

Holistic PID reconstruction for ePIC

Alexander Kiselev

ePIC and EIC Physics Readiness Workshop, University of Calabria, Italy 03/18/2026

Instead of an introduction



ho·lis·tic

/hō'listik/

adjective

characterized by the belief that the parts of something are interconnected and can be explained only by reference to the whole.

"the solution demands a holistic approach and a strategic vision of what can be achieved"

- **MEDICINE**

characterized by the treatment of the whole person, taking into account mental and social factors, rather than just the symptoms of an illness.

Instead of an introduction



ho·lis·tic

/hō'listik/

adjective

characterized by the **belief** that the parts of something are interconnected and can be explained only by reference to the whole.

"the solution demands a holistic approach and a strategic vision of what can be achieved"

- **MEDICINE**

characterized by the treatment of the whole person, taking into account mental and social factors, rather than just the symptoms of an illness.

Previous discussions on this topic

[Path to Holistic reconstruction](#) session in Lehigh in summer 2024

[Belle II talk](#) by Umberto Tamponi

[Parallel PID session](#) in Frascati in early 2025

Recent [discussion session](#) during ePIC Collaboration meeting at BNL

The discussion(s) lately converge to a desire of having a set of Log Likelihood (LL, $\log(\mathcal{L})$) PID subsystem estimates for each charged particle track, *for a set of PDG hypotheses*, which can be combined and used for particle identification together with particle species flux information in particular physics analyses later on

Construct $\Delta\log(\mathcal{L})^{\pi K} = \Delta\log(\mathcal{L})^{\pi K}_{\text{pFRICH}} + \Delta\log(\mathcal{L})^{\pi K}_{\text{hpDIRC}} + \Delta\log(\mathcal{L})^{\pi K}_{\text{BTOF}} + \Delta\log(\mathcal{L})^{\pi K}_{\text{dRICH}} + \Delta\log(\mathcal{L})^{\pi K}_{\text{FTOF}}$

Then use e.g. $N_{\sigma}^{\pi K} = \sqrt{2\Delta\log(\mathcal{L})^{\pi K}}$ as an estimator of πK separation, etc



A short wish list for ePIC holistic PID

- Provide best estimate event-level PID using all available information
- Be based on a reasonable statistical foundation
 - And (ideally) on an existing example implementation
- Avoid a combinatorial overhead, to the extent possible
 - And be computationally affordable, in general
- Provide a unified interface to all PID subsystems
 - Either an output data format or algorithmic part or both
- Have no “threshold” behavior vs momentum, etc
 - If pfRICH provides just a 1σ e/π complementary separation at 4 GeV/c, make use of it

Scope of this talk

- Attempt to outline a practical implementation rather than give a general overview
 - Will be conceptually based on LHCb “global” reconstruction scheme

See the original note by R. Forty [here](#)

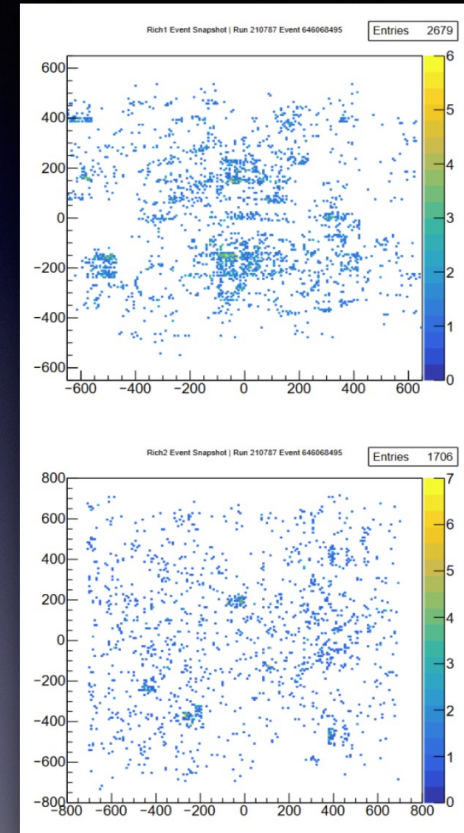
- Charged particle PID only
- No decay particle products with their respective secondary vertices
- Focus on Cherenkov and TOF detectors

- Will be deliberately limited by EICrecon functionality & output
 - *At this stage*, the task is considered solved once EICrecon output for each track has LL estimates for all PID hypotheses of interest

LHCb implementation

Taken from a presentation by C. Jones

- Decode RICH data, forming 3D **hit detection points** in the LHCb global reference frame.
- Take reconstructed track objects and determine their trajectories through the RICH radiator volumes. Primarily for assumed **photon emission point**, but also provides path-length and **impact point on photon detector plane** (raytracing track as if it were a photon).
- Determine, for each Track and mass hypothesis the **expected Cherenkov theta values and resolutions**.
- Reconstruction **candidate Cherenkov photons** (i.e. Theta/Phi angles) by combining the hit detection points with the assumed emission points, with the knowledge of the RICH mirror optics.
- For each reconstructed photon candidate, determine for each mass hypothesis the **expected observable pixel signal yield**.
- Perform a **global event likelihood maximisation**, that iteratively changes the mass hypothesis of each track to find the overall set for the event that is the most likely. Provides the final track mass hypothesis PID information.



LHCb implementation

Taken from a presentation by C. Jones

- For each track in the event, and for a given mass hypothesis, create Cherenkov photons and project them to the detector plane using the knowledge of the geometry of the detector and its optical properties.
- Repeat this for all the other tracks.
- From the photon distribution on the detector plane calculate the probability that a signal would be seen in each pixel of the detector from all tracks.
- Compare this with the observed set of photoelectron signal on the pixels, by creating a likelihood.
- Repeat all the above after changing the set of mass hypothesis of the tracks.
- Find the set of mass hypotheses, which maximize the likelihood.
- Provide for each track delta log likelihood values, that give the overall change in event likelihood changing that track to that given type.

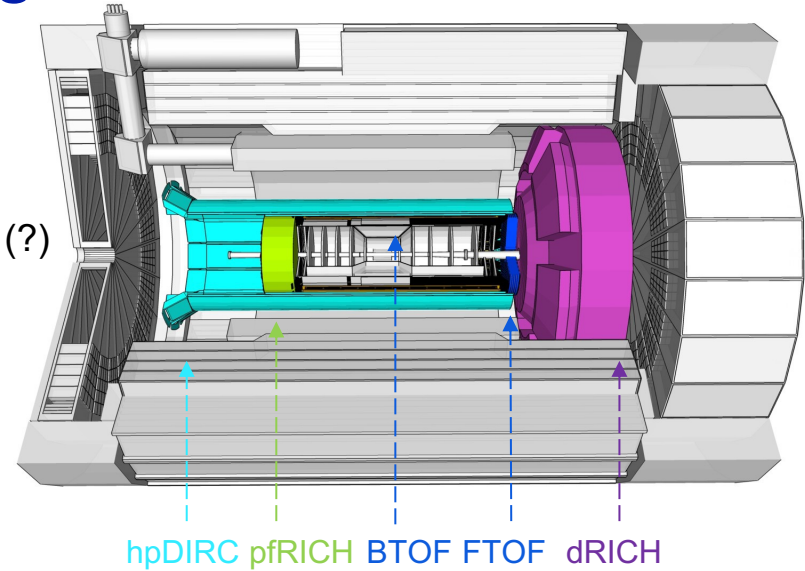
LHCb implementation and its possible use in ePIC

- Includes a combination of direct and inversed ray tracing
- Is obviously designed for a very challenging environment
 - And works fine at these high track and hit densities
- Does not provide a hit-to-track assignment as a final result

- Does not seem to include timing
 - At least in its original formulation
- One may argue, that in a multi-path optical configuration (pfRICH, hpDIRC), the DRT role can be further enhanced to building 1D image contours (“distorted rings”) in the sensor plane as proxies of respective PDFs (probability density functions)
 - And a local deformation of these contours: azimuthal stretching (normalization) and transverse broadening (apparent spatial resolution) at the vicinity of hits contains all the information necessary to evaluate LL contributions

ePIC PID detector subsystems

- Provide full acceptance coverage
- Some of them overlap in acceptance
 - FTOF & dRICH, BTOF & hpDIRC, pfRICH & hpDIRC (?)
- pfRICH and hpDIRC will likely be using the same photosensors
 - Same / similar digitization scheme
- Some of them are in fact “two detectors in one”
 - dRICH (aerogel & gas radiators), pfRICH (imaging and timing capability)
- Time-of-flight performance will depend on a best possible t_0 determination
 - And it better be the same (global) quantity for all subsystems for a given event



+ e/m calorimetry

Further observations

- Our track multiplicities will be “small”
 - Typically, a handful of tracks in each of the forward / central / backward acceptance at most -> nothing close to the LHCb pictures shown on slide 7
- A high-resolution timing capability in 4π is a distinguishing feature of ePIC
 - And, up until recently, has not been such for either LHCb or Belle II
- Software-wise, we know that pFRICH and dRICH can be unified
 - Same is true for the two TOF subsystems
- In a DRT formulation, there may be not that much difference between hpDIRC and pFRICH
 - After all, it is all about constructing a PDF distribution on a similar sensor plane

Preliminary conclusions / suggestions

- May be tricky to implement a $\Delta\log(\mathcal{L}) = \Delta\log(\mathcal{L})_{\text{pfRICH}} + \dots + \Delta\log(\mathcal{L})_{\text{FTOF}}$ like scheme for ePIC
 - [and not necessary perhaps?]
 - Several subsystems depend on a highest-quality global timing reference
 - It is not available before the PID assignments are made
 - It may not be easy / wise to retain a raw timing-related information for iterations
 - Just a single LL number does not allow one to reverse engineer the same LL estimate for a changed t_0 (although one may consider retaining $d\{\Delta\log(\mathcal{L})\}/dt$ derivatives for iterations?)
- ePIC charged track multiplicities in 4π are small
 - Certainly not a showstopper for using *all subsystems at once* for PID assignments

Preliminary conclusions / suggestions

- t_0 estimate provided by beam optics and vertexing may or may not be sufficient to perform PID assignments on a per-subsystem basis first, and iterate later
 - Even if it is, the core PID algorithm across a subset of subsystems can be the same
- If it turns out that e.g. a dRICH is marginally sensitive to t_0 (and, most importantly, does not contribute much to its determination), it can of course be factorized out
 - Even in this case, the core PID algorithm for dRICH can be the same

Assumptions

- Tracking and vertexing steps have been performed already
 - N tracks with a common vertex are provided as an input for PID code
 - Projections and path length estimates are available (and we ignore their errors, for now)
- With an added beam optics constraint, a $\{\mathbf{V}_{IP}, t_0\}$ is always available
 - With a 4x4 covariance matrix, even for single-track events
- Digitization step has been performed already for all subsystems
 - And is sufficient for matching with PDFs to provide meaningful LL calculations

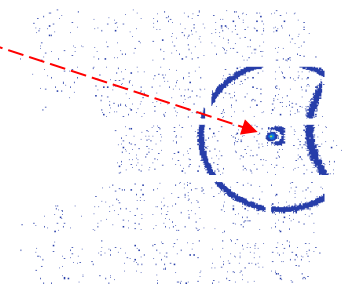
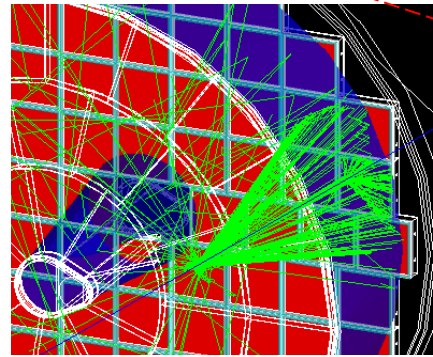
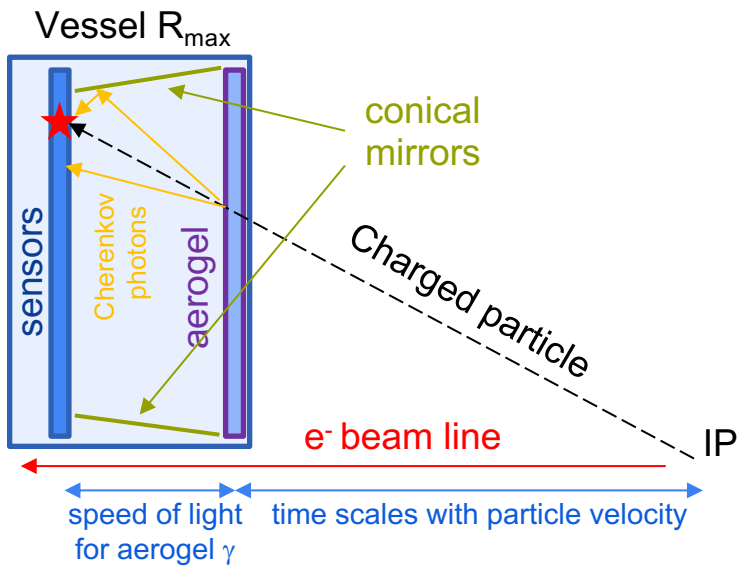
Assumptions

- We are interested only in $\{e, \mu, \pi, K, p\}$ particle hypotheses evaluation
- Calorimetry *optionally* provides combined ML estimates for these hypotheses
 - For instance, to help reducing a combinatorial overhead
- Measurement errors are considered Gaussian
- Backgrounds and noise are assumed “static” (do not vary from event to event)

Some of these are just to present a clear case; can be (easily?) lifted later

pfRICH case as a testbed

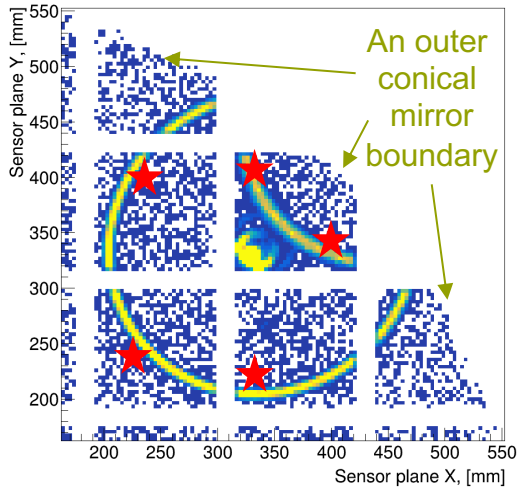
- Perhaps the trickiest one
 - Same aerogel Cherenkov photons used for imaging and timing
 - Multiple optical path combinations for any given pair of emission and detection points
 - Timing also provided by a Cherenkov photon flash where a charged particle passes through the HRPPD window



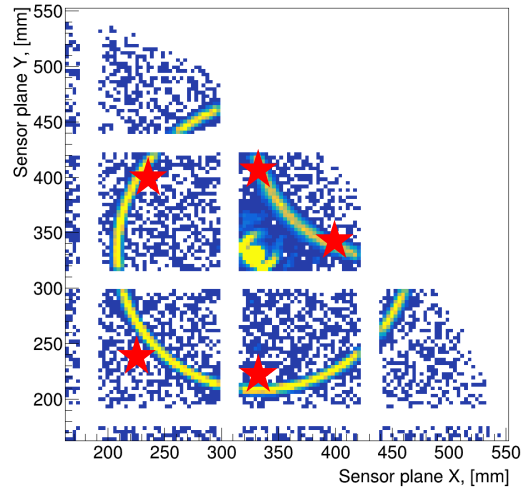
pfRICH case, aerogel photons

- For each track in this acceptance, and for each of the PID hypotheses of interest, create a PDF map of where the Cherenkov photons can land
- Either analytically (?) or via a (fast) simulation or from a lookup table (?)
- This step takes care of a multiple optical path issue in IRT (but also of two separate imaging radiators in case of the dRICH)

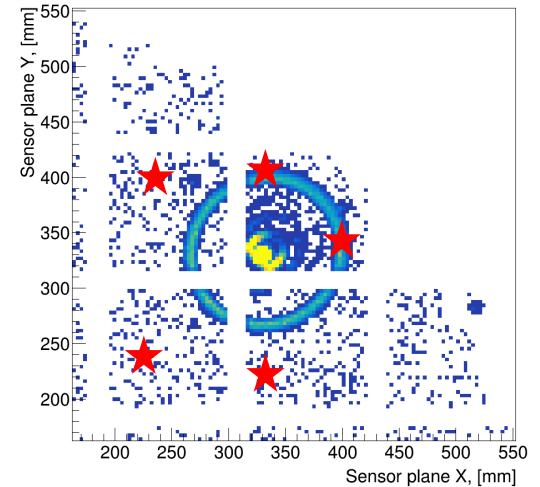
★ Hits from a single 2 GeV/c track



positron



pion

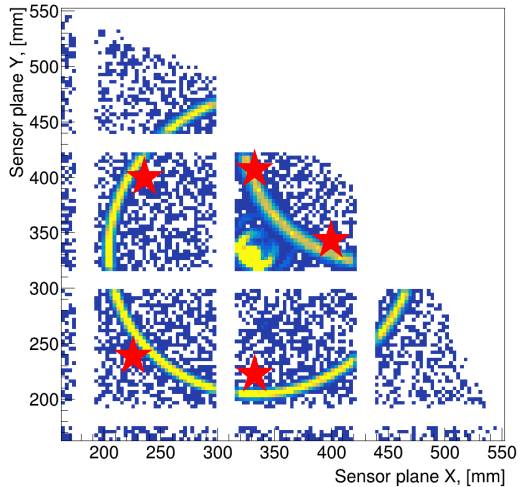


kaon

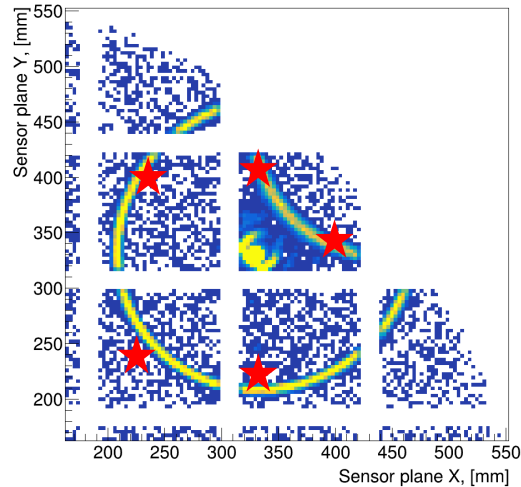
Here: 10k events with the same kinematics for each of the three hypotheses used to generate “PDF maps”

pfRICH case, aerogel photons

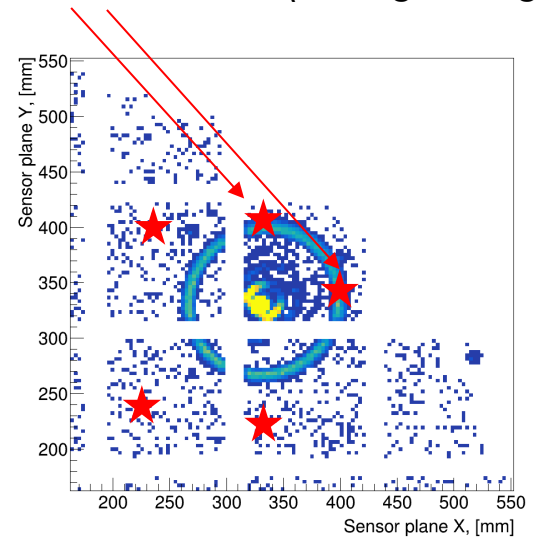
- Judging by eye, a pion hypothesis fits the hits “a bit better” than a positron one
 - *By how much exactly better?*
- Kaon hypothesis is a clear outlier
 - *And (if the track was in fact a pion) it is not getting these two hits either (wrong timing)*



positron



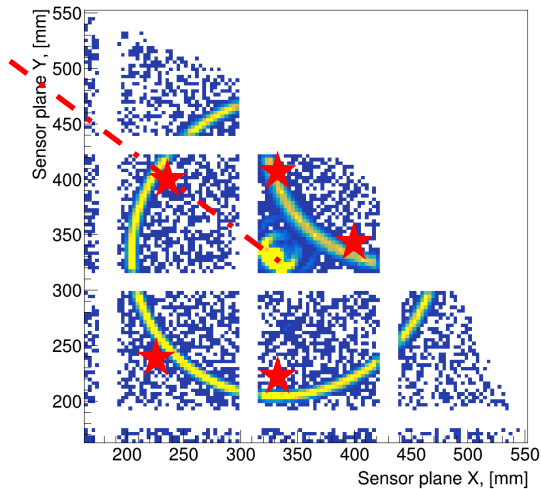
pion



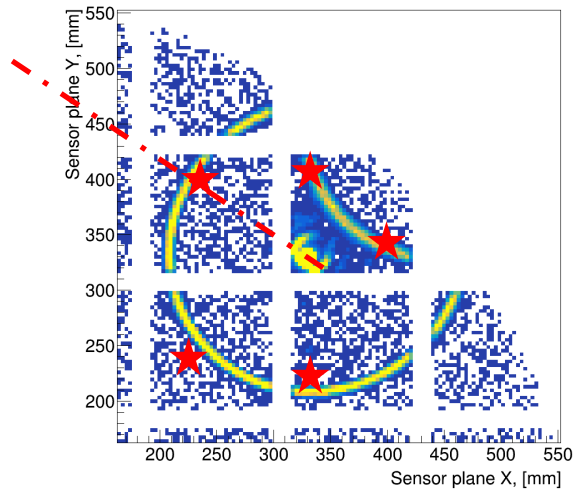
kaon

pfRICH case, aerogel photons

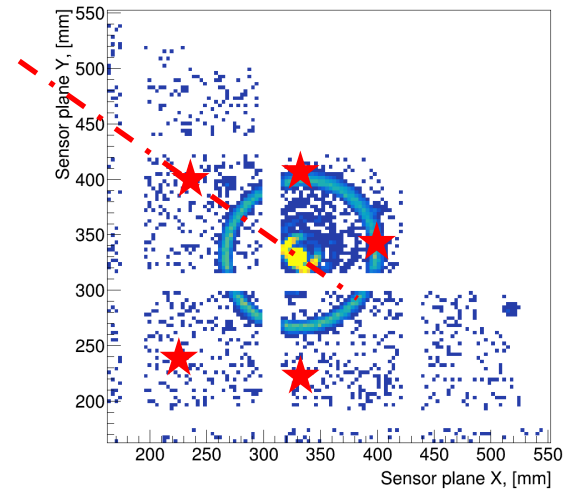
- What matters for a "good match" is a *radial* hit displacement with respect to the 1D PDF contour in space (along the red lines shown for one of the hits) and a timing difference, in a metric defined by a contour width and a hit resolution
- In a Gaussian approximation, these are sufficient to build a $\log(\mathcal{L})$ quantity (?)
- *No hit-to-track assignment is required; just compare $\{e, \pi, K\}$ $\log(\mathcal{L})$ values*



positron



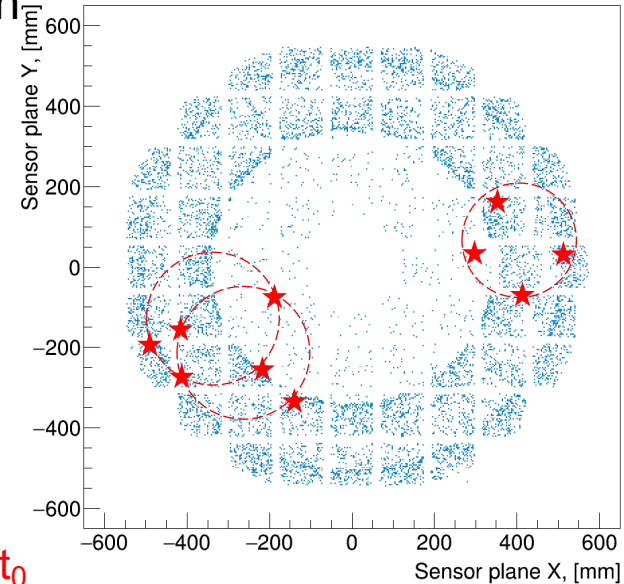
pion



kaon

Algorithm logic for a given detector subsystem

- For each track, select hits within a predefined “road” in space *and time* from a generated PDF contour for at least one particle type hypothesis
- Identify groups of tracks sharing the same hits
 - May then in principle deal with each group separately
- Assign initial PDG types for every track
 - Use EmCal info, guess based on TOF, ask AI, etc
- Iterate minimizing overall LL by changing PDGs
 - But if all PDGs are fixed in every iteration, why not *re-fit* a t_0 estimate every time new, along with changing the PDGs?
- And if so, why not doing this for all track groups of this detector subsystem (pFRICH) *at once*?



Two apparently disjoint groups of tracks & hits in pFRICH

Hit-to-track association for a given PDG hypothesis

edm4eic::PIDSubsystemHitToTrackAssociation:

Description: "A PID subsystem hit coupling to a PIDSubsystemTrack under a given PDG hypothesis (assumes Gaussian errors)"

Author: "Joe PID Writer"

Members:

- float path1 // particle trajectory length before emitting a photon (time scales with velocity)
- float path2 // Cherenkov photon trajectory length (if any; normalized to media <n>'s)
- float dr // spatial component of a hit distance to a PDF contour
- float dt // hit timing difference to a PDF contour
- edm4eic::Cov2f covariance // [dr,dt] covariance matrix (diagonal, except for perhaps hpDIRC)

OneToOneRelations:

- edm4eic::PMTHit hit // relationship to some standard digitized hit structure

**Should just contain enough information to recalculate
LL-terms at a different t_0 during iterative process**

The rest of the data model

edm4eic::PIDSubsystemTrackPDGHypothesis:

Description: "A PDG hypothesis info for a PIDSubsystemTrack"

Author: "Joe PID Writer"

Members:

- int32_t PDG // PDG code hypothesis
- float norm // some normalization reflecting $\langle N_{pe} \rangle$ (?)

OneToManyRelations:

- edm4eic::PIDSubsystemHitToTrackAssociation hits // hits associated with this PDG hypothesis

edm4eic::PIDSubsystemTrack:

Description: "A PIDSubsystemTrackGroup member track"

Author: "Joe PID Writer"

OneToOneRelations:

- edm4eic::Track track // charged particle track

OneToManyRelations:

- edm4eic::PIDSubsystemTrackPDGHypothesis PDGs // PDG hypothesis information

If we got that far, note that it is nowhere written that these two track / hit groups belong to the same detector subsystem (pfRICH in this case)

edm4eic::PIDSubsystemTrackGroup:

Description: "A group of tracks with shared hits in a PID subsystem"

Author: "Joe PID Writer"

Members:

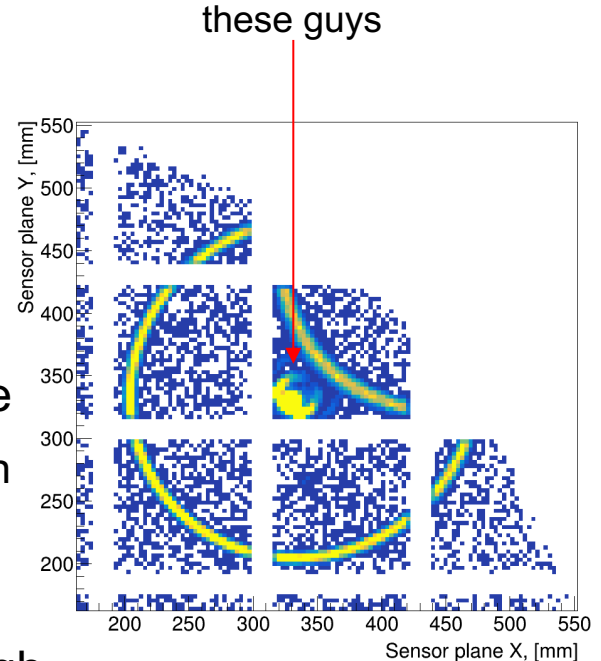
- float nsigma // n-sigma cut applied to hit-to-track associations

OneToManyRelations:

- edm4eic::PIDSubsystemTrack tracks // charged particle tracks with their hit associations

pfRICH case, window flash photons

- Well, they are kind of similar to the imaging ones
 - One can create a PDF map where to expect them
 - A group of more or less independent hits produced
- A point-to-point rather than a point-to-contour distance
 - Not a big deal to account for, even in a simplified Gaussian error based scheme
- Need to maintain a different path length scheme though
 - See path1 and path2 fields on slide 21, as well as slide 14



Seems to be covered by the same algorithm?

dRICH case

- Two radiators should not be a problem
 - If gas and aerogel rings can overlap for a given track, may need to keep both options
- Spherical sensor plane is not a problem either
 - Will still result in a 1D PDF contour in a 3D $\{x,y,t\}$ space
- A long gas radiator may spoil timing for these photons entirely
 - But timing should still work for aerogel

- Will dRICH be able to contribute to a t_0 determination?
 - Well, $\sim 100\text{ps} / \sqrt{15}$ is pretty much in the right ballpark
- Alternatively, can treat dRICH as an independent subsystem

Yet seems to be covered by the same algorithm

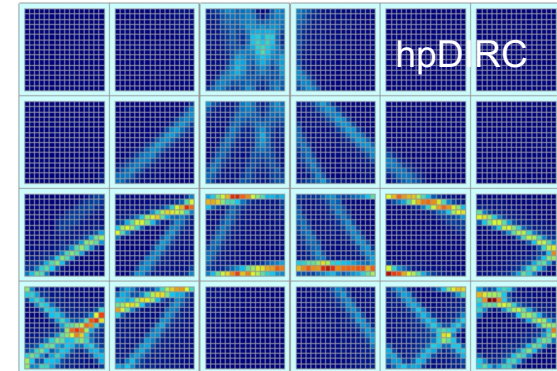
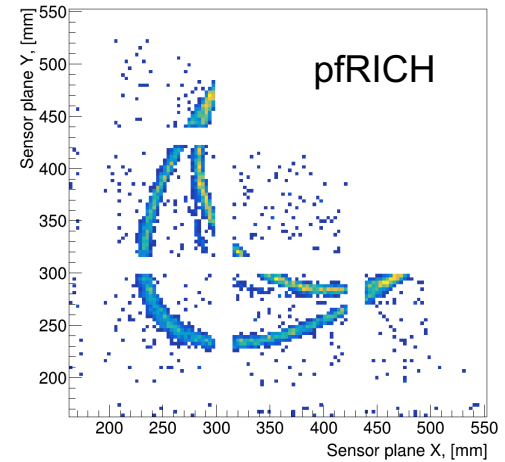
BTOF / FTOF case

- Let's forget that TOF hits have been associated with tracks already
 - Perhaps it is a good idea anyway (PID determination a la LHCb does not require hit-to-track association; no need to resolve a possible ambiguity for tracks in a close proximity to each other at the BTOF / FTOF location)
- In this case, there is no difference (?) with pFRICH HRPPD window photon flashes
 - Same type of a track-related PDF, pixels associated with it, etc

Can be covered by the same algorithm (?)

hpDIRC case

- Well, we all know that DIRC reconstruction is tricky and special ...
- ... but it's tempting to claim that the case (in DRT formulation) is similar to pfRICH
- A familiar track projection on input, and a 1D PDF contour on a similar sensor matrix
- hpDIRC picture looks more frightening, but *it also has a third dimension (time)*
- A possible space-time correlation in a PDF contour



Can potentially be covered by the same algorithm (?)

Instead of a summary

The end goal of such a development

- For a PID assignment, treat all ePIC Cherenkov & TOF detectors in a same (or a very similar) way in EICrecon
- *Per default, fit a global event t_0 together with a simultaneous PID assignments to all charged particle tracks across all PID subsystems in the event at once*
 - In addition, for each track give $\log(\mathcal{L})$ estimates for other four PID hypotheses
- Be able to optionally handle any subsystem (or any group of tracks) independently if needed
 - Using perhaps an AI or a different algorithm
- Seamlessly account for contributions of all PID subsystems, even outside of their “nominal” kinematic range
- Account for calorimetry info, in some way
- Be able to retrieve $\log(\mathcal{L})$ contributions of individual subsystems *a posteriori*
 - Just by fixing a t_0 value to its optimal value and re-running for a subset of detectors?

Output ElCrecon format

edm4eic::PIDOutput:

Description: "An event-level PID output for a single charged particle track"

Author: "Joe PID Writer"

Members:

- int32_t	PDG	// reconstructed (most likely) PDG code
- float	electron	// electron LL
- float	muon	// muon LL
- float	pion	// pion LL
- float	kaon	// kaon LL
- float	proton	// proton LL

VectorMembers:

- uint16_t	tracks	// bitmask of tracks used
- uint16_t	subsystems	// bitmask of detector subsystems used

OneToOneRelations:

- edm4eic::Track	track	// charged particle track
------------------	-------	---------------------------

➤ Can be extended to N-particle groups if needed