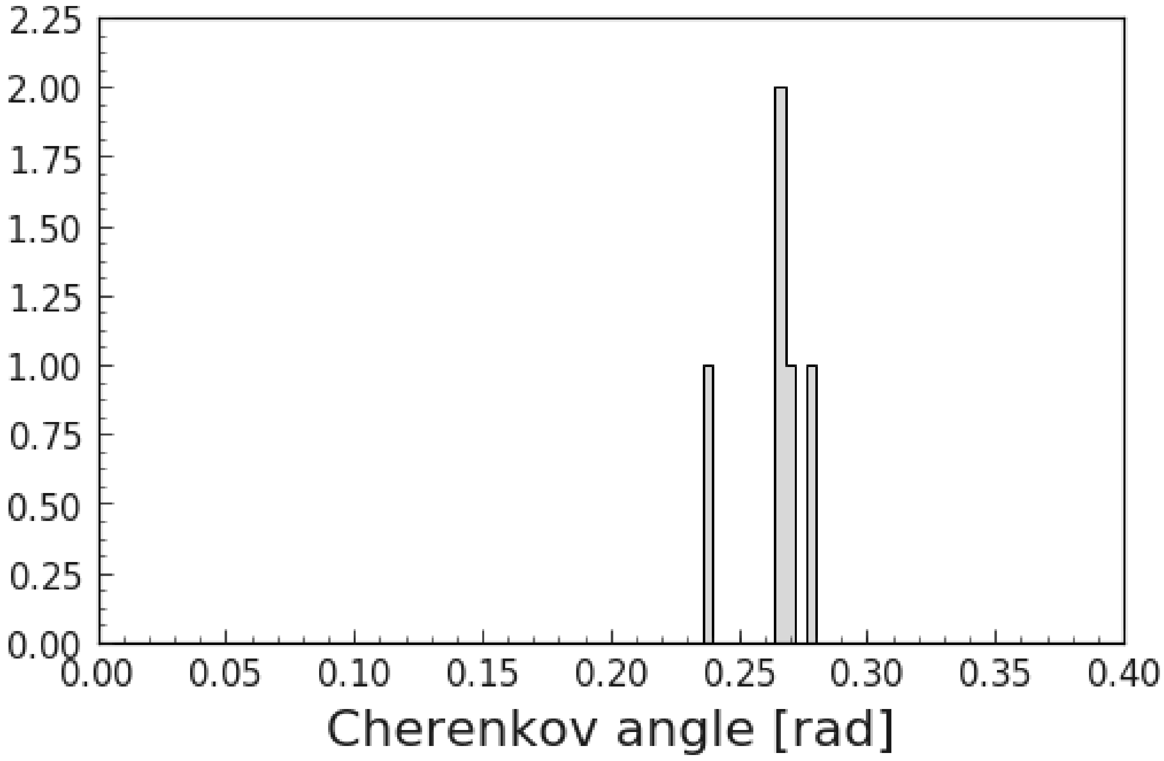


Toy Likelihood with an ARICH-like device

Detector level



Reconstruction level

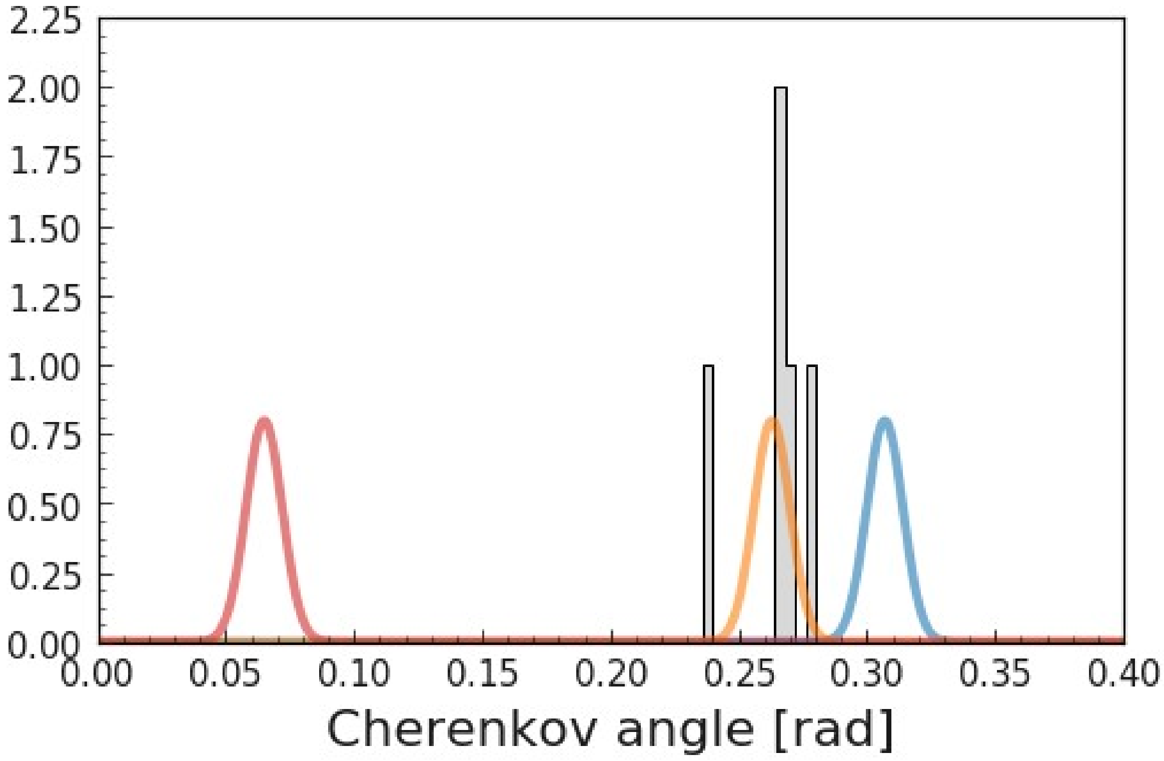


Toy Likelihood with an ARICH-like device

Compare the observed distro with the expected one for each particle type

pi, K and p here. Which is which?

Reconstruction level



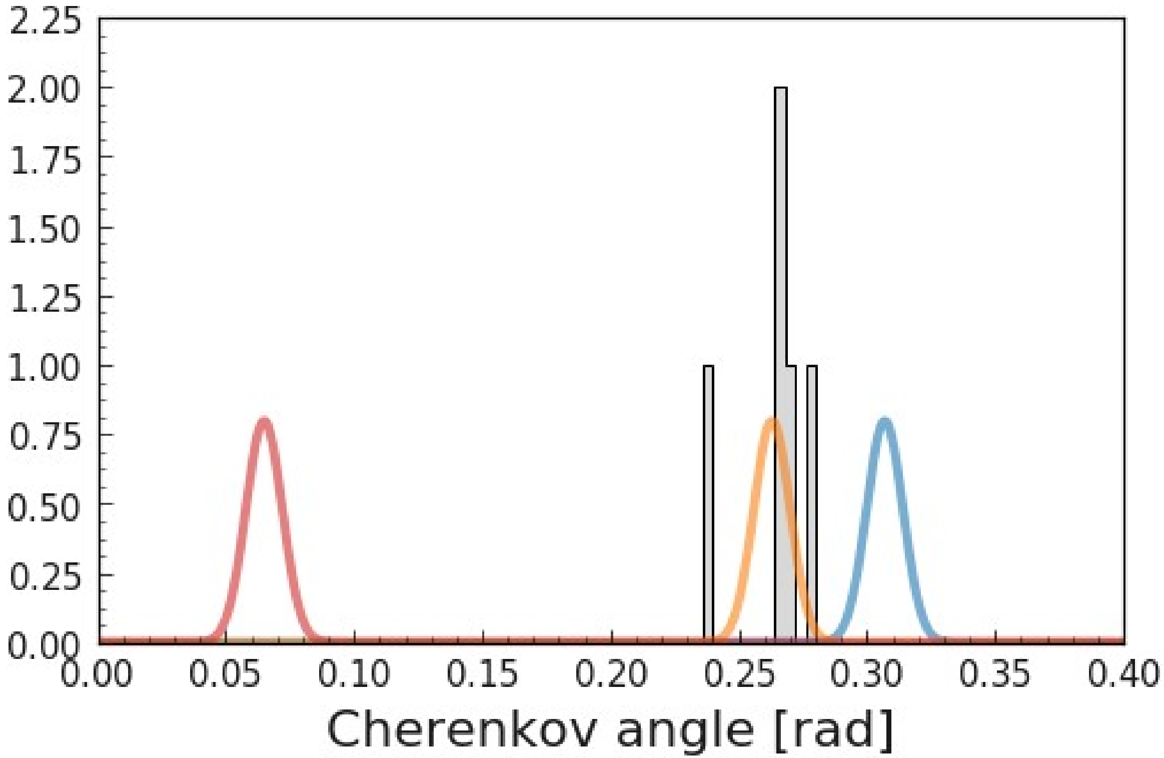
Toy Likelihood with an ARICH-like device

Compare the observed distro with the expected one for each particle type

pi, K and p here. Which is which?

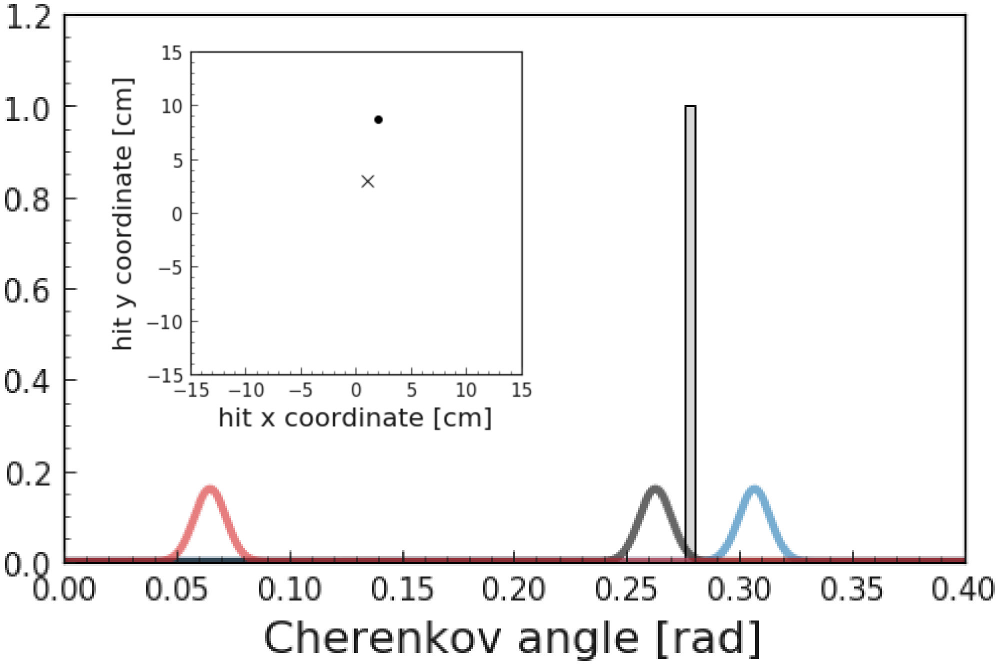
LL(π) = -165
LL(K) = -28
LL(p) = -1805

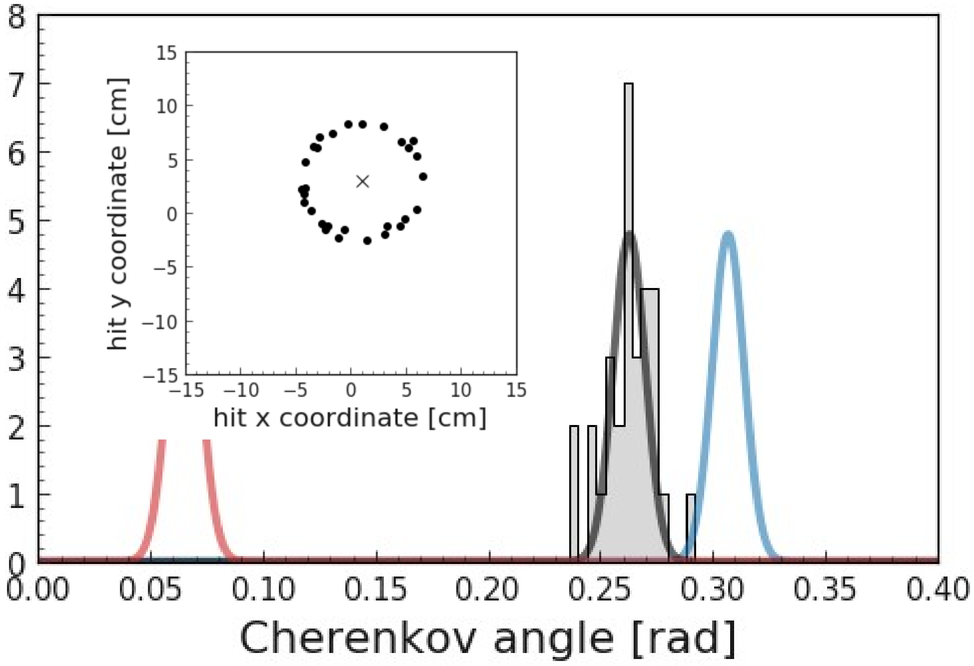
Reconstruction level



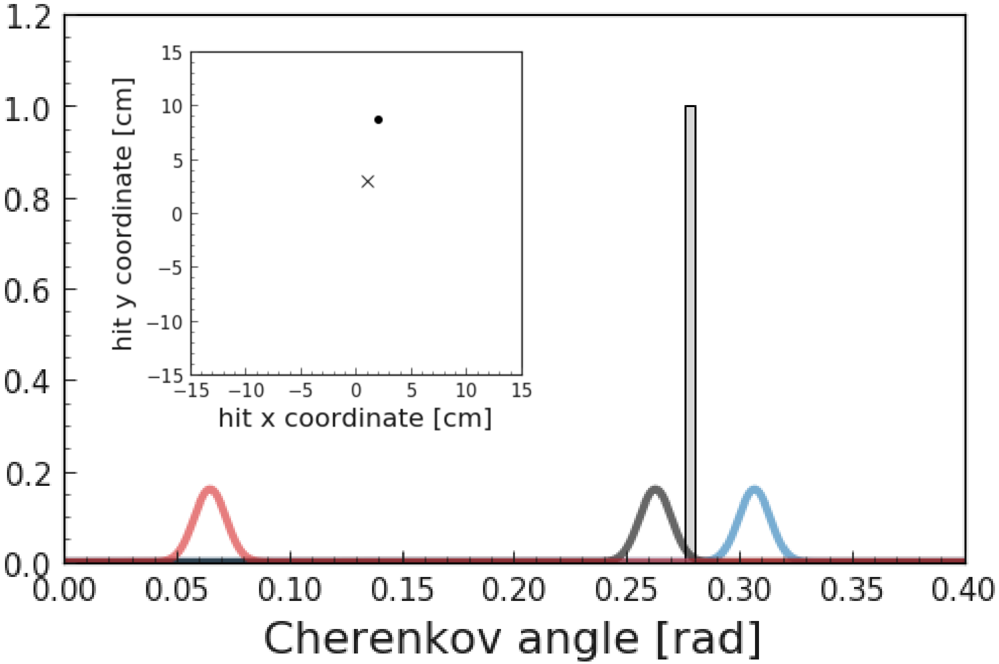
Fun with likelihoods

Will a kaon always have a high kaon LL?

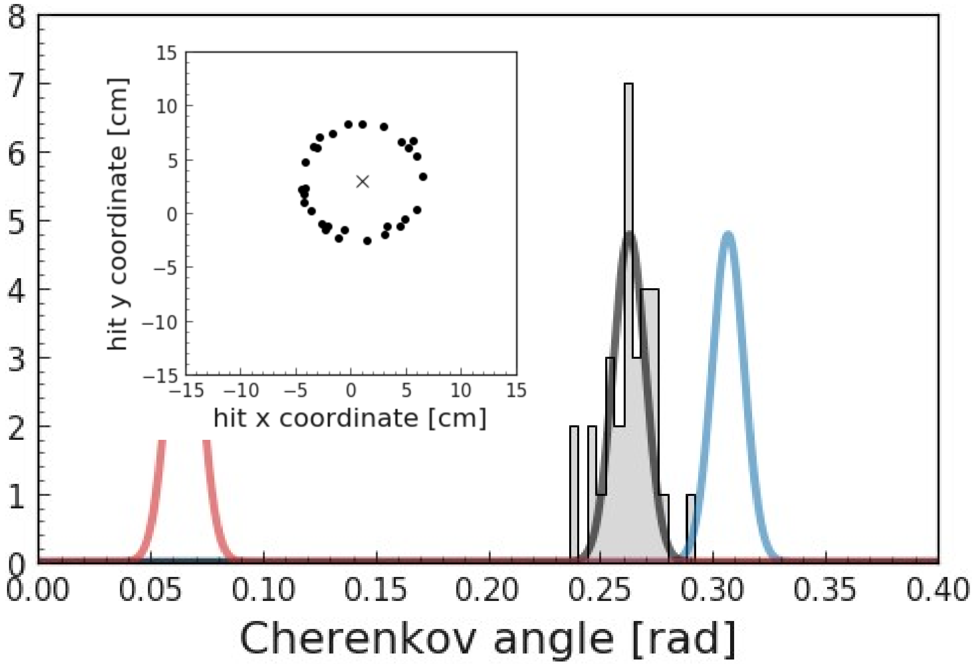




Will a kaon always have a high kaon LL?



LL(K) = -6



LL(K) = -135

Getting a PID value: DeltaLL

How do we compare different hypotheses?

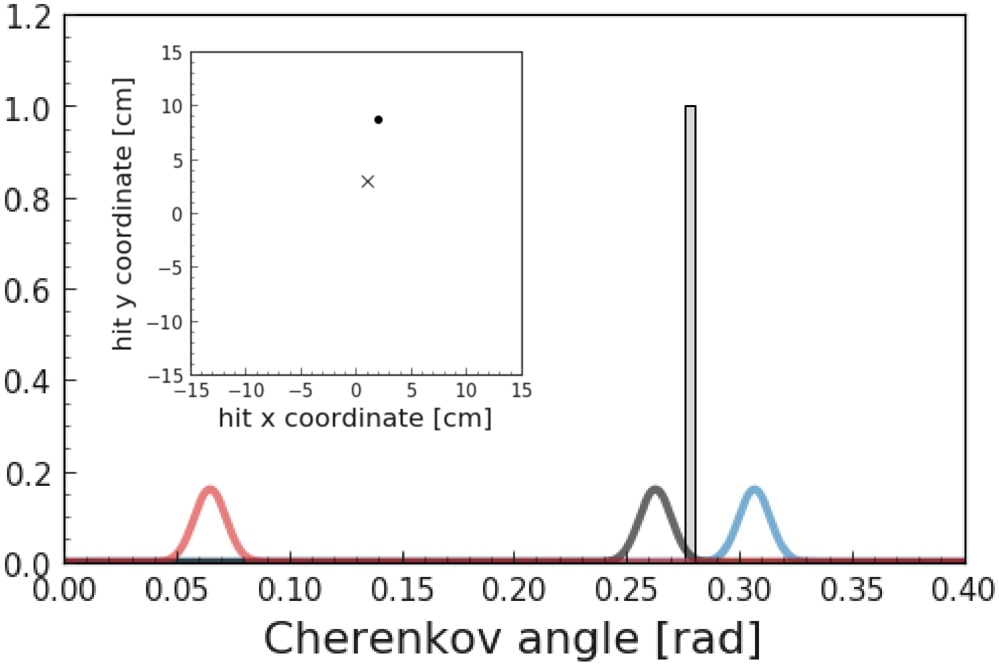
→ The better the mass hypothesis fits the data, the larger the likelihood is.

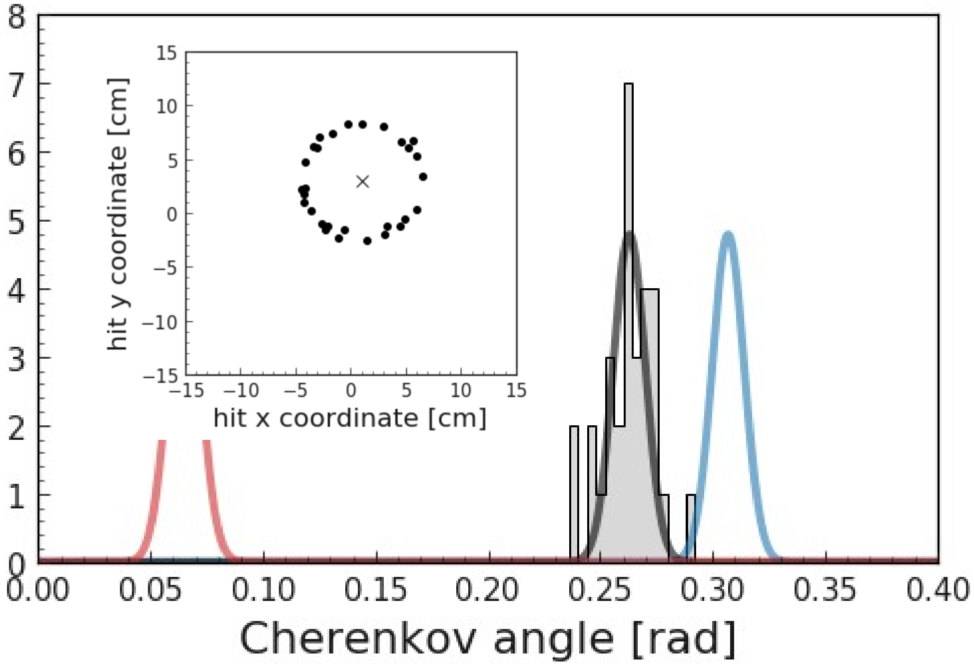
→ The most basic comparison **is the Log-likelihood difference**

$$\Delta LL = \log L_A - \log L_B$$

→ **ΔLL tells you which one of two hypotheses is the most likely**

DeltaLL examples





DLL(K,π) = 4.6 **DLL(K,π) = 590**

Combining all the information

How can we combine in a coherent way all the signals from the sub-detectors?

1) Each detector fits the distribution of its hits with six PDFs (one per species)

2) The outcome of each fit is quantified in a (Log)-likelihood value

Combining all the information

How can we combine in a coherent way all the signals from the sub-detectors?

1) Each detector fits the distribution of its hits with six PDFs (one per species)

2) The outcome of each fit is quantified in a (Log)-likelihood value

3) The for each mass hypothesis, we sum the LogLikelihoods of the sub-detectors to construct **a single particle likelihood**

$$\log \mathcal{L}_\pi = \log \mathcal{L}_\pi^{\mathrm{SVD}} + \log \mathcal{L}_\pi^{\mathrm{CDC}} + \log \mathcal{L}_\pi^{\mathrm{TOP}} + \log \mathcal{L}_\pi^{\mathrm{ARICH}} + \log \mathcal{L}_\pi^{\mathrm{ECL}} + \log \mathcal{L}_\pi^{\mathrm{KLM}}$$

Weighting the detector info

Imagine two detectors D1 and D2

$\rightarrow$ D1 can separate π and μ

$\rightarrow$ D2 cannot

Should we weight the subdetectors in the combined LL?

Weighting the detector info

Should we weight the subdetectors in the combined LL?

Imagine two detectors D1 and D2

$\rightarrow$ D1 can separate π and μ

$\rightarrow$ D2 cannot

$$\Delta \log L(\pi, \mu) = \log L(\pi) - \log L(\mu)$$

Weighting the detector info

Should we weight the subdetectors in the combined LL?

Imagine two detectors D1 and D2

→ D1 can separate π and μ

→ D2 cannot

$$\Delta \log L(\pi, \mu) = \log L(\pi) - \log L(\mu)$$

$$\log L(\mu) \rightarrow \log L_{D1}(\mu) + \log L_{D2}(\mu)$$

$$\log L(\pi) \rightarrow \log L_{D1}(\pi) + \log L_{D2}(\pi)$$

Weighting the detector info

Should we weight the subdetectors in the combined LL?

Imagine two detectors D1 and D2

→ D1 can separate π and μ

→ D2 cannot

$$\Delta \log L(\pi, \mu) = \log L(\pi) - \log L(\mu)$$

$$\Delta \log L(\pi, \mu) = \Delta \log L_1(\pi, \mu) + \Delta \log L_2(\pi, \mu)$$

Question: Does D2 contribute?

Weighting the detector info

Should we weight the subdetectors in the combined LL?

NO, Likelihoods are “self-weighting”

Side B: PID probabilities

From likelihoods to probabilities

DLL is a powerful tool to understand a detector's performance

However, how do you quantify the "PID level" of a particle in an understandable way?

From likelihoods to probabilities

DLL is a powerful tool to understand a detector's performance

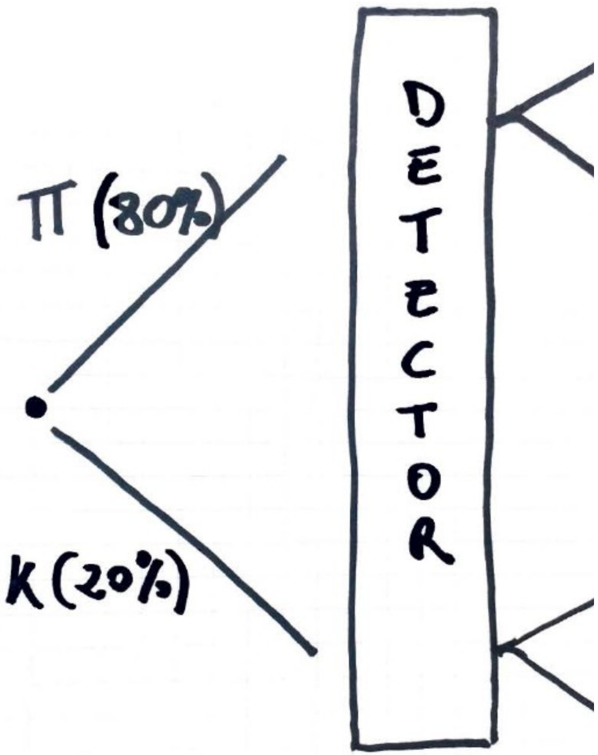
However, how do you quantify the “PID level” of a particle in an understandable way?

PID is inherently a Bayesian problem.

I observe a “kaon-like” signal, and want to know what's the probability for that signal to be really generated by a kaon

From likelihoods to probabilities

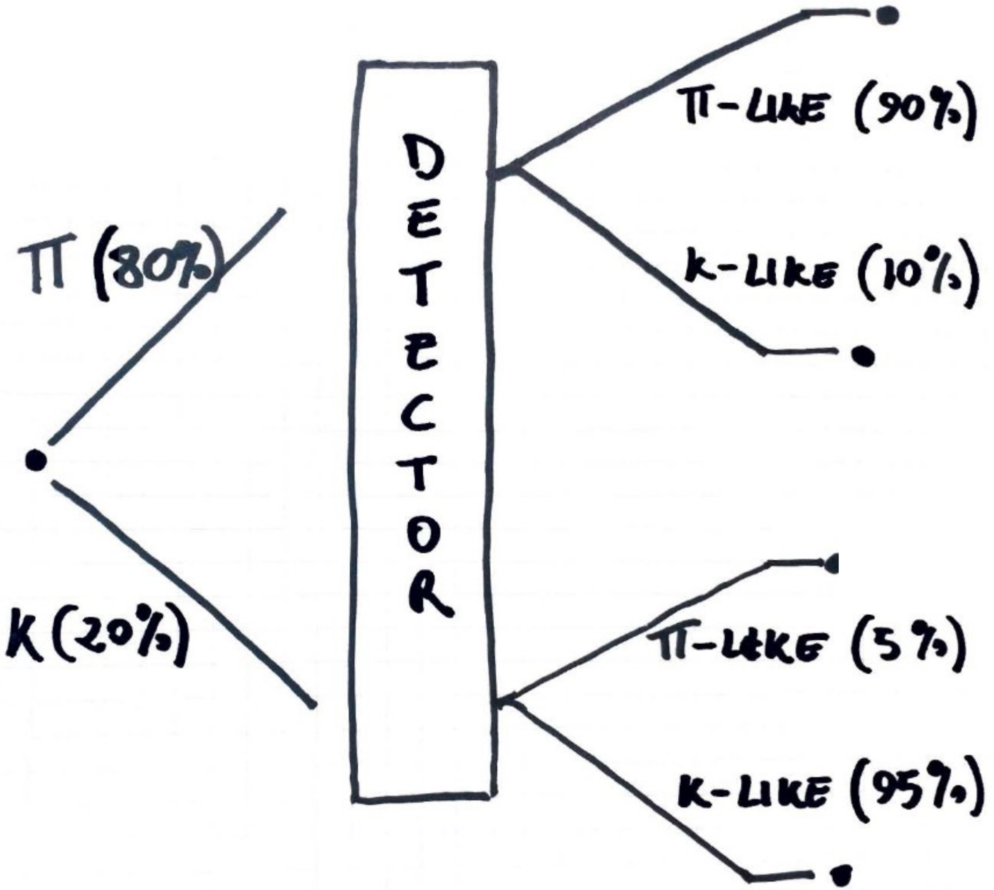
A very simple example: a magic universe where only pions and kaons exist.



Let's assume that our data sample contains 20% kaons and 80% pions (how do we know it?)

From likelihoods to probabilities

A very simple example: a magic universe where only pions and kaons exist.

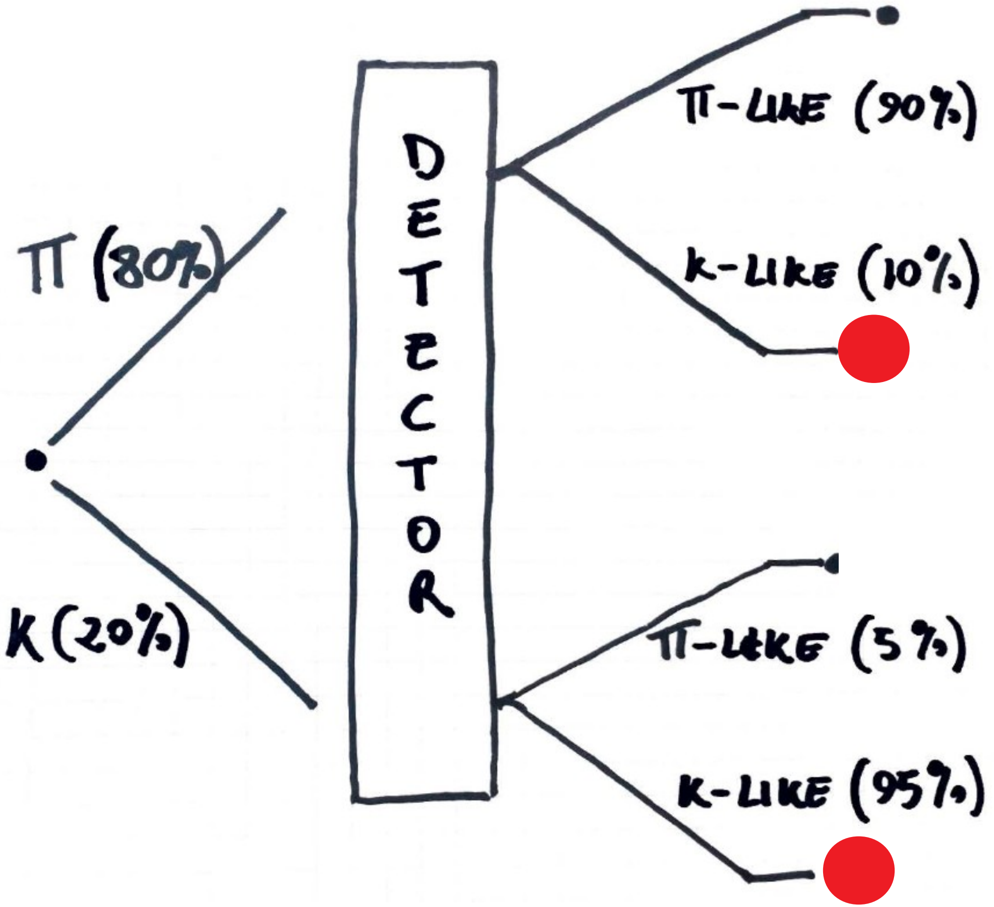


Let's assume that our data sample contains 20% kaons and 80% pions (how do we know it?)

The detector has a certain probability of assigning pion or kaon ID, depending on the original particle

From likelihoods to probabilities

A very simple example: a magic universe where only pions and kaons exist.



We observe a "kaon-like" signal. What's the probability for that particle to be a kaon?

From likelihoods to probabilities

A very simple example: a magic universe where only pions and kaons exist.

We observe a "kaon-like" signal. What's the probability for that particle to be a kaon?

Posterior probability

$$P(\text{S is from K}) = \frac{P(\text{K gives S}) \cdot P(K)}{P(\text{K gives S}) \cdot P(K) + P(\pi \text{ gives S}) \cdot P(\pi)}$$

Prior probability

From likelihoods to probabilities

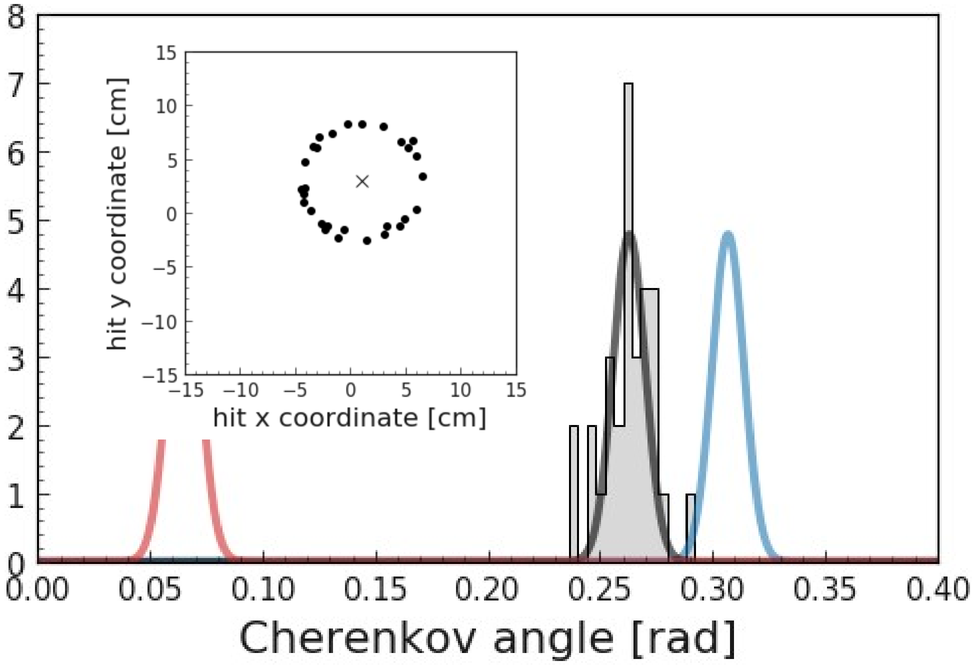
The likelihood value is actually a proxy (i.e. is proportional) exactly to the conditional probability!

$$P(\text{S is from K}) = \frac{L(K) \cdot P(K)}{L(K) \cdot P(K) + L(\pi) \cdot P(\pi)}$$

Belle II default PID variables are posterior probabilities

Binary PID

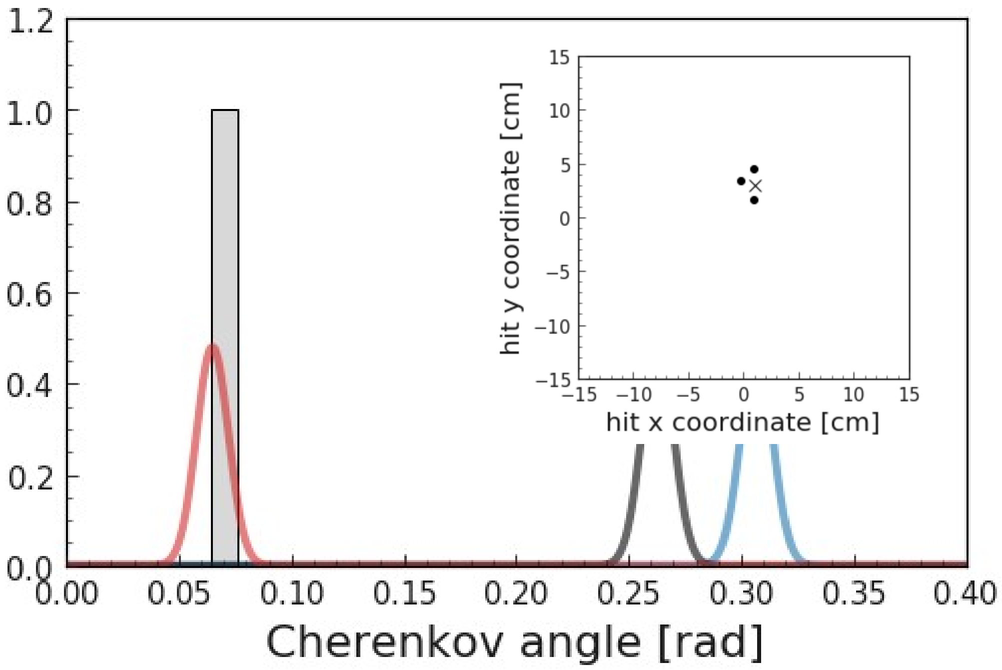
?



DLL(K,π) = 568
P(K, π) = 1

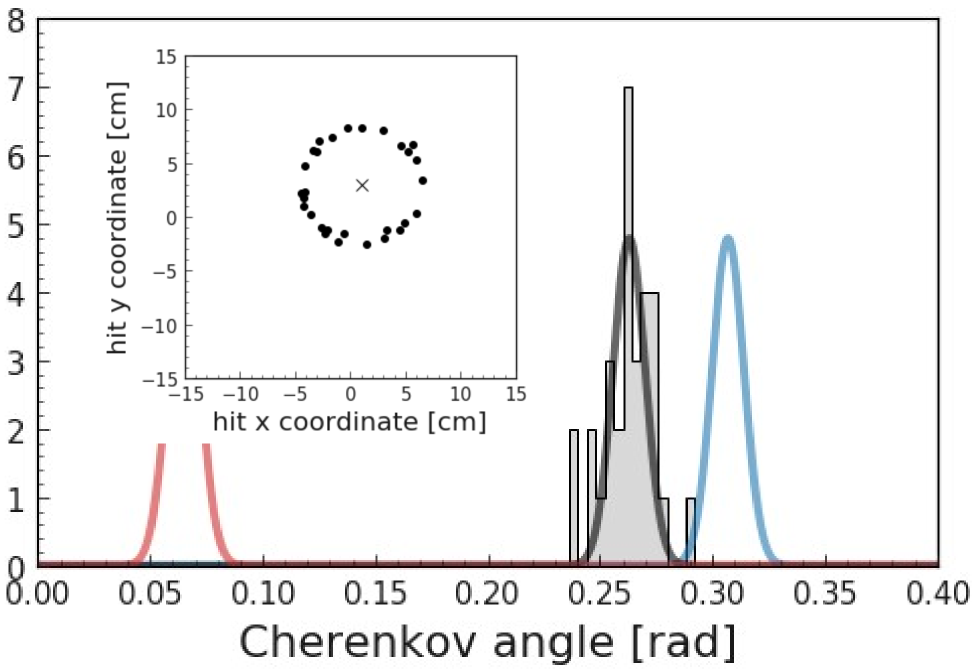
DLL(K,π) = 590
P(K,π) = 1

Binary PID



DLL(K,π) = 568

P(K, π) = 1

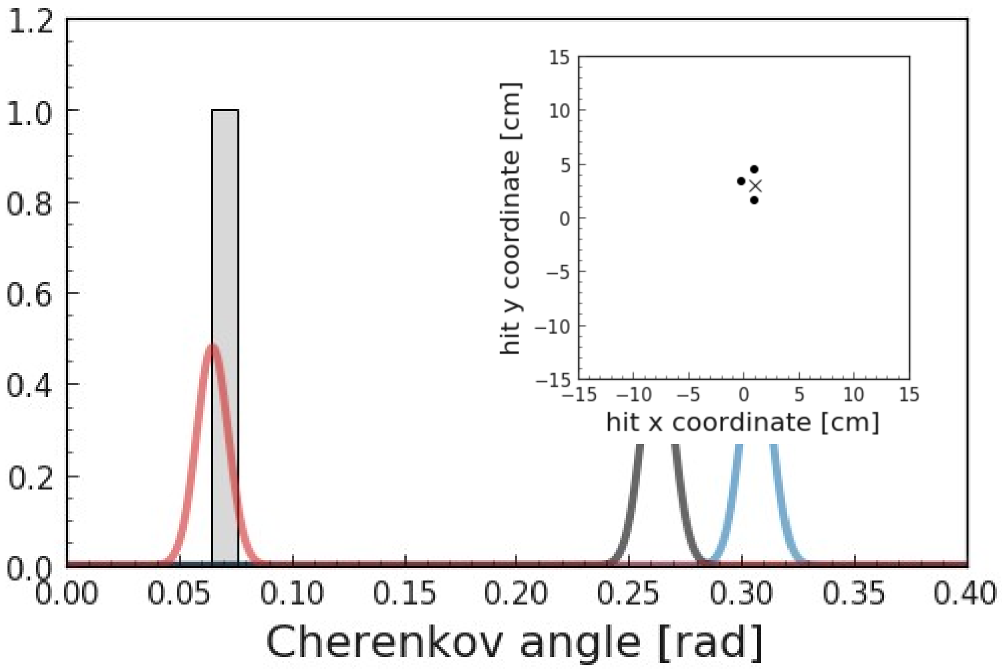


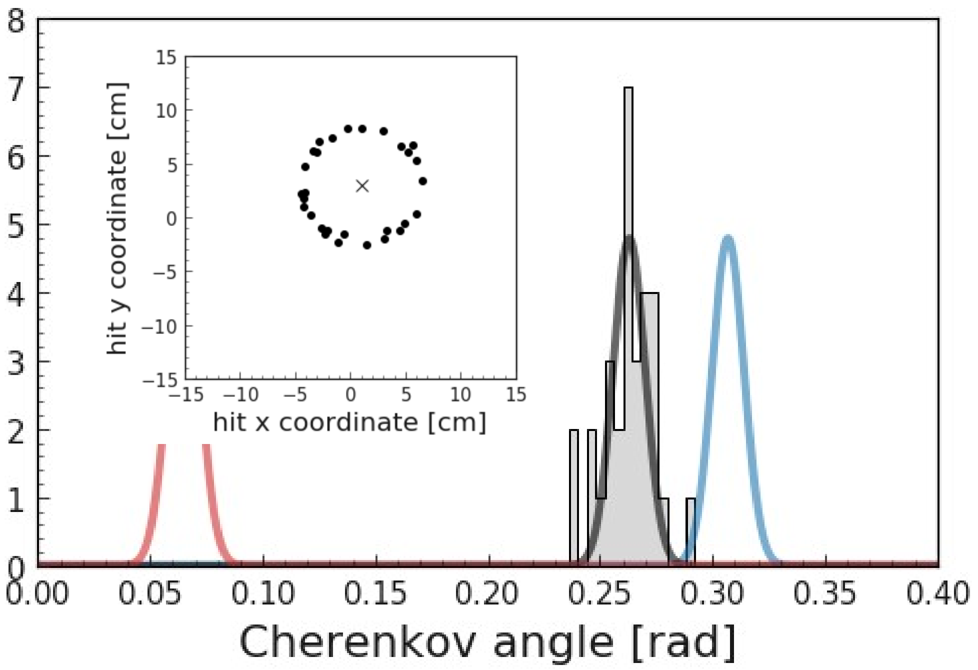
DLL(K,π) = 590

P(K,π) = 1

Delta LL limitations

What is going on here?





DLL(K,π) = 2375
P(K, π) = 1

DLL(K,π) = 612
P(K, π) = 1

“Global” and “Binary” PID

“Binary PID” is a special case of the “global PID”

$$Pid(K,\pi) = \frac{L(K)P(K)}{L(K)P(K) + L(\pi)P(\pi)}$$

$$Pid(K) = \frac{L(K)P(K)}{\sum_{i=e,\mu,\pi,K,p,d} L(i)P(i)}$$

Can you see what the only difference is?

Why priors matter

Likelihood values are meaningless without a reference

PID probabilities are meaningless without a prior scheme

Bonus track: using PID

PID variables in basf2

Basic variables:

```
electronID, muonID, pionID, kaonID, protonID, deuteronID

pidPairChargedBDTScore(pdgCodeHyp, pdgCodeTest)
```

Can you find the documentation yourself?

PID variables in basf2

Basic variables:

```
electronID, muonID, pionID, kaonID, protonID, deuteronID

pidPairChargedBDTScore(pdgCodeHyp, pdgCodeTest)
```

“Expert” variables

```
pidLogLikelihoodValueExpert(pdgCode, detectorList)

pidDeltaLogLikelihoodValueExpert(pdgCode1, pdgCode2, detectorList)

pidPairProbabilityExpert(pdgCodeHyp, pdgCodeTest, detectorList)

pidProbabilityExpert(pdgCodeHyp, detectorList)
```

Performances

Few metrics are used to characterize the performances of a PID detector

$\rightarrow$ **Efficiency: ability to correctly assign the ID**

$\varepsilon(\mathrm{K}) = \mathrm{N}(\mathrm{K}$ identified as $\mathrm{K})/\mathrm{N}(\text{real } \mathrm{K})$

Equal, by definition, to the "probability of a kaon to be called kaon"

Performances

Few metrics are used to characterize the performances of a PID detector

→ **Efficiency: ability to correctly assign the ID**

$\varepsilon(\mathrm{K}) = \mathrm{N}(\mathrm{K}$ identified as $\mathrm{K})/\mathrm{N}(\text{real } \mathrm{K})$

Equal, by definition, to the "probability of a kaon to be called kaon"

→ **Mis-ID probability: ability not to assign the incorrect ID**

$\text{Mis-ID}(\mathrm{K}) = \mathrm{N}(\text{non-}\mathrm{K}$ identified as $\mathrm{K})/\mathrm{N}(\text{non } \mathrm{K})$

Equal, by definition, to the "probability for a non-kaon to be called kaon"

Performances

Few metrics are used to characterize the performances of a PID detector

→ **Efficiency: ability to correctly assign the ID**

$\varepsilon(\mathrm{K})$ = N(K identified as K)/N(real K)

Equal, by definition, to the "probability of a kaon to be called kaon"

→ **Mis-ID probability: ability not to assign the incorrect ID**

Mis-ID(K) = N(non-K identified as K)/N(non K)

Equal, by definition, to the "probability for a non-kaon to be called kaon"

→ **Fake rate: fraction of particles with the wrong ID**

F(K) = N(non-K identified as K)/N(identified as K)

Equal, by definition, to the "fraction of non-kaons in my collection of kaons"

Fake rates VS mis-ID probability

The fake rate is (to a certain extent) a Bayesian idea

Given that I have something that looks like a kaon, what are the chances for this to really be a kaon and not a pion?

Let's assume:

Mis-ID probability $\pi \to K \sim 1\%$

Kaon efficiency $\sim 100\%$

2% of kaons and 98% pions in the data

Fake rates VS mis-ID probability

INFN Istituto Nazionale di Fisica Nucleare
SEZIONE DI TORINO

The fake rate is (to a certain extent) a Bayesian idea

Given that I have something that looks like a kaon, what are the chances for this to really be a kaon and not a pion?

Let's assume:

Mis-ID probability $\pi \rightarrow K \sim 1\%$

Kaon efficiency ~ 100%

2% of kaons and 98% pions in the data

Bayes theorem:

P(my "kaon" is kaon) = 1 x 0.02 / (1 x 0.02 + 0.01 x 0.98) ~ 67%

Fake rate = 33%

One slide summary

Istituto Nazionale di Fisica Nucleare
SEZIONE DI TORINO

1) PID variables are probabilities

$\rightarrow$ Bayes theorem with Likelihoods are conditional probabilities

$\rightarrow$ Priors are constant (for now)

2) Don't confuse fake rate with mis-ID probability

3) Things will improve in future

$\rightarrow$ Priors will be implemented

$\rightarrow$ ML to properly deal with high order correlations