Bayesian Likelihood Combination for PID detectors - ALICE and Belle II

Oskar Hartbrich (ORNL)

With many thanks to
U. Tamponi (INFN Torino)

Acknowledgements

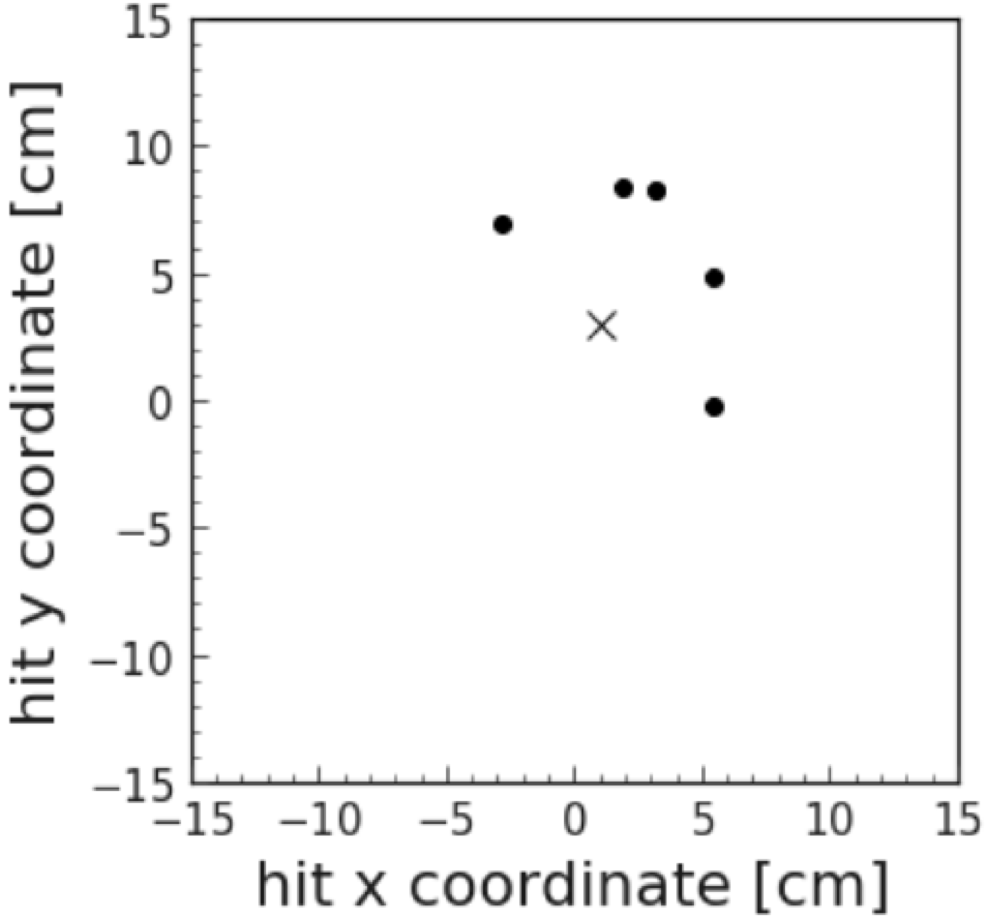
- Based on introductory material Umberto Tamponi (INFN Torino) and OH prepared for “Belle II Start Kit” workshops
- Excellent paper by ALICE colleagues on the very same topic (and much more! Must read!)
  - Particle identification in ALICE: a Bayesian approach

Particle Identification

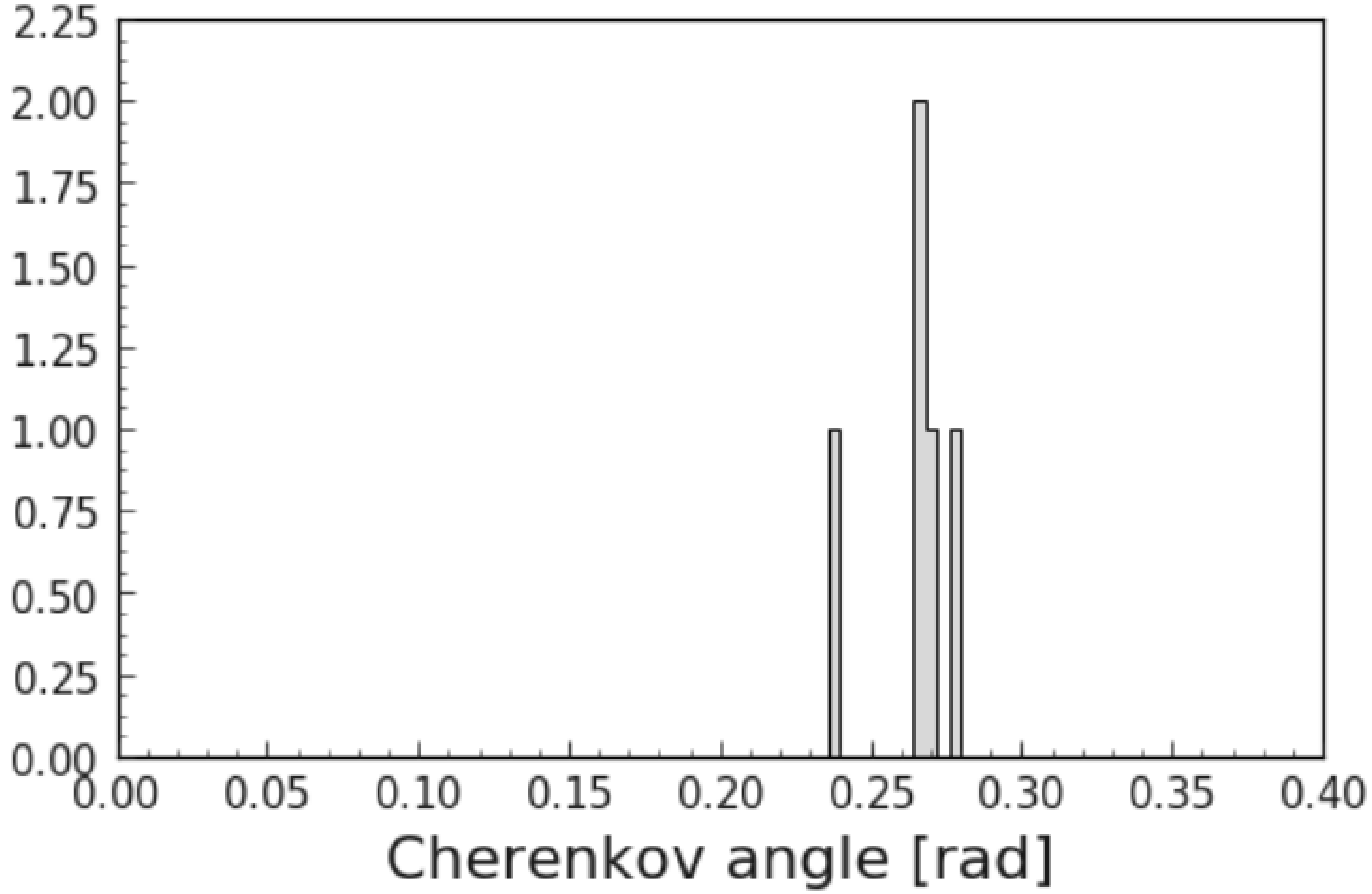
- PID is a rest mass measurement (in practice)
  - Rest masses are known, so classification into discrete types:
    $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$, $p^{\pm}$, ($D^{\pm}$)
- Each detector gives *some* answer for a given track, but how do we coherently combine this information?
  - What "format" is needed for that information?
- (Log-)Likelihoods for each particle hypothesis
  - Def.: Likelihood is a probability value obtained for a single measurement from an (arbitrary dimensional) PDF
  - PDFs are normalized to unity (!)

Toy Likelihood with a RICH-like Device

Detector level

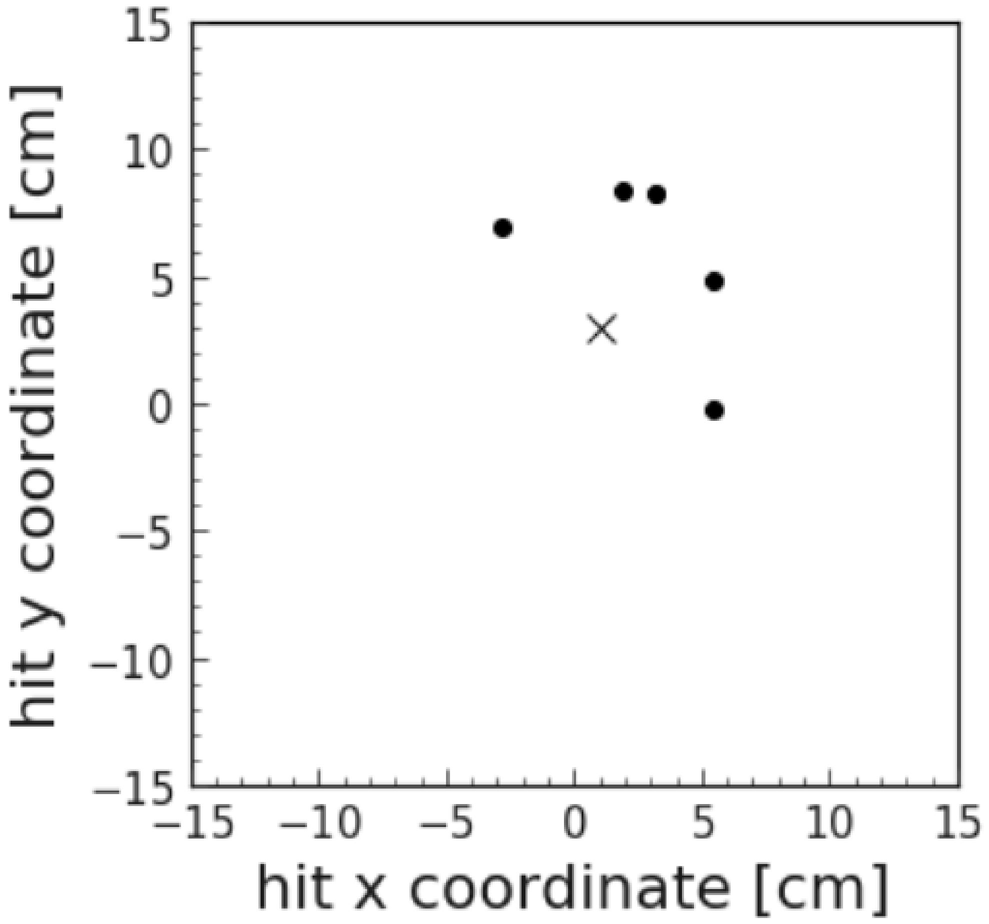


Reconstruction level

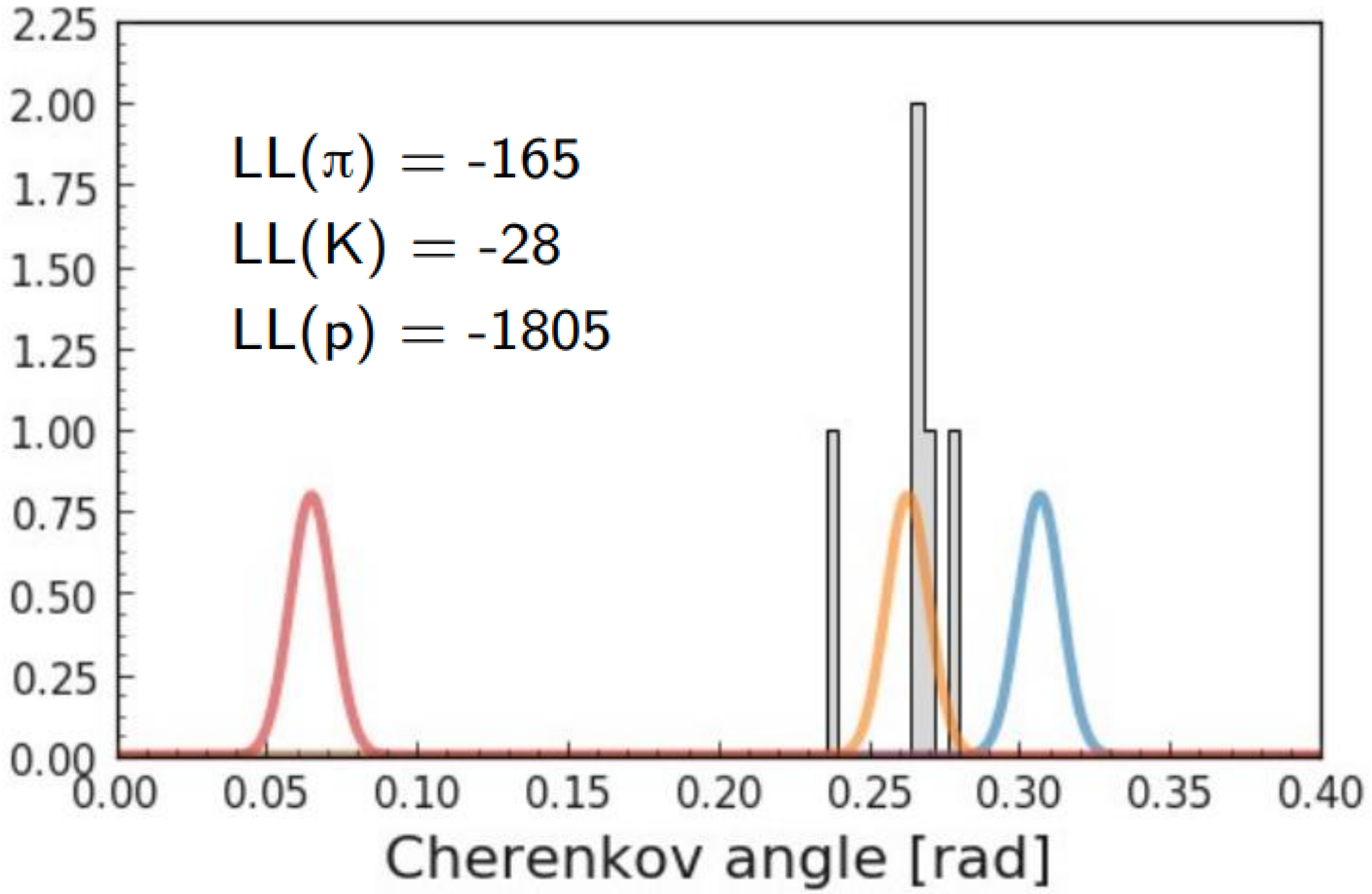


Toy Likelihood with a RICH-like Device

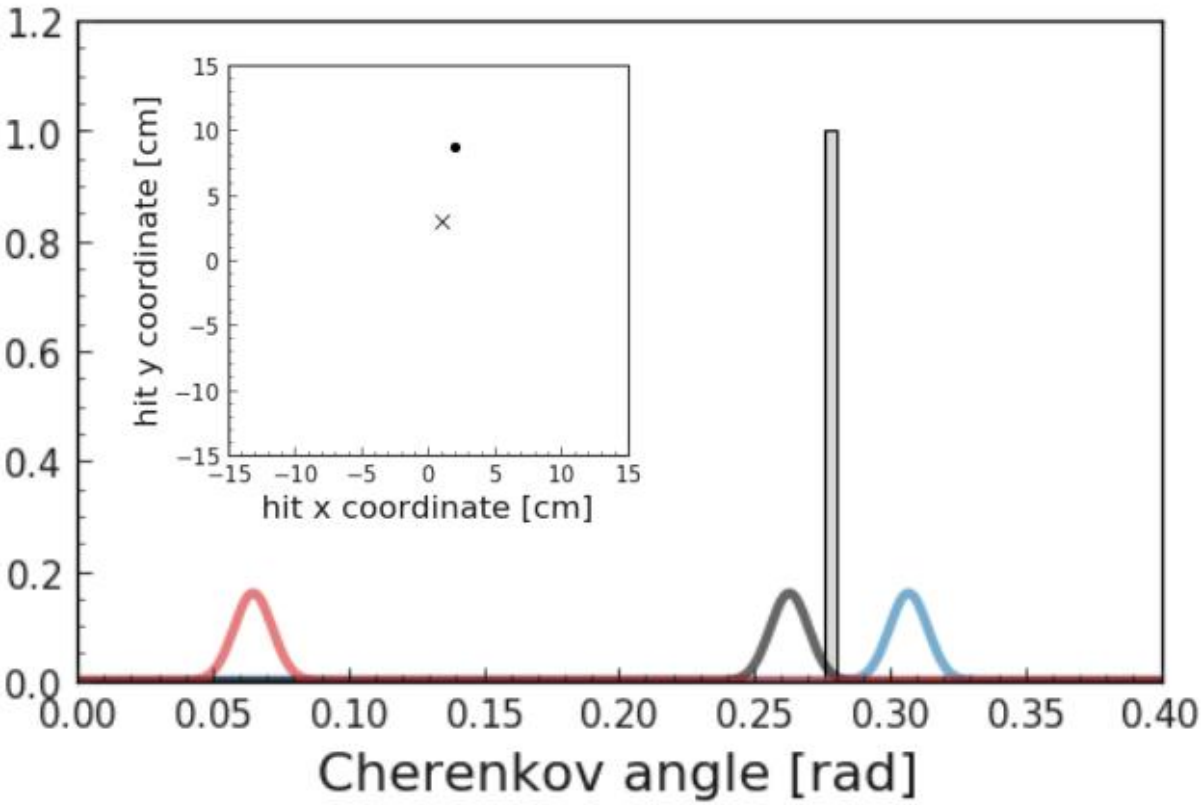
Detector level



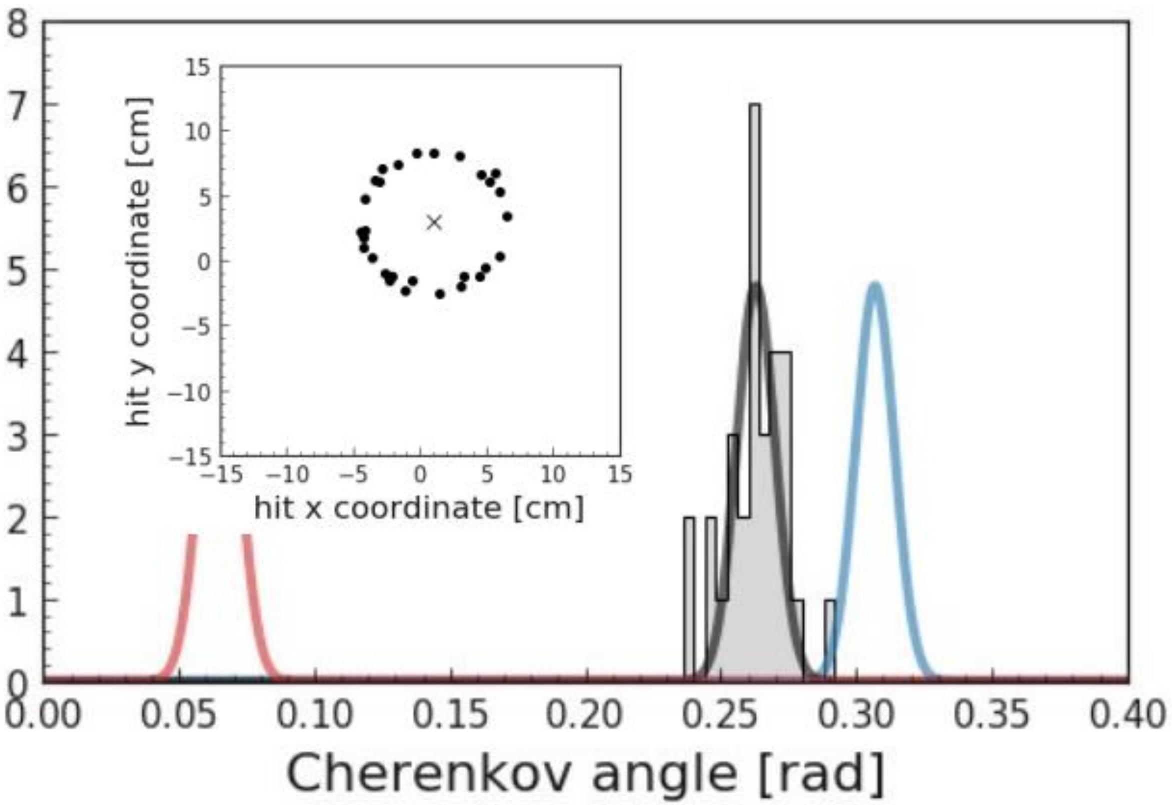
Reconstruction level



Absolute Likelihood Values are Meaningless



LL(K) = -6



LL(K) = -135

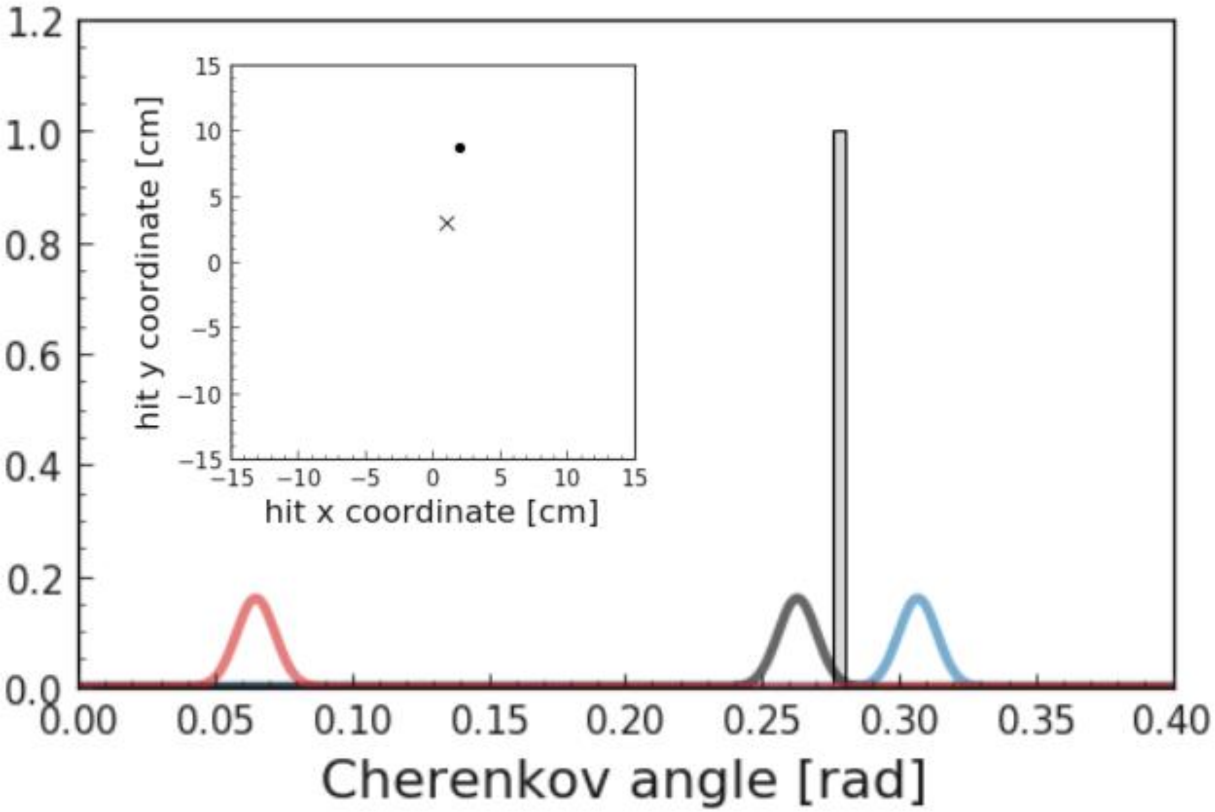
Comparing Hypotheses: ΔLL

- The difference in Log-Likelihoods holds all the information:

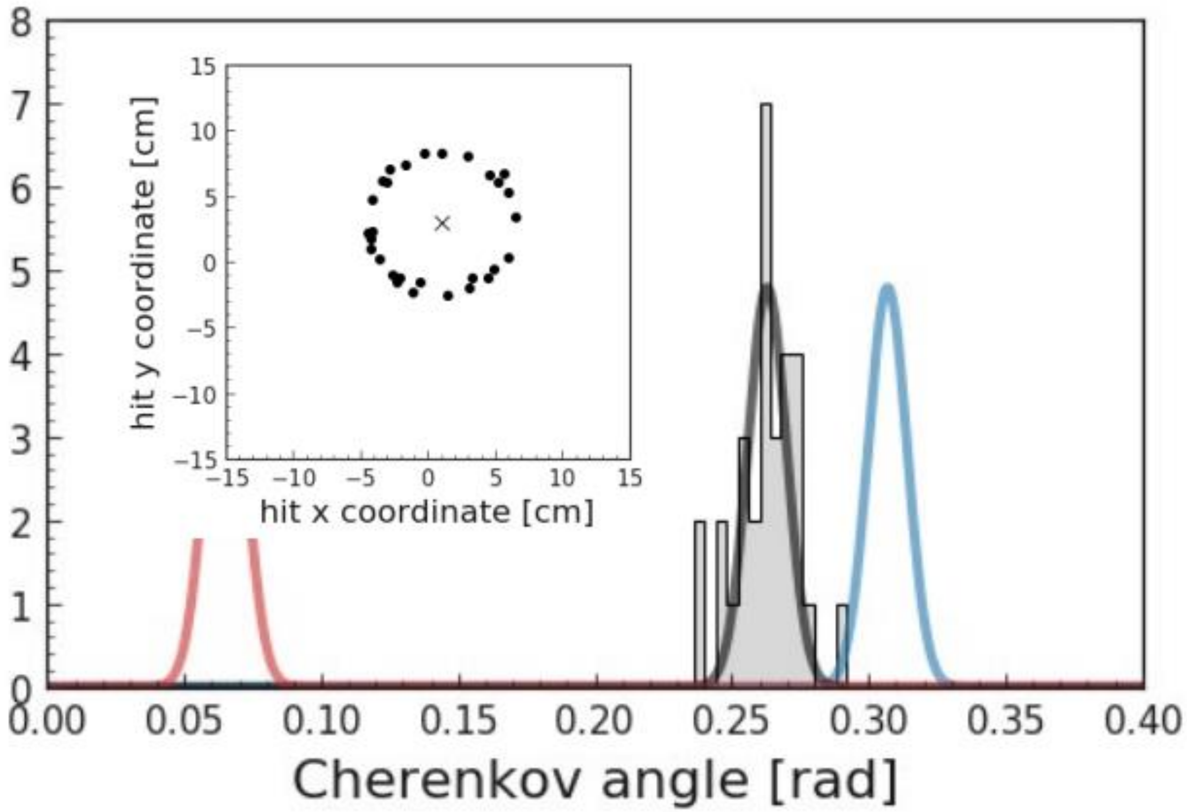
$$\Delta LL = \log L_A - \log L_B$$

- ΔLL tells you which one of the two hypotheses is more likely

Absolute Likelihood Values are Meaningless



DLL(K,π) = 4.6



DLL(K,π) = 590

Combining Likelihoods

- Combining Log-Likelihoods from different detector systems is just summing them up:

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

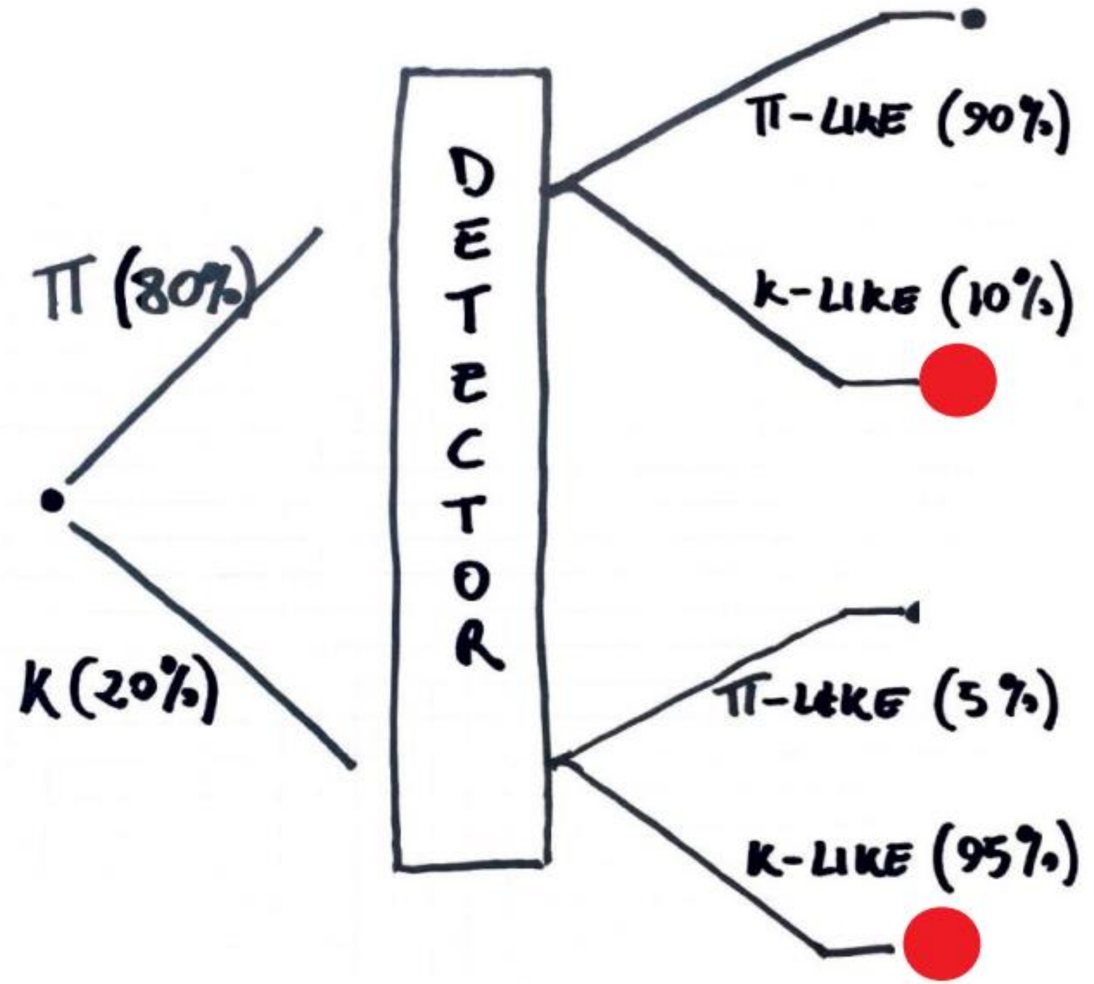
  - (this is the same as adding up LLs from individual photons in the previous toy-RICH examples)

- Then ΔLL works just the same as before.

From Likelihoods to Probabilities

- ΔLL is powerful to understand performance, but hard to interpret on its own
- Need a quantification of "PID level" for each track
- PID is a Bayesian problem
  - We observe a "kaon-like" signal
  - What is probability that this was generated by a kaon

From Likelihoods to Probabilities – A Simple Example



- Toy Universe:
  80% pions, 20% kaons
- We observe a kaon-like signal
- What is the probability that this was an actual kaon?
- Bayes' Theorem

From Likelihoods to Probabilities – A Simple Example

- Toy Universe:
  80% pions, 20% kaons
- We observe a kaon-like signal
- What is the probability that this was an actual kaon?
- Bayes' Theorem

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 – A Simple Example

- The (combined) Likelihood value is proportional to the conditional probability!

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

- And that is all we need!

“Binary” PID vs. “Global” PID

- Binary PID: $$Pid(K,\pi) = \frac{L(K)P(K)}{L(K)P(K) + L(\pi)P(\pi)}$$

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

- Binary PID will yield very misleading results if the true particle is neither of the binary choices…

Summary

- Each detector should provide a (Log-)Likelihood value for all particle hypotheses for each track
  - We combine the information by adding up LLs
- Bayes' Theorem transforms individual Likelihoods into useful probabilities when priors are provided
- "The Physics" dictates priors, but we need a way for analyses to provide priors for the given phase space
  - The ALICE paper describes a very cool scheme to "bootstrap" priors
- To make it easy, the software should primarily provide particle probabilities

FAQ

- "We know detector **X** is not great in phase space **Y**. Should we not add some weight to reduce its impact?"
  - Not needed! Likelihoods are "self weighting".
- What about cases in which the "PDF" is not normalized?
  - … then it's not a PDF
  - Roger Barlow, "Extended maximum likelihood"
  - Executive Summary: one can subtract a term to make it compatible
  - N.b.: I have yet to see a case where this is necessary in PID, will be happy to discuss
- Adding LLs is only "correct" when measurements are uncorrelated
  - True, but a higher order effect. If turns out to be significant: throw it into your favorite ML tool.