FARICH+AI
Experience with RECO/ID/FASTSIM

CEPC Workshop Nov 5-10, 2025

Fedor Ratnikov

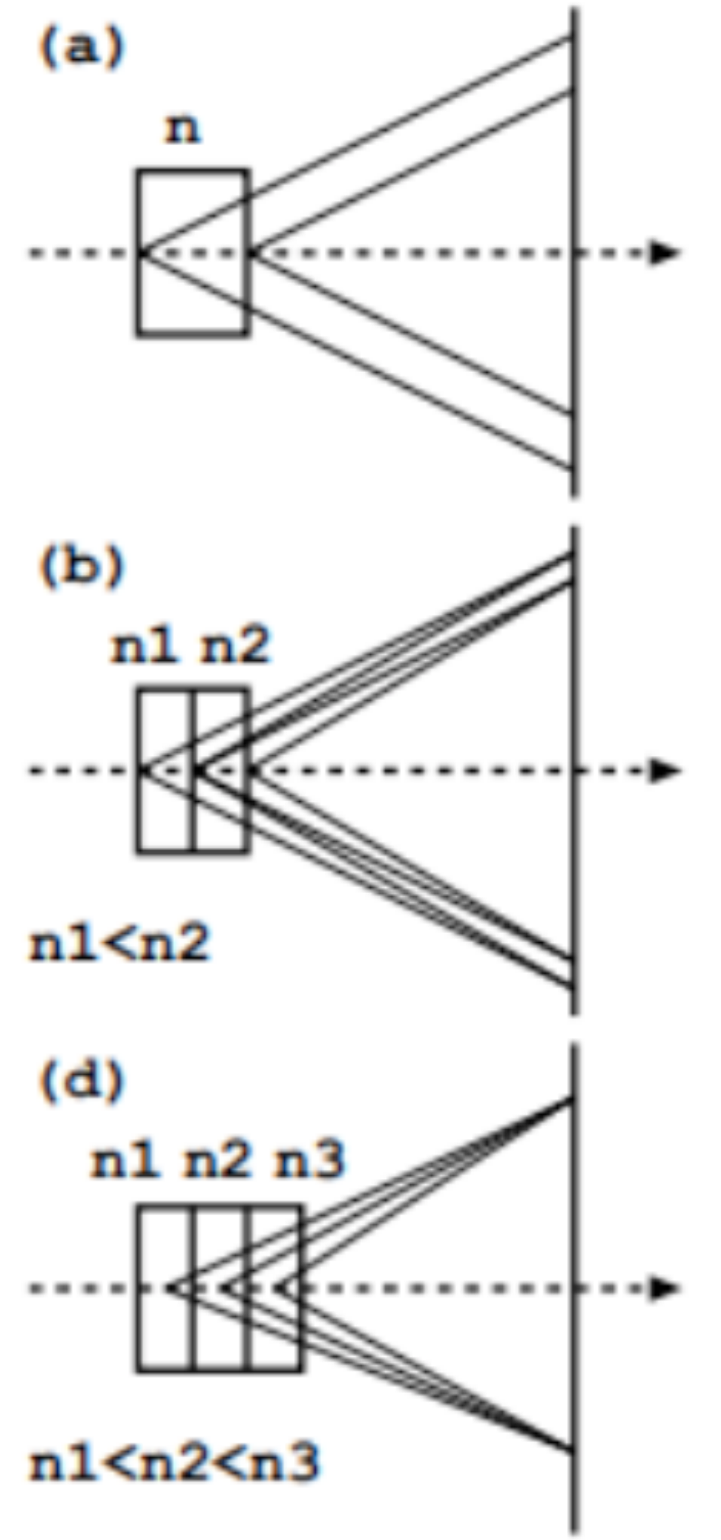
HSE University, Moscow RU

LAMBDA • HSE

What is FARICH?

Focusing Aerogel RICH

- combines benefits of
  - compactness of proximity RICH
  - higher photon yeld of optical RICH



Nice Stuff Produced in Novosibirsk

Barnyakov et.al.
https://doi.org/10.1016/j.nima.2019.05.088

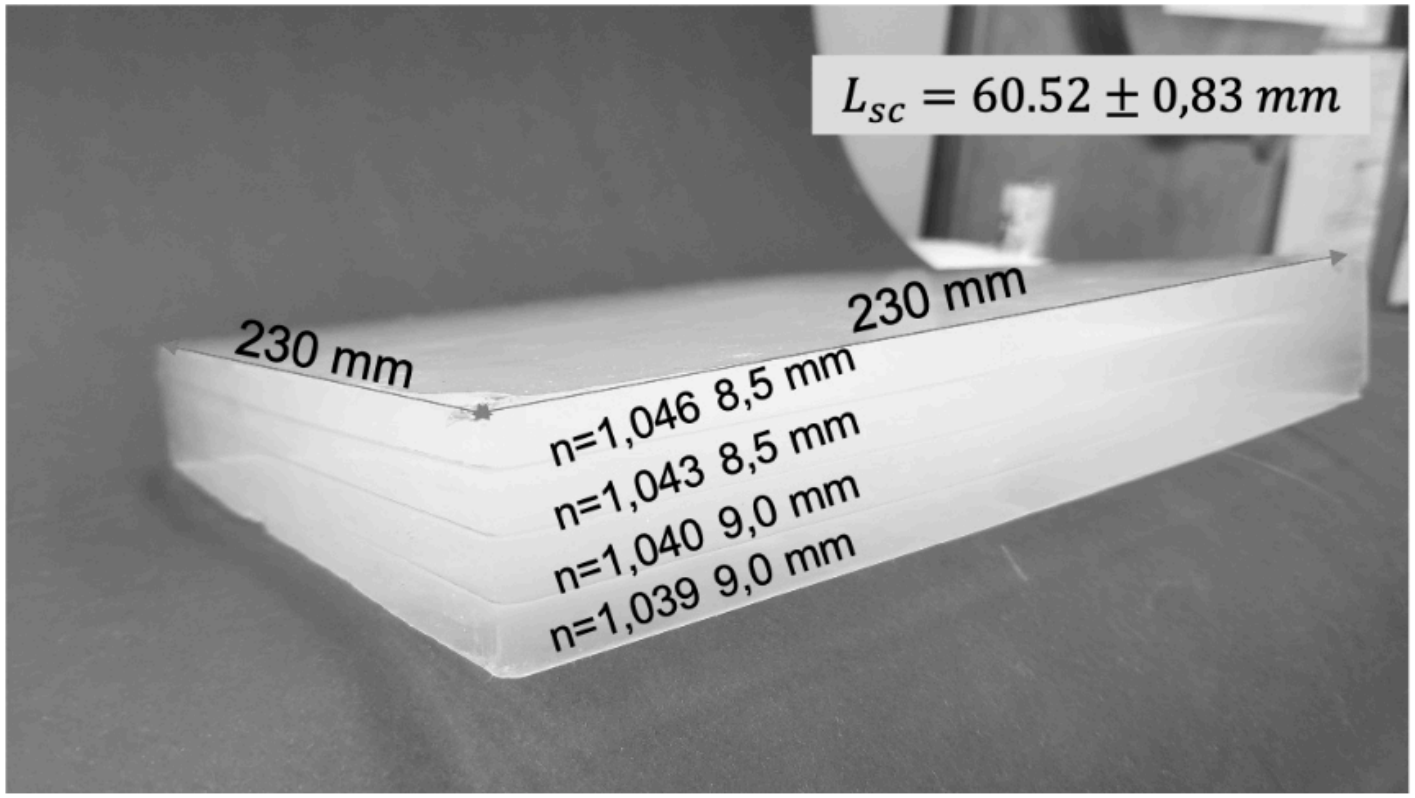


Fig. 2. The largest four-layer focusing aerogel block.

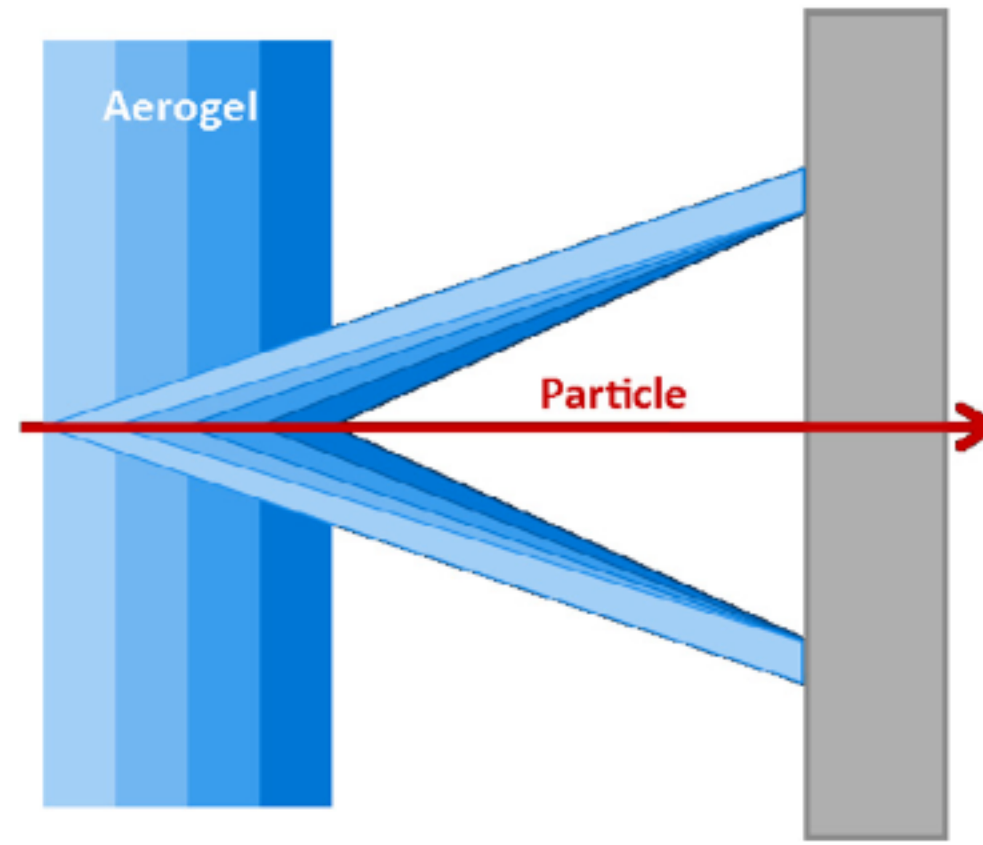
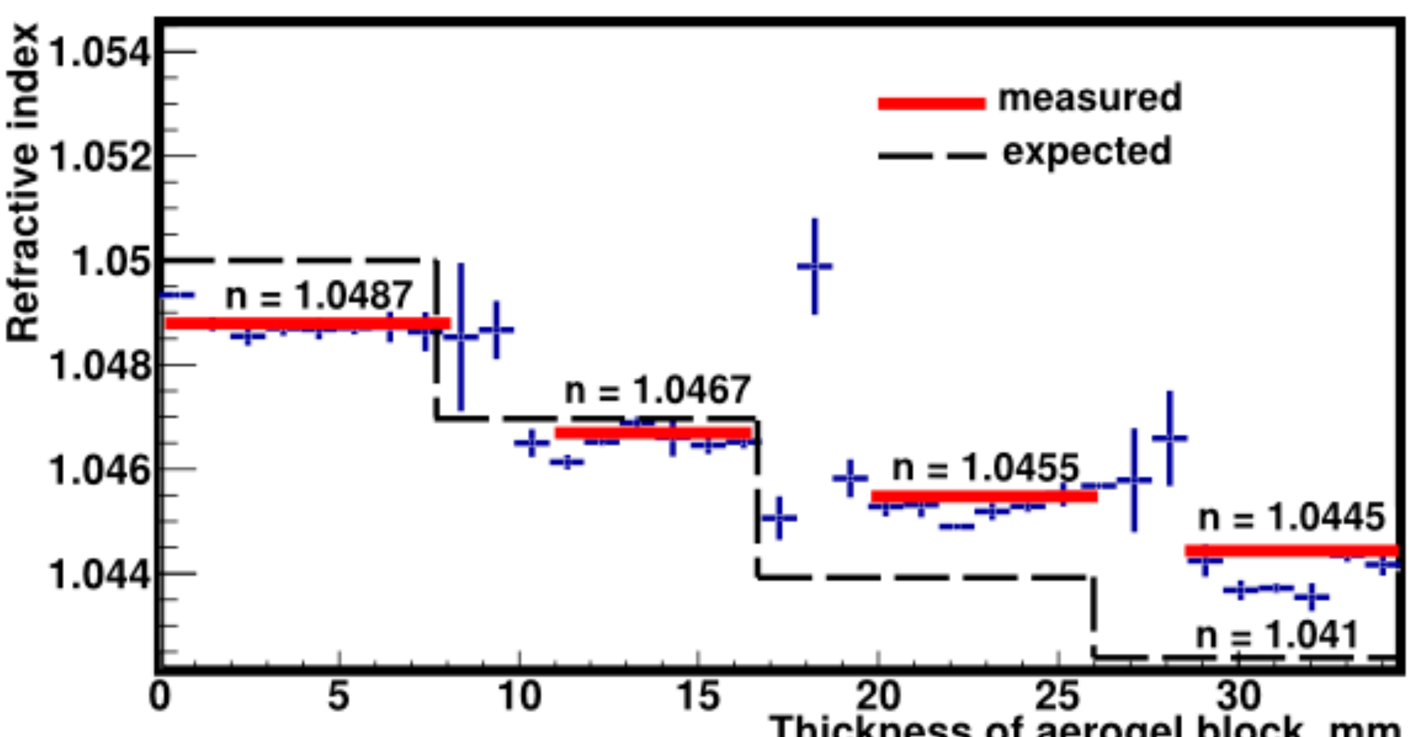


Fig. 1. The Focusing Aerogel RICH (FARICH) concept.



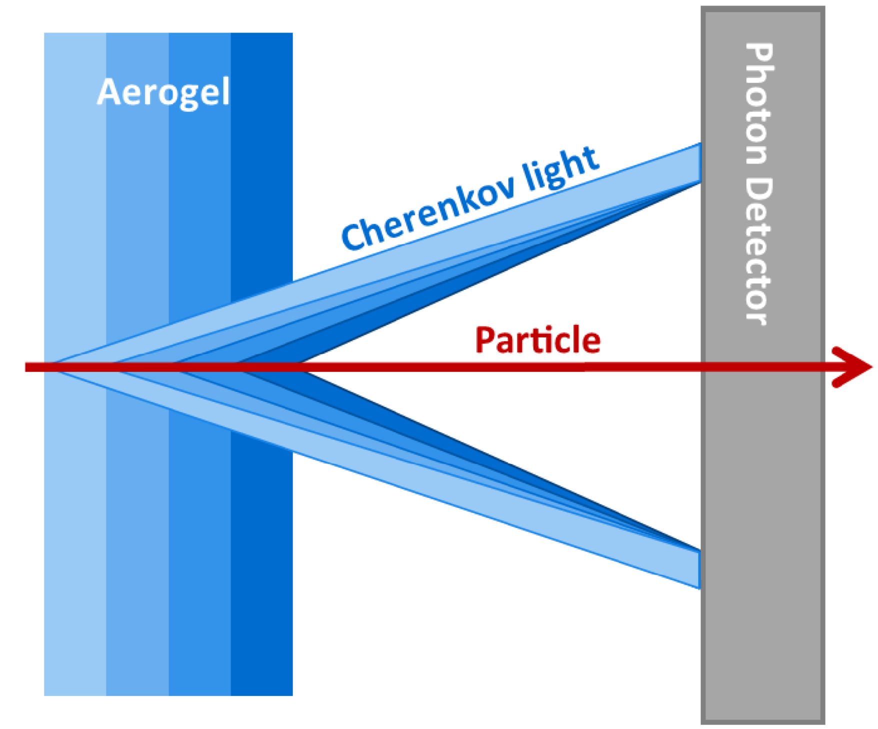
What Can We Do With FARICH

Possible tasks:

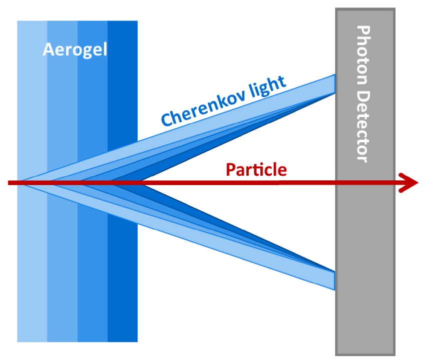
- identify signal presence
  - unseeded, at online

    do it faster than tracking
- reconstruct β (regression)
  - seeded, offline

    use for particle ID
- particle ID (classification)
  - $\mu^{\pm}/\pi^{\pm}/K^{\pm}$ separation
- fast simulation
  - bypass Cherenkov photons generation and propagation



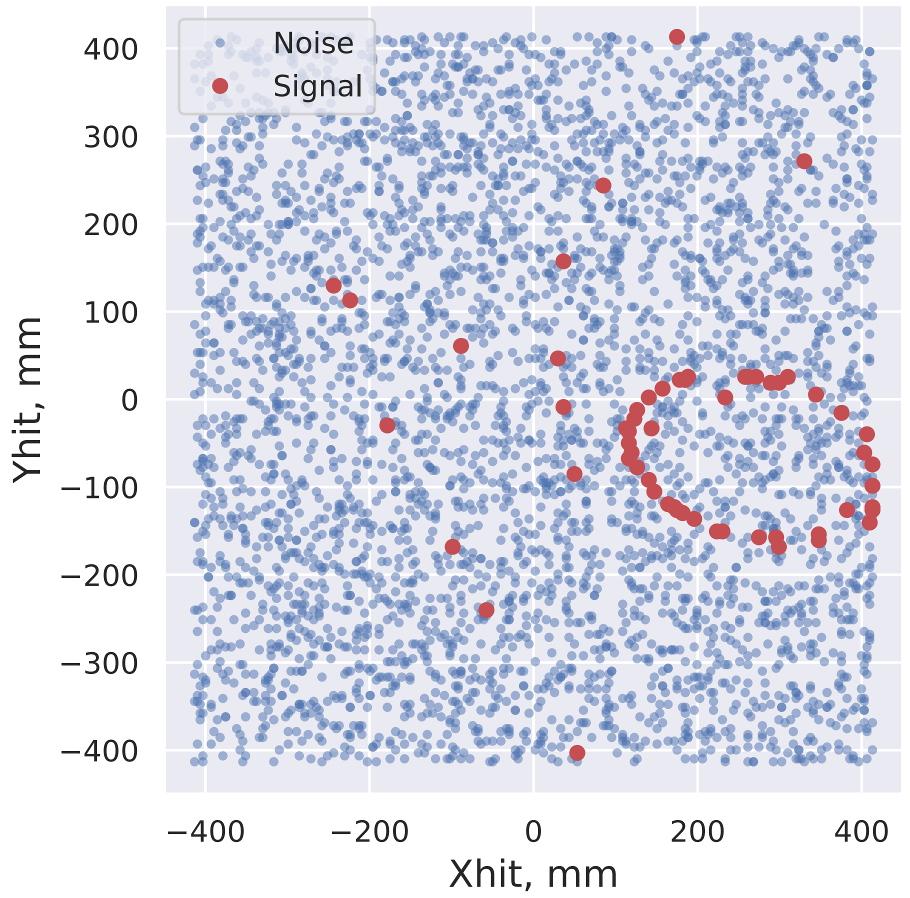
Problem: Sensors Noise

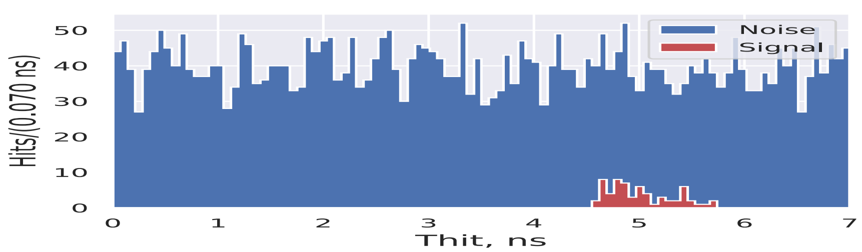


Tight space requires compact sensors

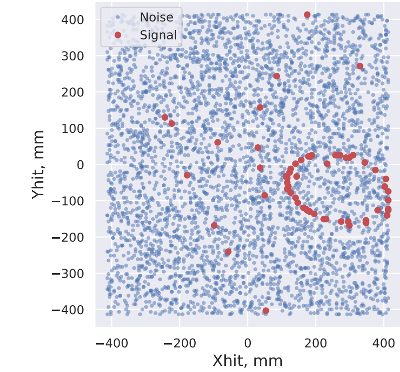
SiPM are noisy

- $10^6\ Hz/mm^2$ without cooling
- $10^5\ Hz/mm^2$ with moderate cooling
- $10^4\ Hz/mm^2$ with agressive cooling





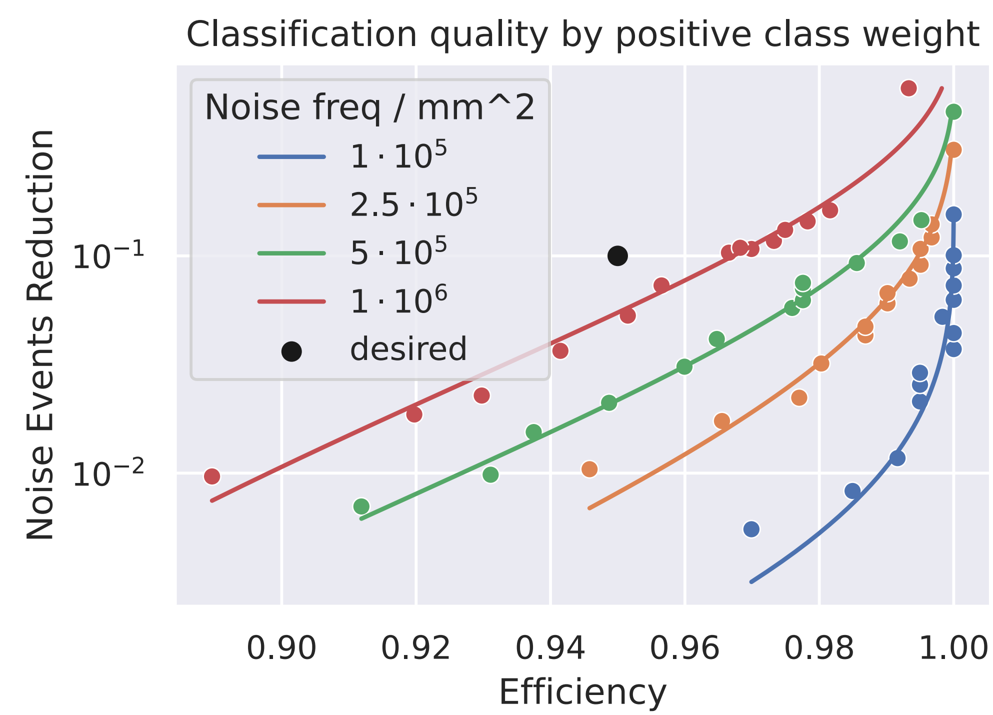
#1: Identify Signal Presence



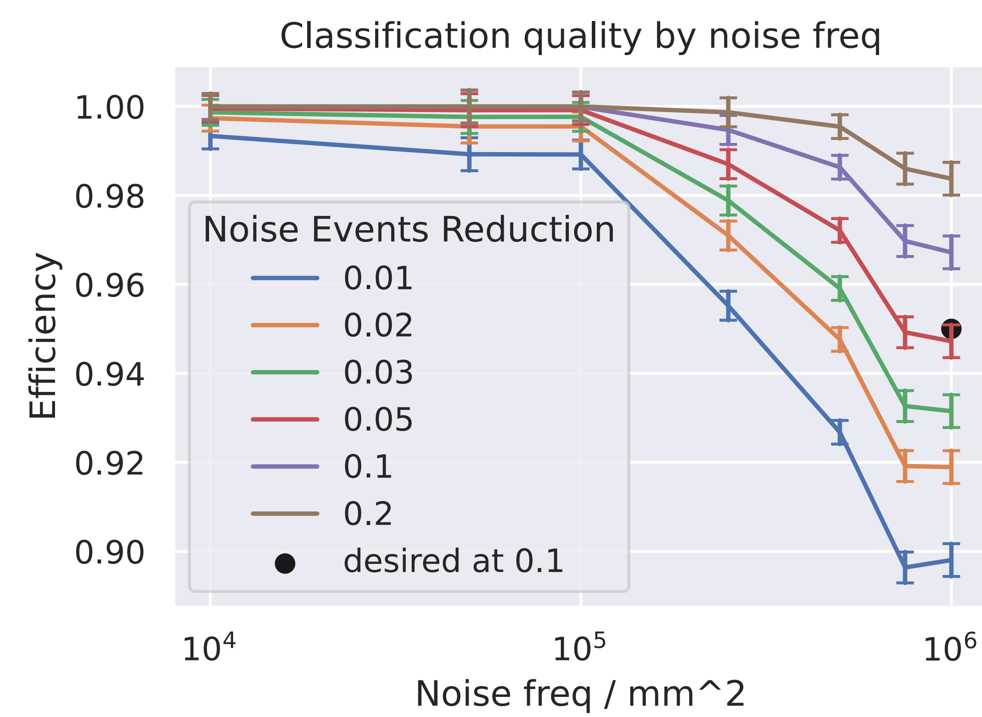
Project hits on the regular grid

Apply computer vision to recognize ring pattern

- ResNet-18 CNN (*) for binary classification
  - does or doesn't exist a signal in the given bounding box?
- No track seed is used
  - assume it is not available at this level

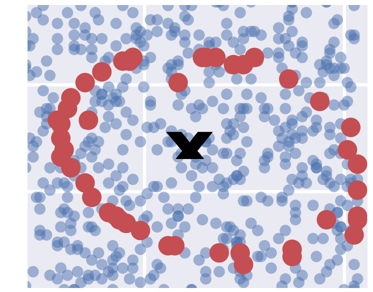


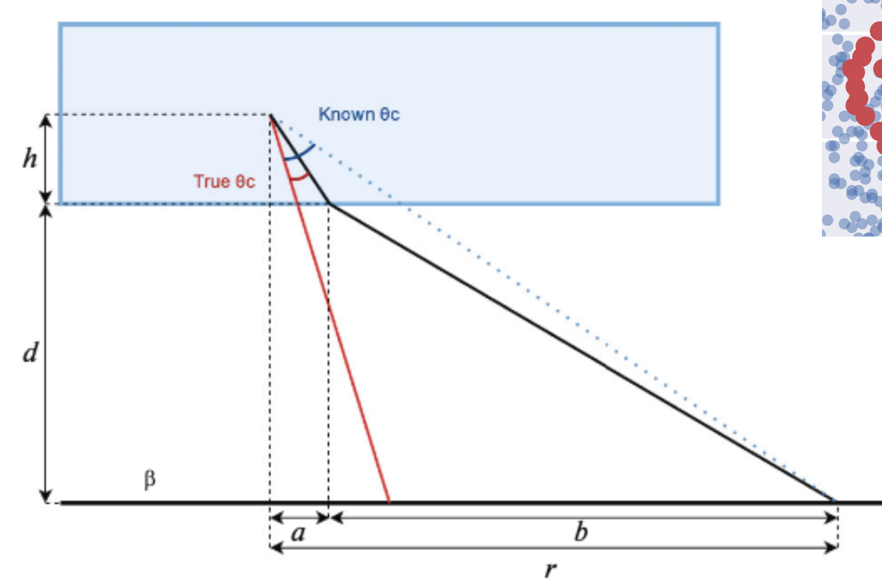
Can achive target factor of 10 background suppression @ $\varepsilon \sim 95\ \%$



(*) ResNet-18 is a specific type of neural network that uses "shortcuts" to learn much more effectively, especially for complex tasks like recognizing objects in images.

#2: Reconstruct Signal Once Identified



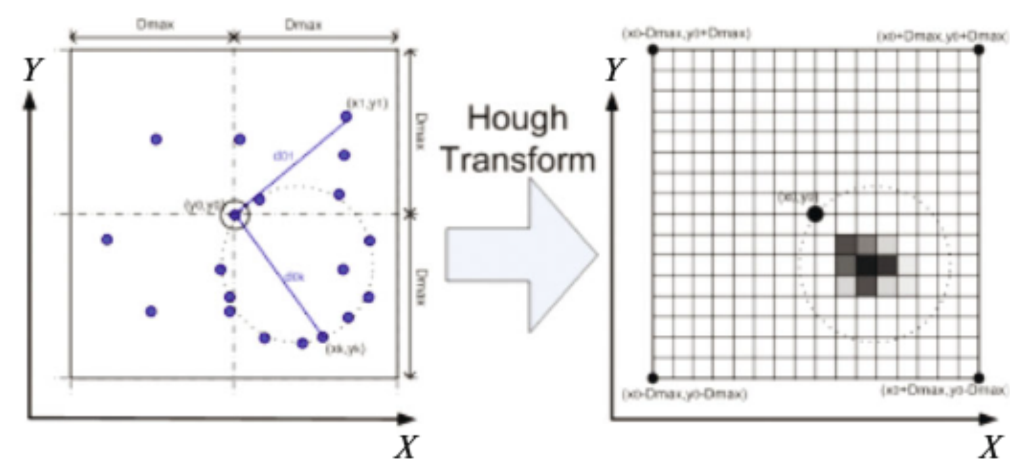


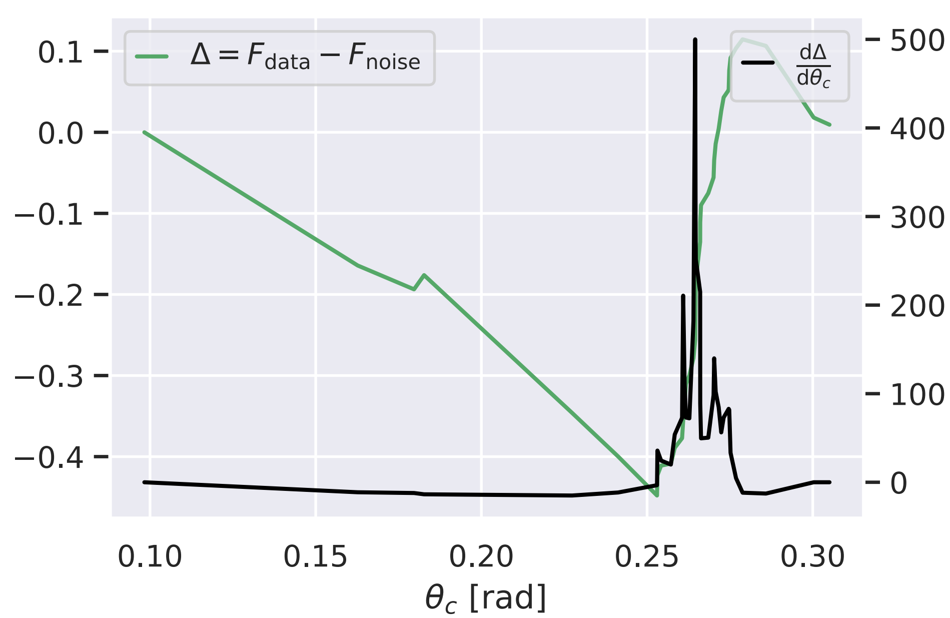
Now have signal seed from track

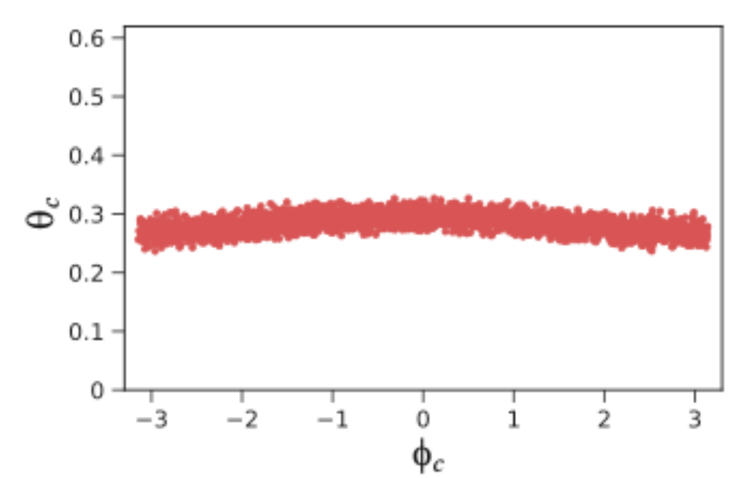
Still project hits on the grid

- Baseline a la Hough transform

- Apply refraction correction

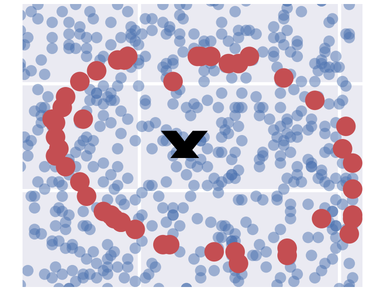




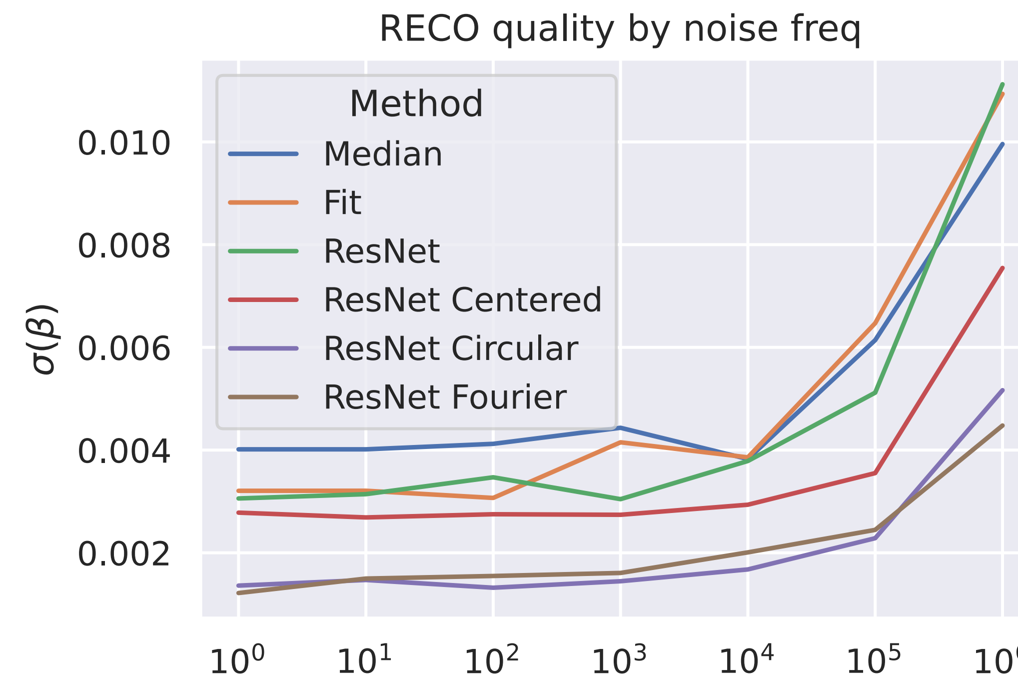


#2: Reconstruct Signal Once Identified

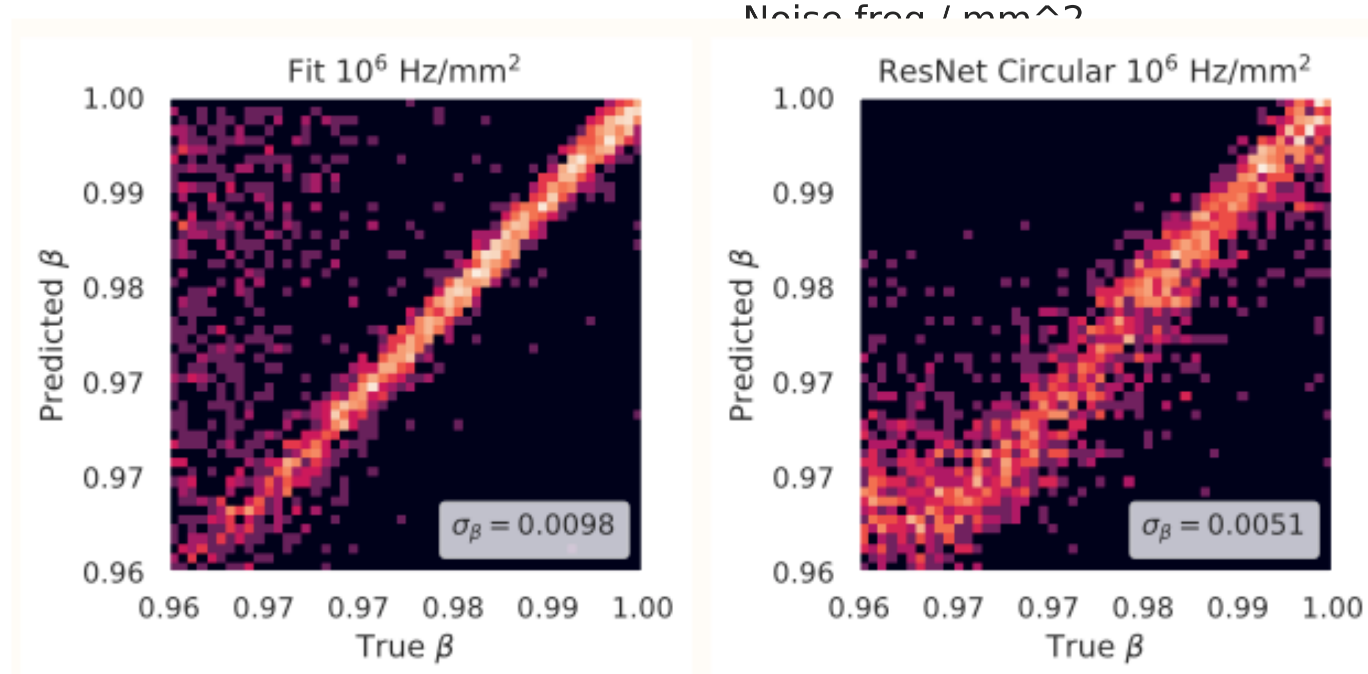
Now have signal seed from track



Still project hits on the grid



- ResNet-18 CNN for β regression
  - pre-process data using track prior

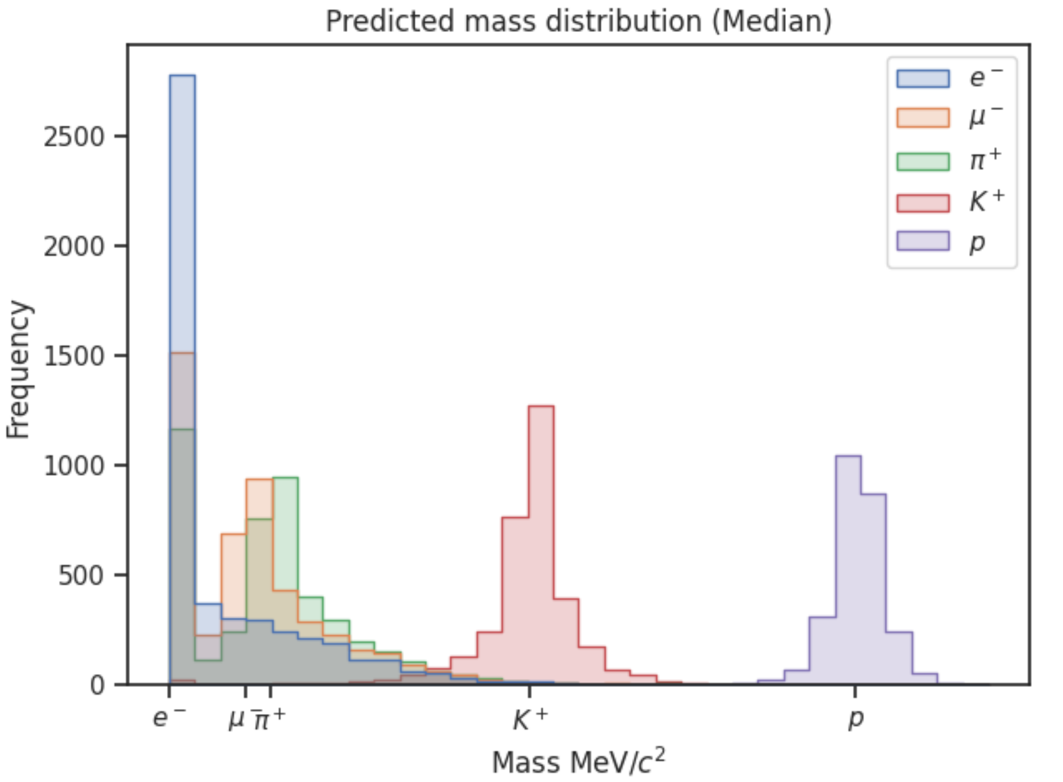


Can achive reasonable β resolution

#3: Classify Signal - Particle ID

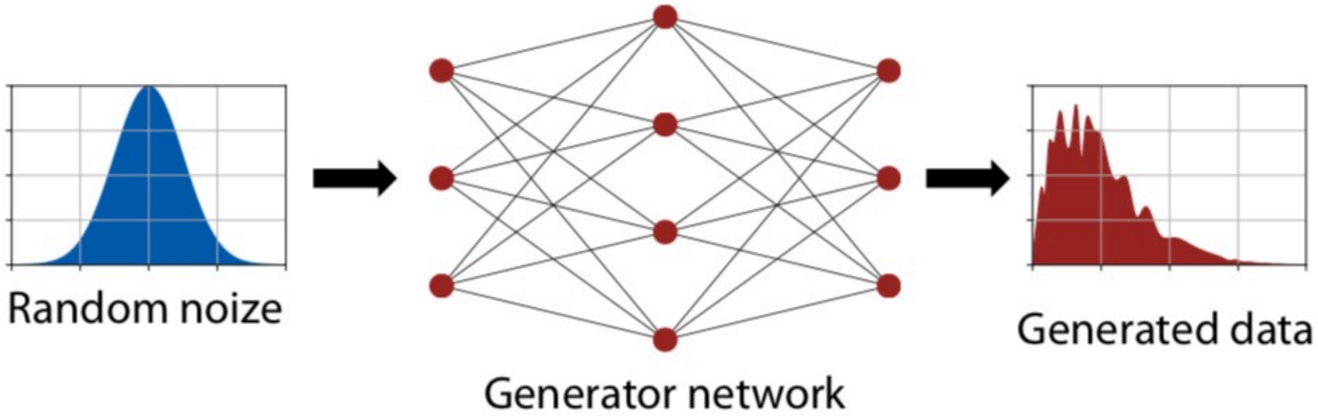
Reconstructed θ_c may be converted to β, which may be converted into m, thus separating particle types

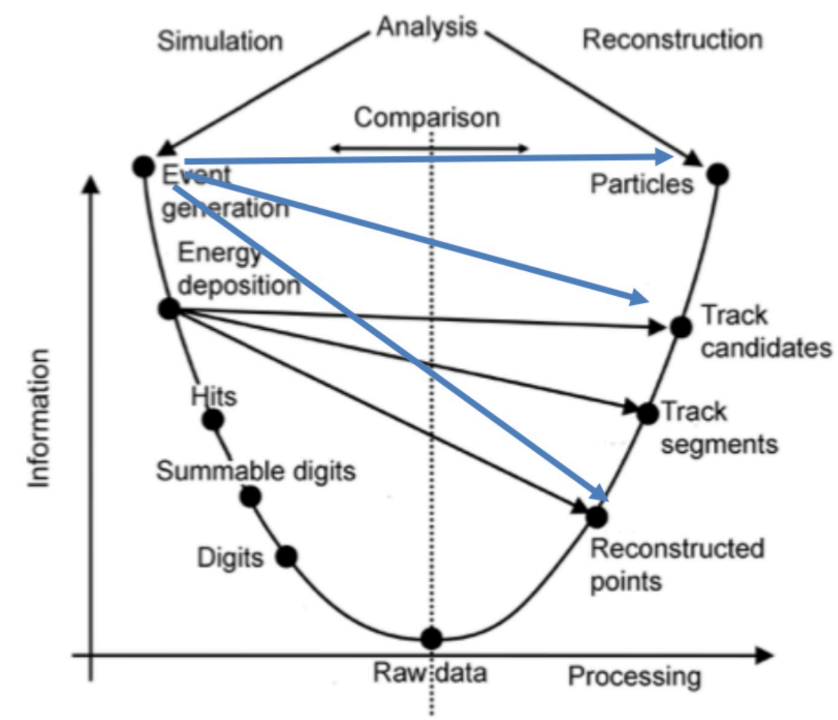
- this chain hard to avoid for classic methods
  - hit map→reco(θ_c)→reco(β)→reco(m)→ PID
- from ML perspective this is a classification problem
  - hit map→PID (multiclass)
  - this allows to cut corners and get better performance



Method	$\mathbb{P}$(true π, pred K)	$\mathbb{P}$(true K, pred π)	π, K AUROC	Total accuracy	σ_β
NN	$\mathbf{0.016 \pm 0.005}$	$\mathbf{0.010 \pm 0.005}$	**0.997**	**0.65**	**N/A**
Median	0.06 ± 0.02	0.02 ± 0.02	0.989	0.64	0.0008
MLE	0.20 ± 0.02	0.12 ± 0.02	0.883	0.52	0.0018
Hough	0.13 ± 0.02	0.26 ± 0.02	0.817	0.49	0.0062

Fast Simulation





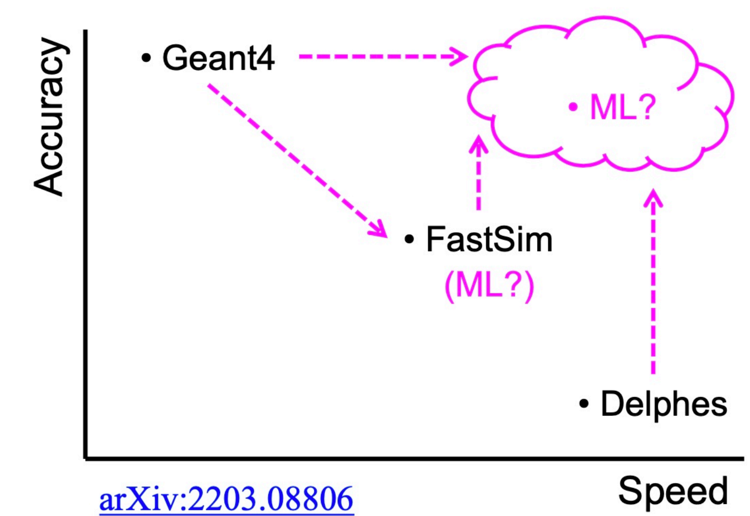
mia parabola" Figure by Federico Carminati, independent parallel inventions by Vincenzo Innocente & K.C.

GEANT is Generative Model

- converts few initial particle parameters into macroscopic detector response
  - processed on microscopic level - pretty intensive calculations

FARICH simulation includes

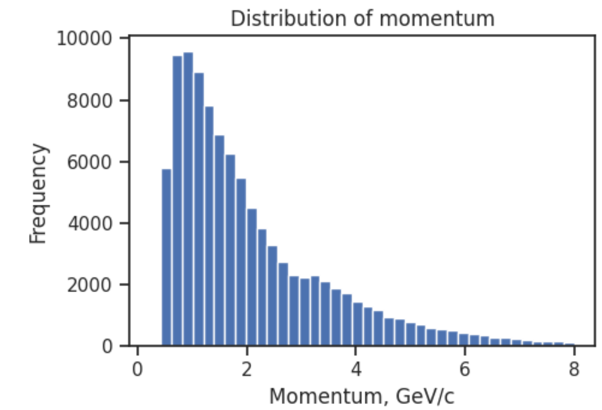
- passing particle through radiator
- cherenkov photons emission
- photons transport, reflection, refraction, diffusion, …
- (sensitive plane readout)
- …

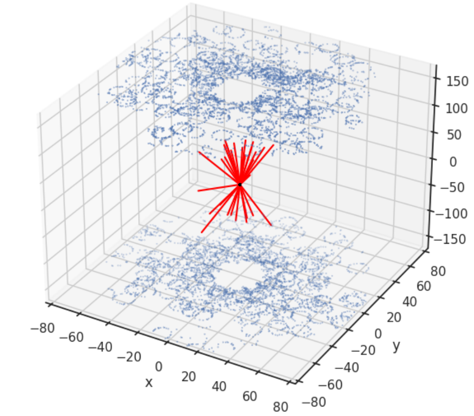


arXiv:2203.08806

Let's try to bypass internal intermediate details

FARICH Fast Simulation



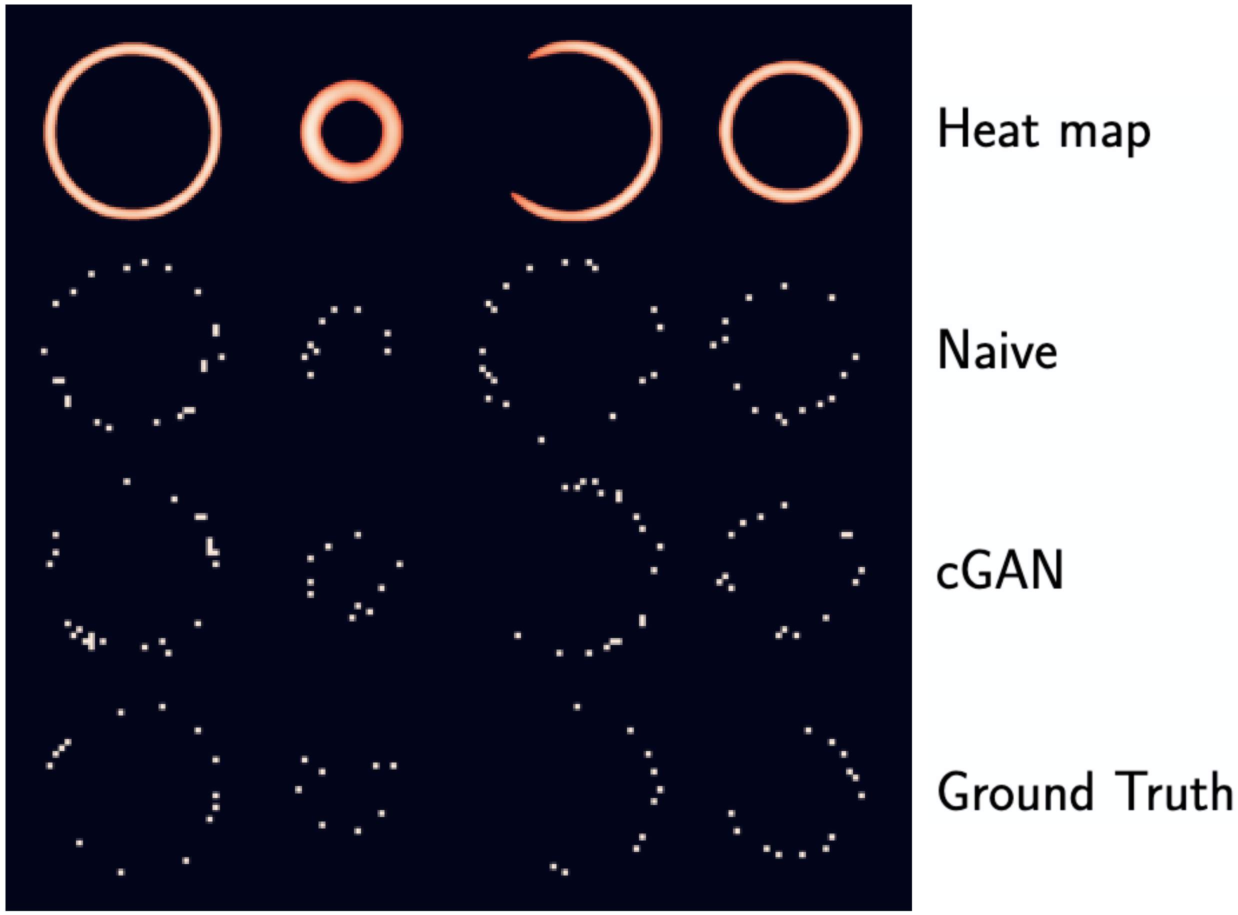


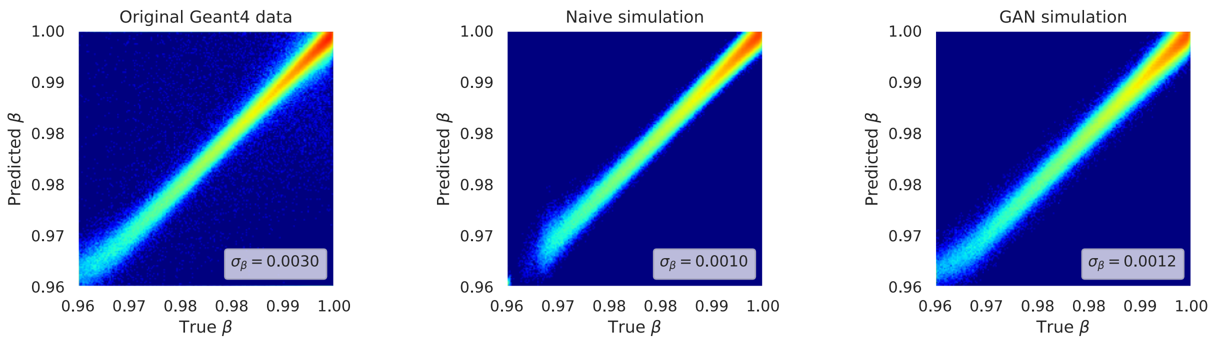
Let’s try to bypass internal intermediate details

Problem: hit map has discrete signals

- generative models are good for smooth distribution

Solution: generate not photon hits but probability map

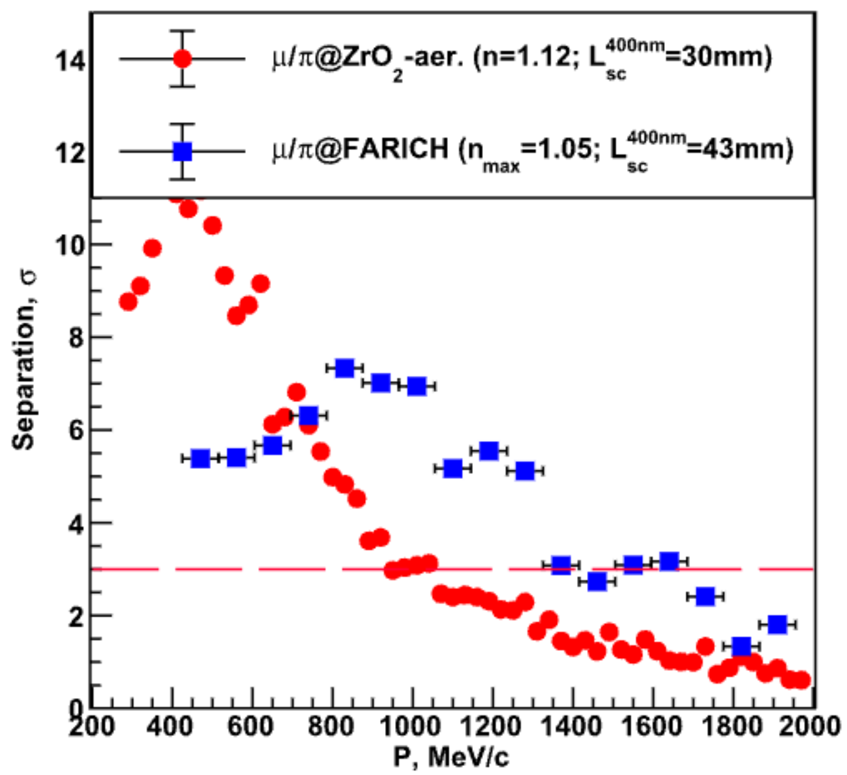


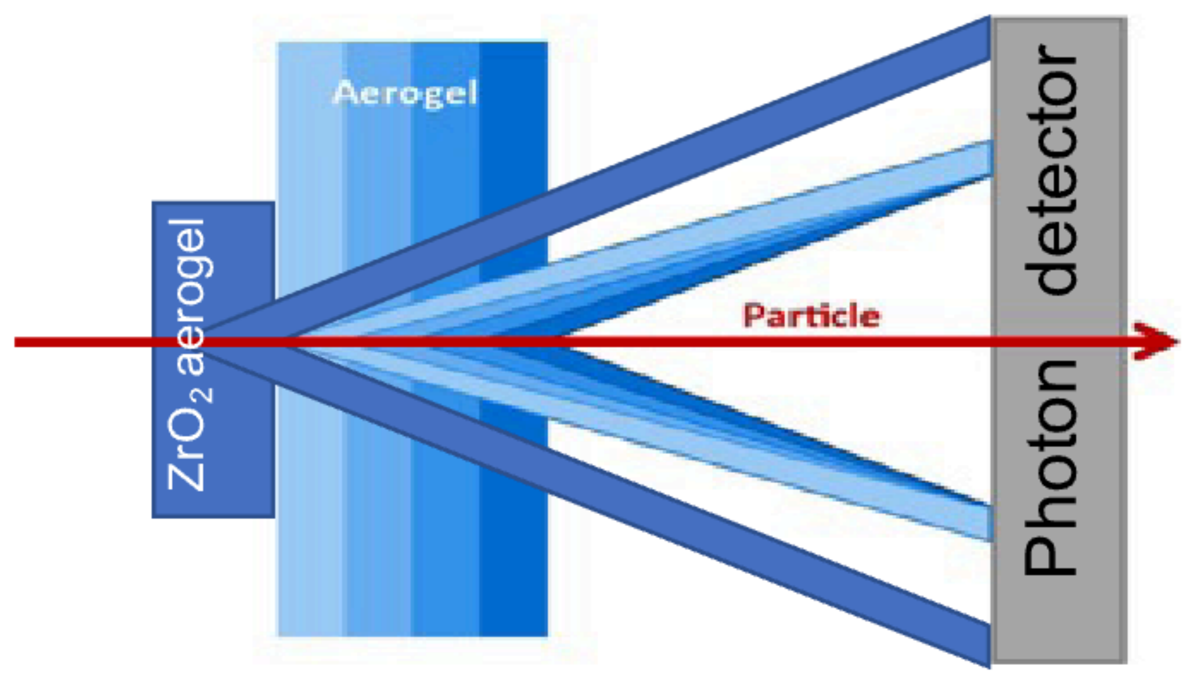


Method	Reco δ (TVD) ↓
Naive	0.30
cGAN	**0.19**

Further FARICH Ideas

Barnyakov et.al.
https://doi.org/10.1142/S0217751X24420120





Excellent problem for ML based PID!

Summary

- FARICH is an interesting detector tunable for PID in different conditions
- Computer vision approaches of the AI are well suted to this kind of data
- We checked different aspects of FARICH related software stack with AI approach and get promising results
- Real estimations must be done in situ in the actual detector physics, configuration and pipeline

Dataset: "Simple" FARICH GEANT4 Simulation

- photodetector
  - 30×30 matrix of SiPM ($57600 = 30 \times 30 \cdot 8 \times 8$ total channels);
  - 3.16×3.16 mm^2 pixels;
  - 1 mm gap between matrices.
- radiator
  - 4 layers, $n_{\mathrm{max}} = 1.05$;
  - 35 mm total depth;
  - 200 mm in front of the photo detector.
- π^- with varying angles $[0^\circ, \ldots, 45^\circ]$ and velocities $[0.957; 0.999]$

1.2M events

- photon hit coordinates
- hit times
- flat random noise is dynamically applied on top of the events

Experiment: Classification

Classification quality (AUCROC) by momentum and θ_p @ 10^6 Hz/mm^2

θ_p [deg] \ Momentum [MeV / c]	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200	2300
44	0.938	0.967	0.989	0.972	0.992	0.997	0.992	0.998	0.999	0.988	0.983	1	1	0.969	0.975	1	1	1	1
42	0.934	0.969	0.978	0.976	0.993	0.996	0.975	0.984	0.999	0.991	1	0.958	0.996	1	1	1	1	0.99	1
40	0.942	0.978	0.985	0.987	0.996	0.994	0.985	0.98	0.998	0.999	0.998	0.974	0.971	0.992	1	1	1	0.99	1
38	0.95	0.975	0.981	0.993	0.986	0.998	0.998	1	0.998	0.993	0.996	1	0.992	0.982	1	1	1	1	1
36	0.947	0.981	0.988	0.994	0.998	1	1	0.997	0.999	0.995	0.945	1	1	1	1	1	1	1	1
34	0.961	0.99	0.994	0.987	0.997	0.995	0.998	0.995	0.989	1	1	0.999	1	1	1	1	1	1	1
32	0.927	0.98	0.989	0.997	0.994	0.998	0.999	1	0.991	1	1	1	1	0.972	1	1	1	1	1
30	0.936	0.979	0.996	0.999	0.993	0.991	0.998	1	1	1	1	1	1	1	1	1	1	1	0.956
28	0.944	0.987	0.988	1	0.998	1	0.987	0.992	1	1	0.998	0.956	1	1	1	1	1	0.969	1
26	0.965	0.984	0.997	0.997	0.998	0.999	1	0.998	1	1	0.997	1	0.997	1	1	1	1	1	1
24	0.942	0.984	0.999	0.996	0.997	1	1	1	1	0.999	0.976	1	1	1	1	1	1	1	1
22	0.954	0.988	1	0.994	1	1	1	0.999	1	0.997	1	1	1	1	1	1	1	1	1
20	0.966	0.987	1	0.999	1	0.999	0.999	1	1	1	1	1	1	1	1	1	1	1	1
18	0.95	0.991	0.996	0.998	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	0.954	0.99	0.998	1	1	1	0.99	0.999	0.994	1	1	1	1	1	1	1	1	1	1
14	0.957	0.989	0.996	0.999	1	1	0.998	1	1	1	1	1	1	1	0.981	1	1	1	1
12	0.96	0.989	0.997	0.999	1	0.988	1	1	1	1	1	1	0.979	1	1	1	1	1	1
10	0.951	0.99	1	0.997	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0.937	0.998	0.999	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	0.921	0.988	0.988	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

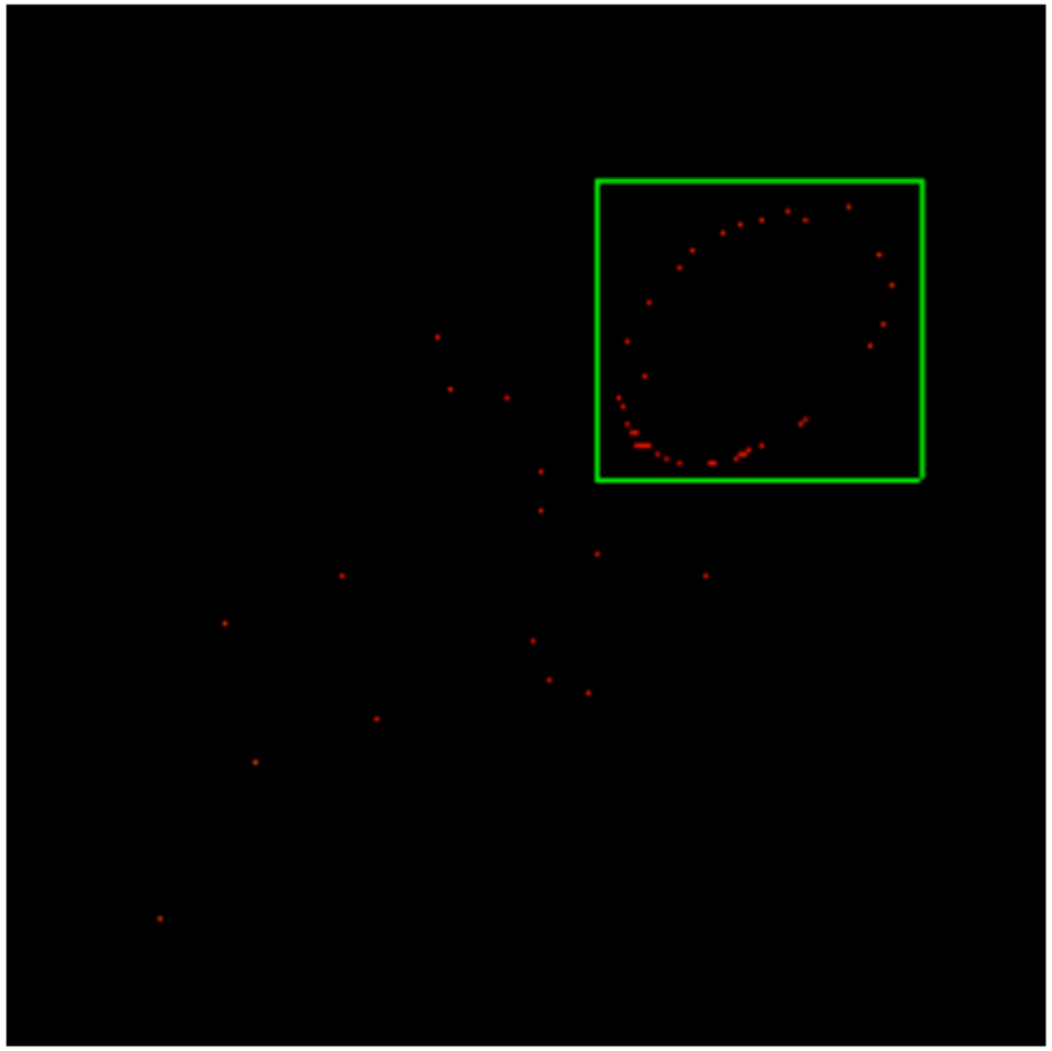
Online Filtering

Approach 2: Bounding box (bbox) regression

- compute ground truth bboxes for signal ellipses
- train CNN to extract bbox coordinates

Metrics:

- Efficiency (area) $= \frac{|B \cap B^{gt}|}{|B^{gt}|}$
- Reduction (area) $= \frac{|B|}{|B_{max}|}$



Ground truth bbox example

Experiment: BBox regression