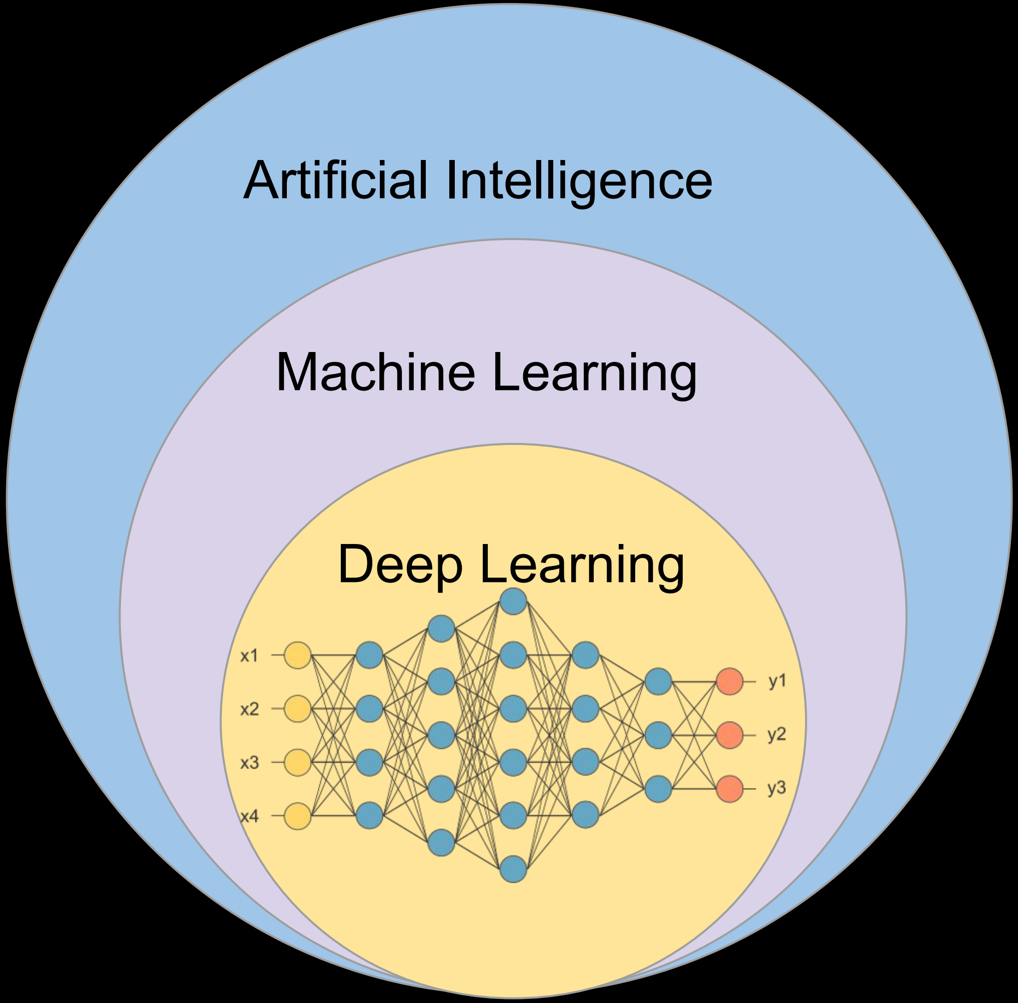
Machine Learning for Imaging Cherenkov Detectors

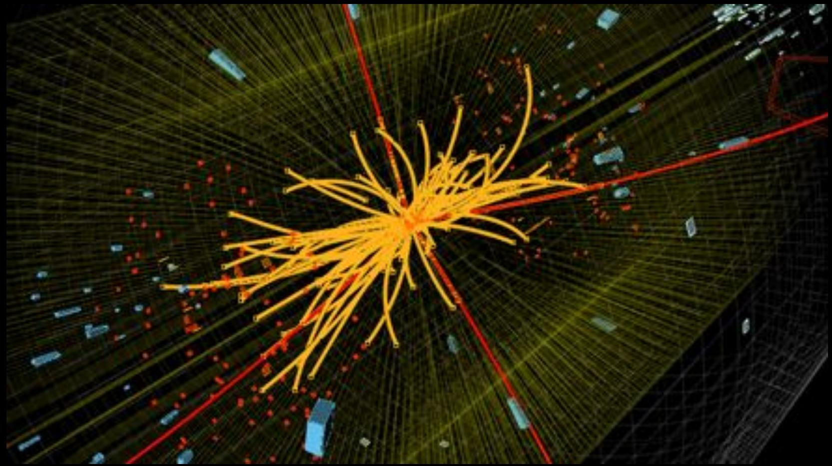
Cristiano Fanelli

MIT

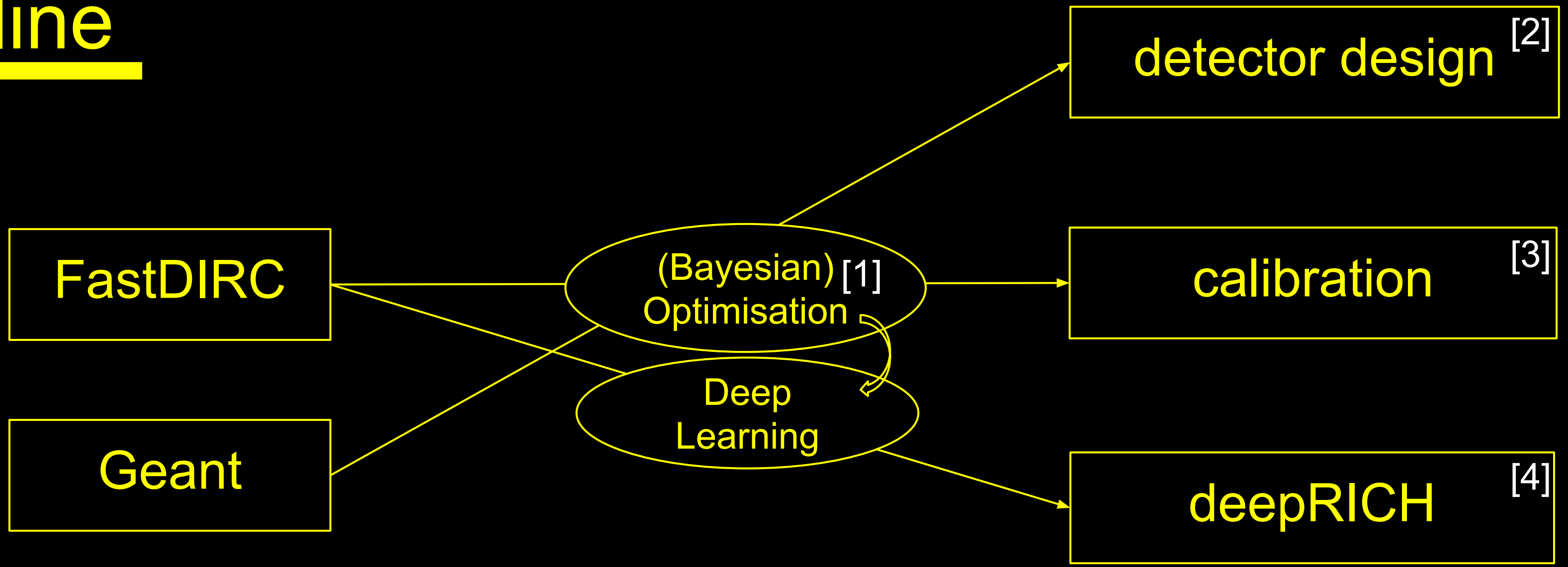
EIC² EIC Center at Jefferson Lab



- DL is a subset of ML which makes the computation of multi-layer NN feasible. When applied to massive datasets and giving massive computer power it outperforms all other models most of the time.
- ML is becoming ubiquitous in nuclear and particle physics.
- DL just started having an impact in nuclear/particle physics



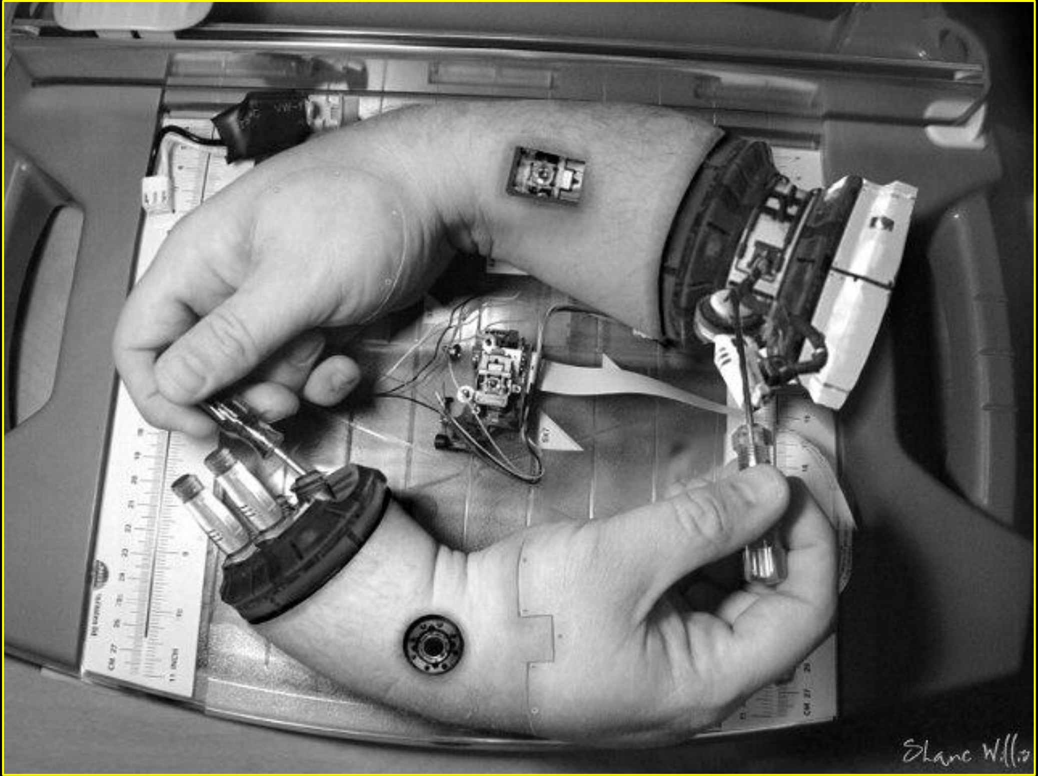
Outline



1. Short intro on BO
2. EIC dRICH detector design NEW!
3. GlueX DIRC optical box calibration using FastDIRC
4. Exploring deep learning for DIRC NEW!

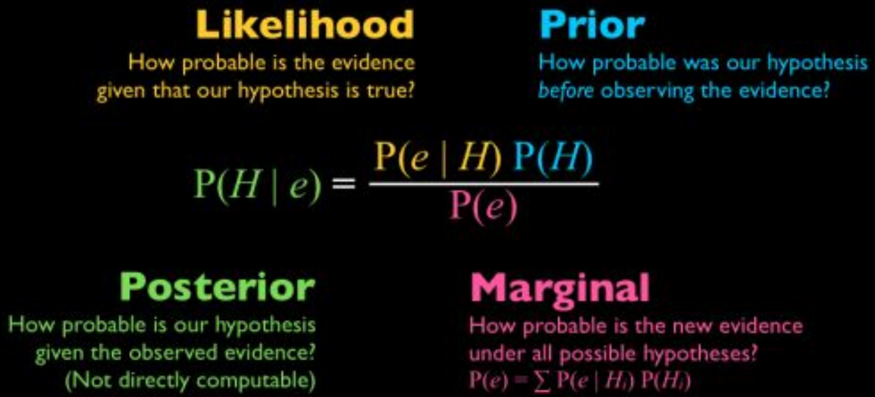
Conclusions

Optimization



Simplest Approaches

- We are not really great at interpreting high-dimensional data
  - Manual Search
    Good luck!
  - Grid Search
    Easy but scales poorly -> curse of dimensionality
  - Random Search
    Faster, but won't guarantee optimal search
- What if we can self-learn the optimal values?
  - Bayesian Optimization
    Takes advantage of the information the model learns during the optimization process.



BO Applications

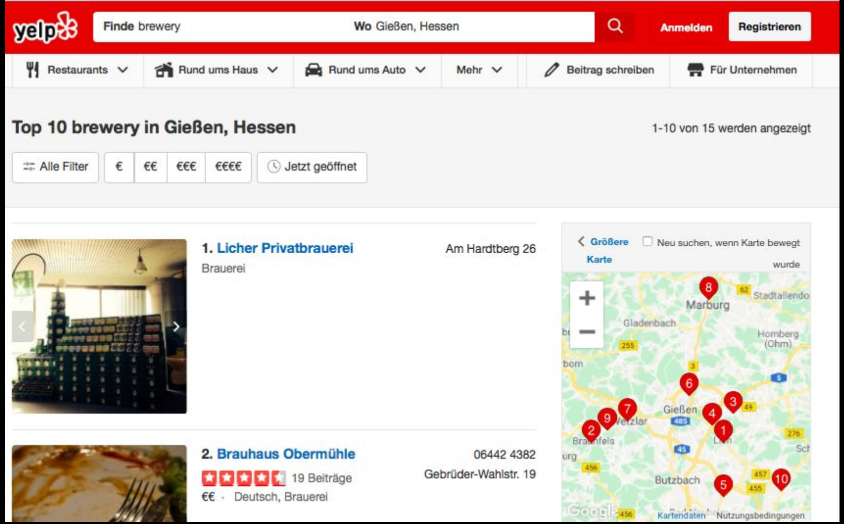
This approach finds a lot of applications:

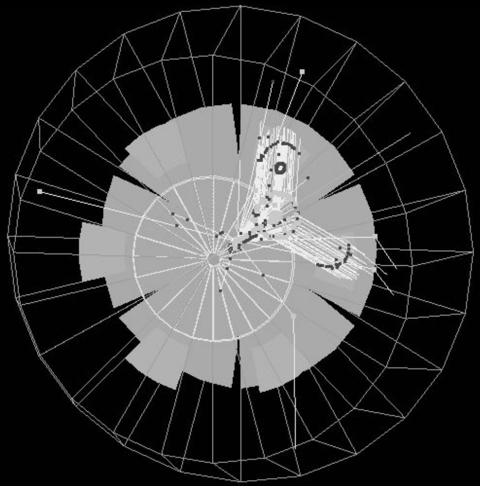
- E.g. Hyperparameters

In particle physics:

- Tuning Simulations [1610.08328])
- Novel directions (this talks):
  - Optimal Design (hardware, ...)

    (EIC dRICH)
  - Calibration (cf. GlueX DIRC)

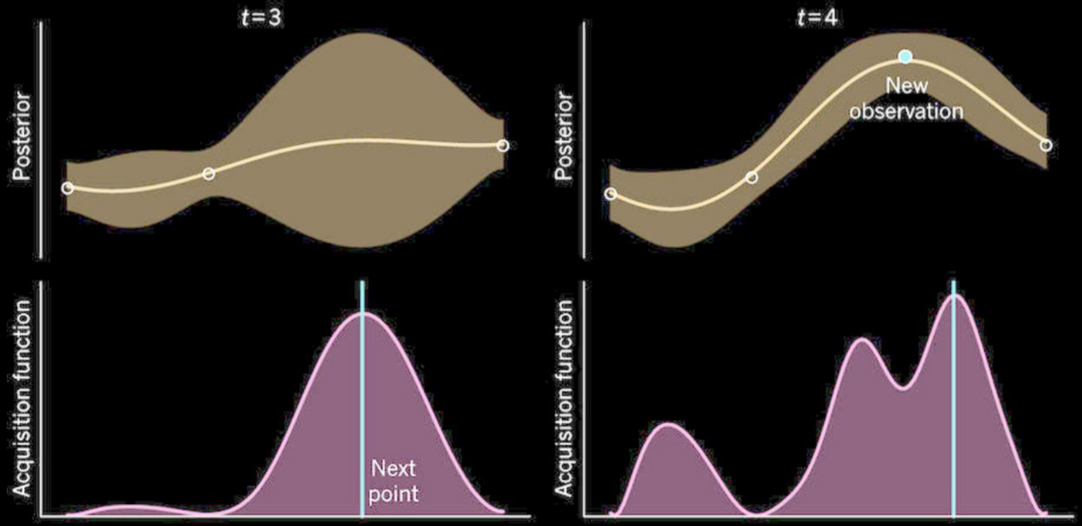


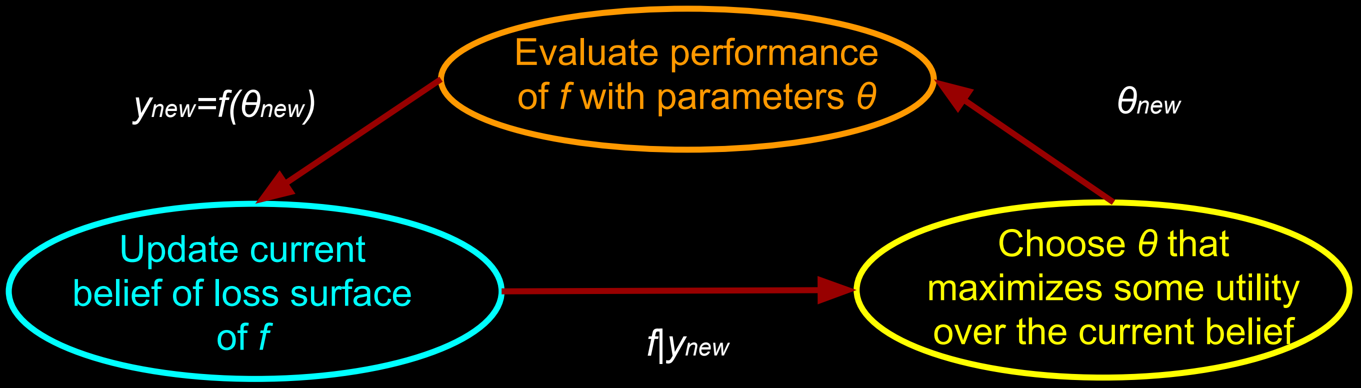


Can work with noisy, non-differentiable black-box functions

How it works

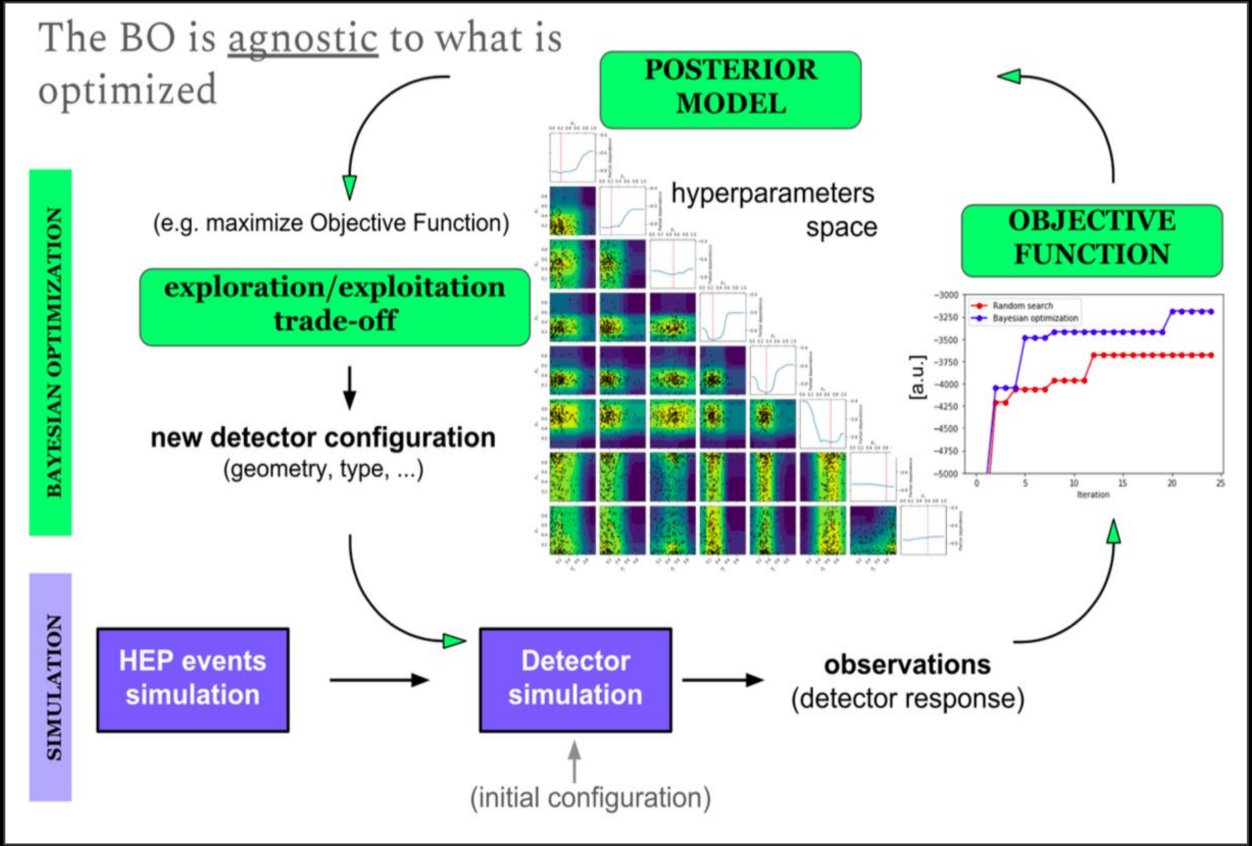
- BO is a strategy for global optimization.
- After gathering evaluations BO builds a posterior distribution used to construct an **acquisition function.**
- This cheap function determines what is **next query point.**



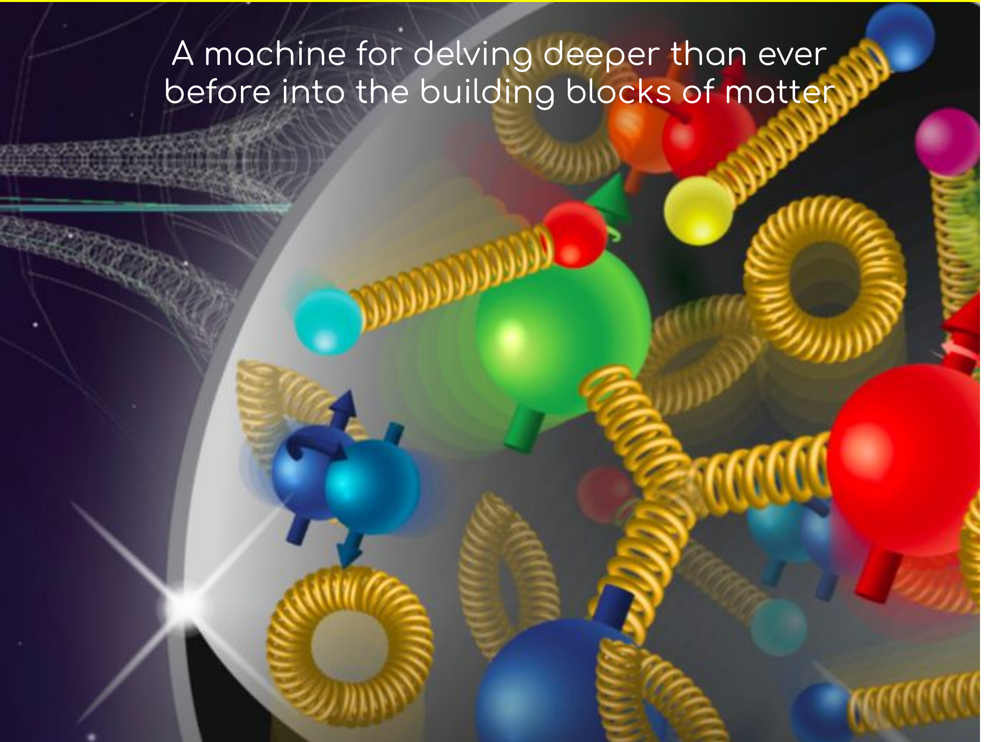


Detector Optimization

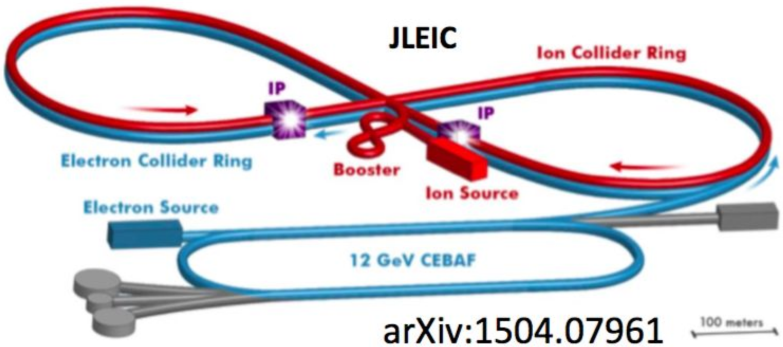
- Optimization of detector design is quite complex problem that can be accomplished with BO
- Multi-purpose detector requires large-scale simulations of the main processes to make decision
- Goal: satisfy detector requirements and minimize cost R&D



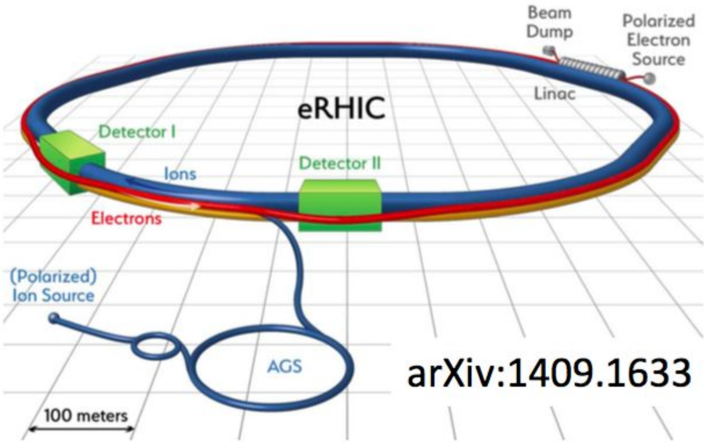
Electron Ion Collider



Building the future EIC is the top long-term priority for medium/high-energy nuclear physics in the U.S.
It already consists of a large international collaboration.



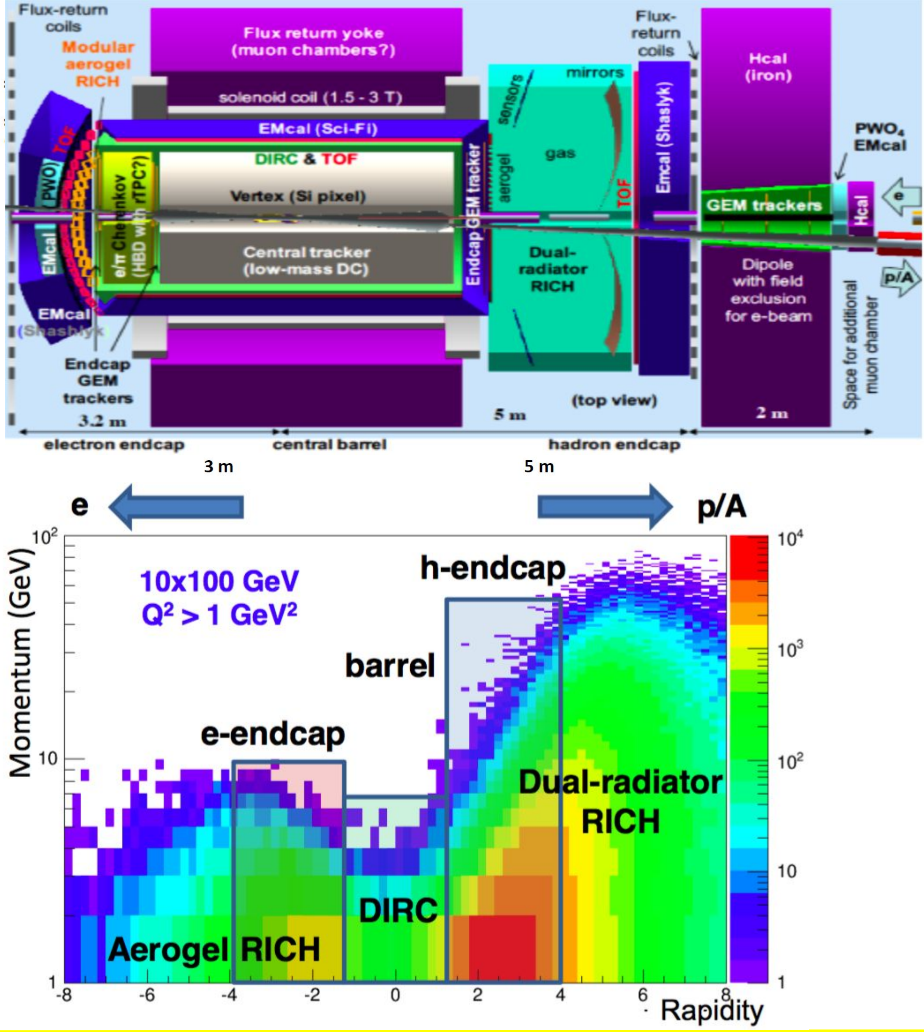
arXiv:1504.07961



arXiv:1409.1633

PID

- h-endcap A dual-radiator RICH is needed to cover continuously momenta up to 50 GeV/c
- **e-endcap**: A small lens focused aerogel RICH for momenta up to 10 GeV/c
- **Barrel**: A DIRC provide a compact and cost effective way to cover momenta up to 6 GeV/c
- **TOF** (and or dE/dx in the TPC) can cover the low momenta region



dRICH

Full momentum, continuous coverage.

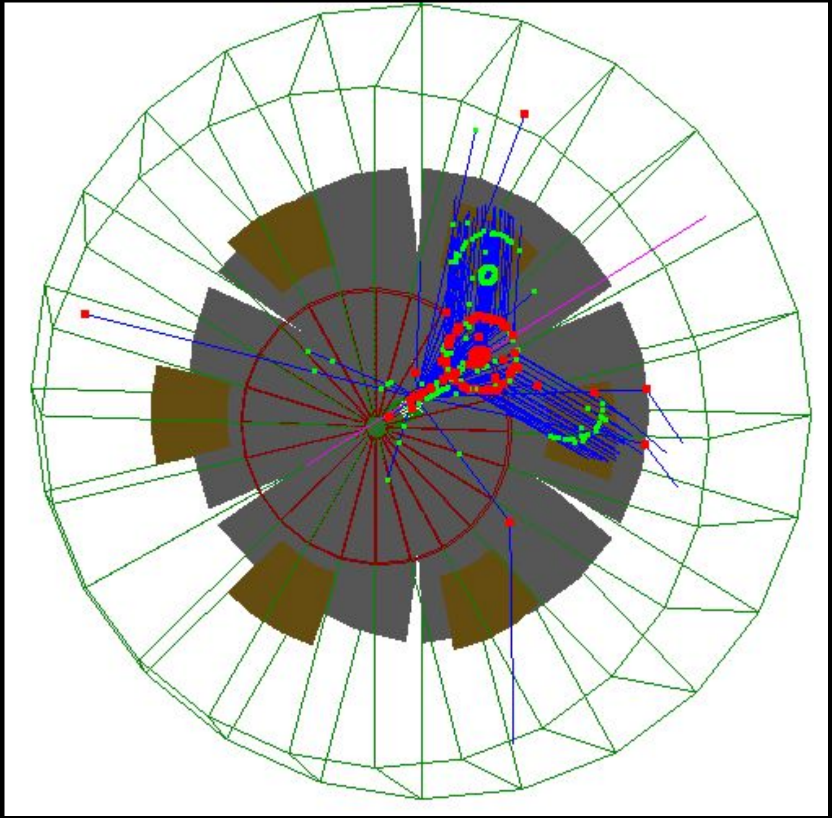
Cost effective

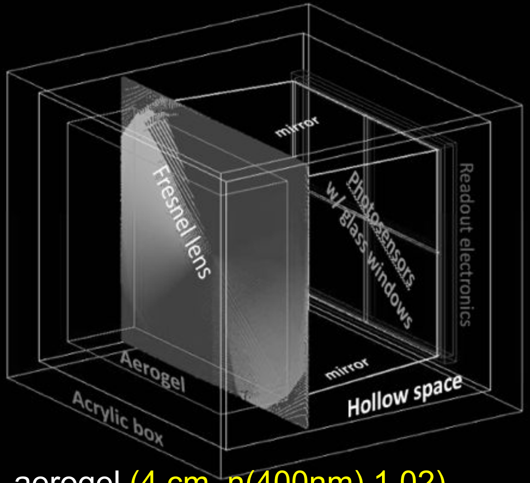
Simple geometry/optics.

6 Identical open sectors (petals)

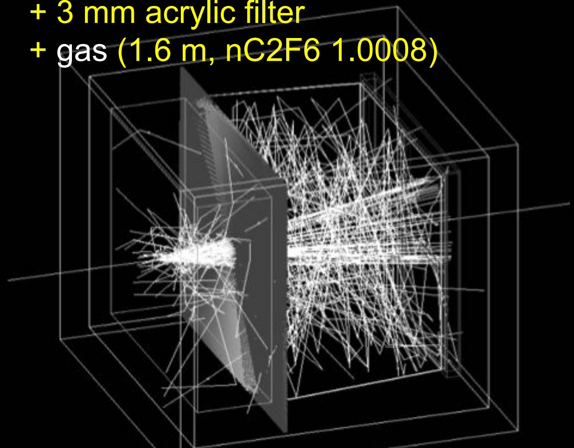
Optical sensor elements: 4500 cm^2/sector, 3 mm pixel

Large Focusing Mirror





aerogel (4 cm, n(400nm) 1.02)
+ 3 mm acrylic filter
+ gas (1.6 m, nC2F6 1.0008)



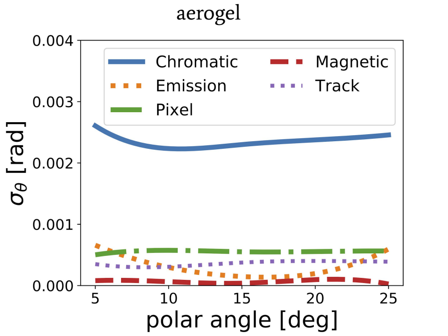
9 GeV Pion simulation

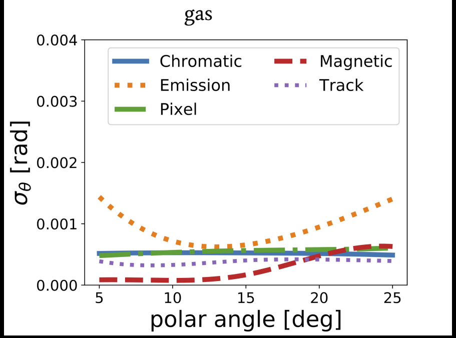
See A. Del Dotto, EICUG2017, and E. Cisbani's talk

dRICH Optimization

$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

parameter	description	range [units]
R	mirror radius	[290.0,300.0] [cm]
pos r	radial position of mirror center	[125.,140.] [cm]
pos l	longitudinal position of mirror center	[-305.,-295.] [cm]
tiles y	shift along y of tiles center	[-5,5] [cm]
tiles z	shift along z of tiles center	[-105,-95] [cm]
tiles x	shift along x of tiles center	[-5,5] [cm]
$n_{aer.}$	refraction index of aerogel	[1.015,1.03]
$t_{aer.}$	aerogel thickness	[3.0,6.0] cm

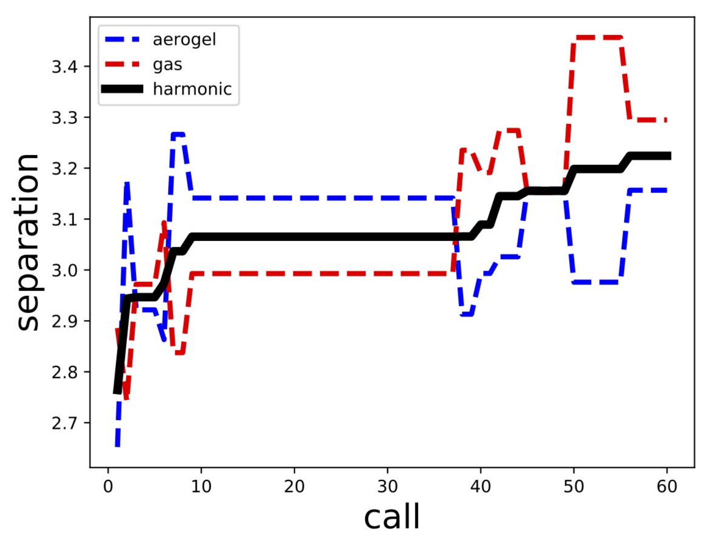




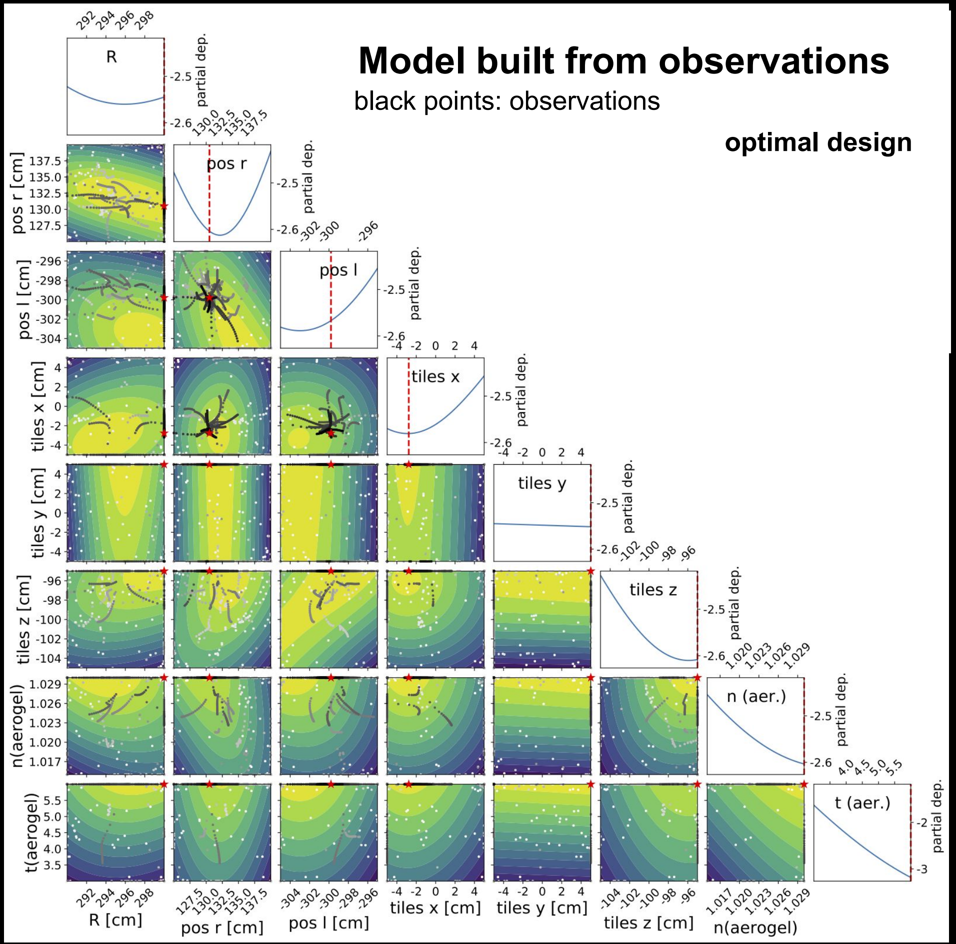
Ranges mainly due to mechanical constraints and optics requirements.

These requirements can change in the next future based on inputs from prototyping.

Results



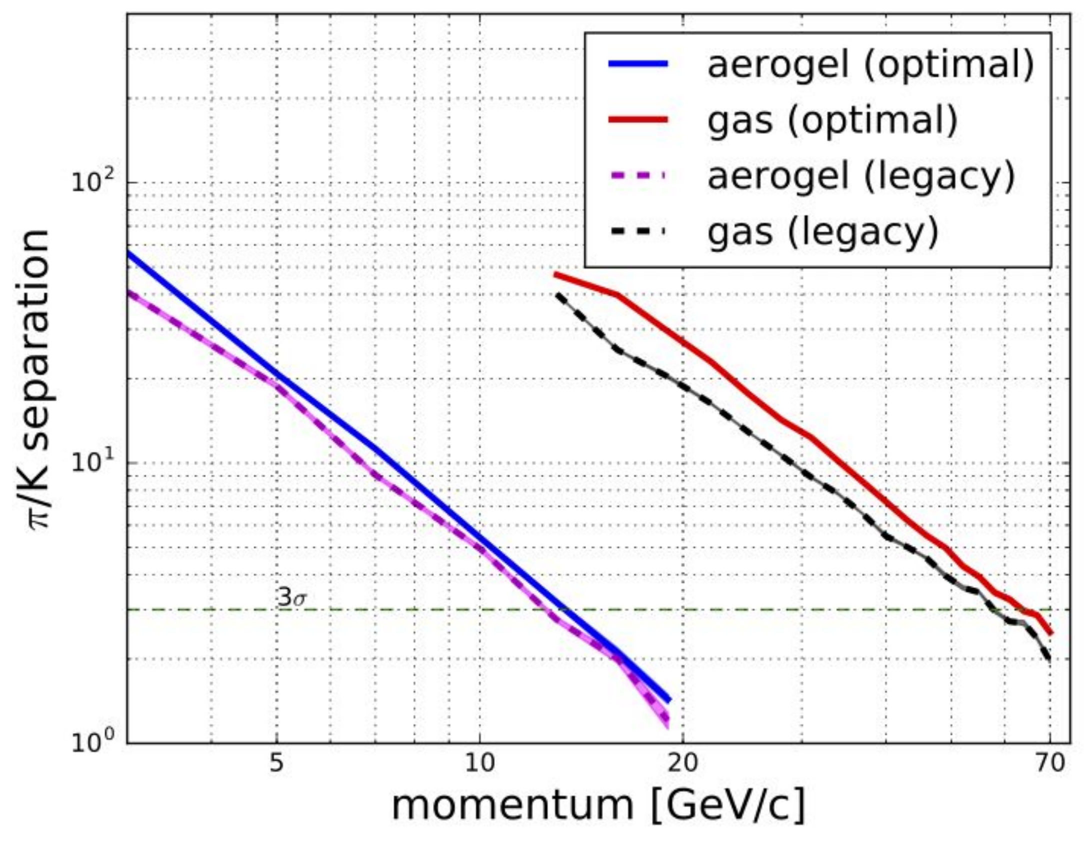
	FoM (h)	$(N\sigma)$ @ 14 GeV/c	$(N\sigma)$ @ 60 GeV/c
BO	3.23	3.16	3.30
legacy	2.9	3.0	2.8

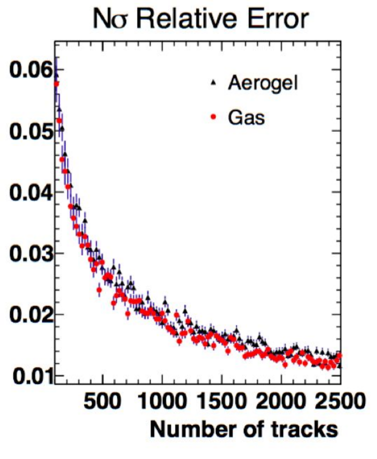


improved "speed" of convergence - tested different regression methods - implemented stopping criteria - determined tolerances

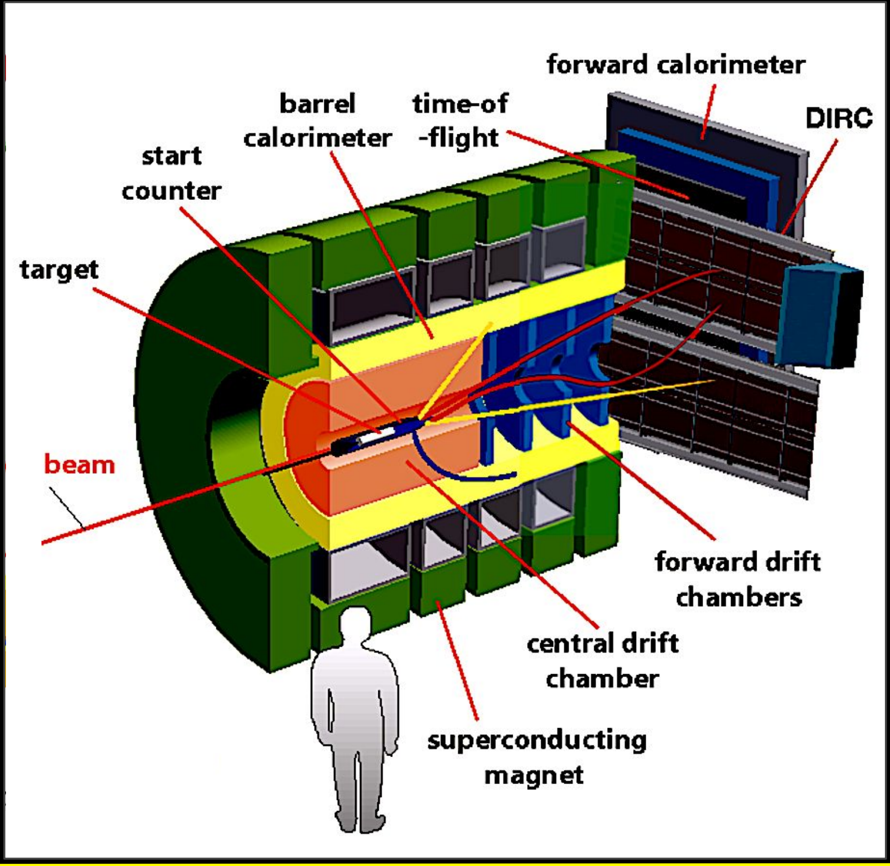
Preliminary

E. Cisbani, A. Del Dotto, CF





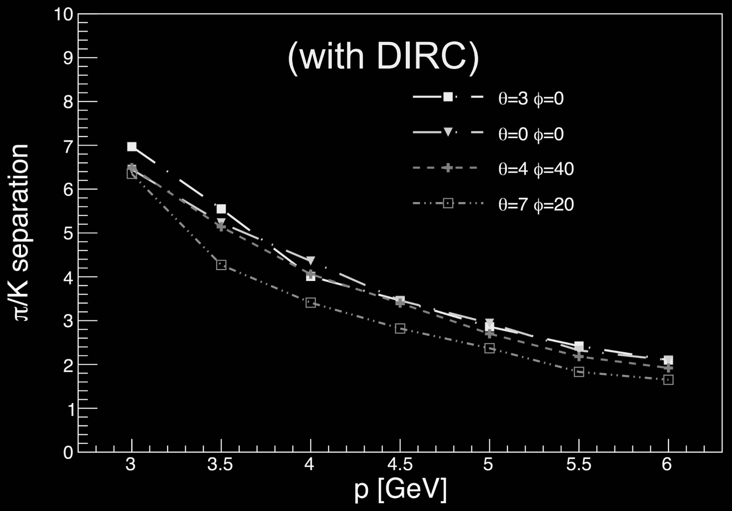
GlueX DIRC calibration

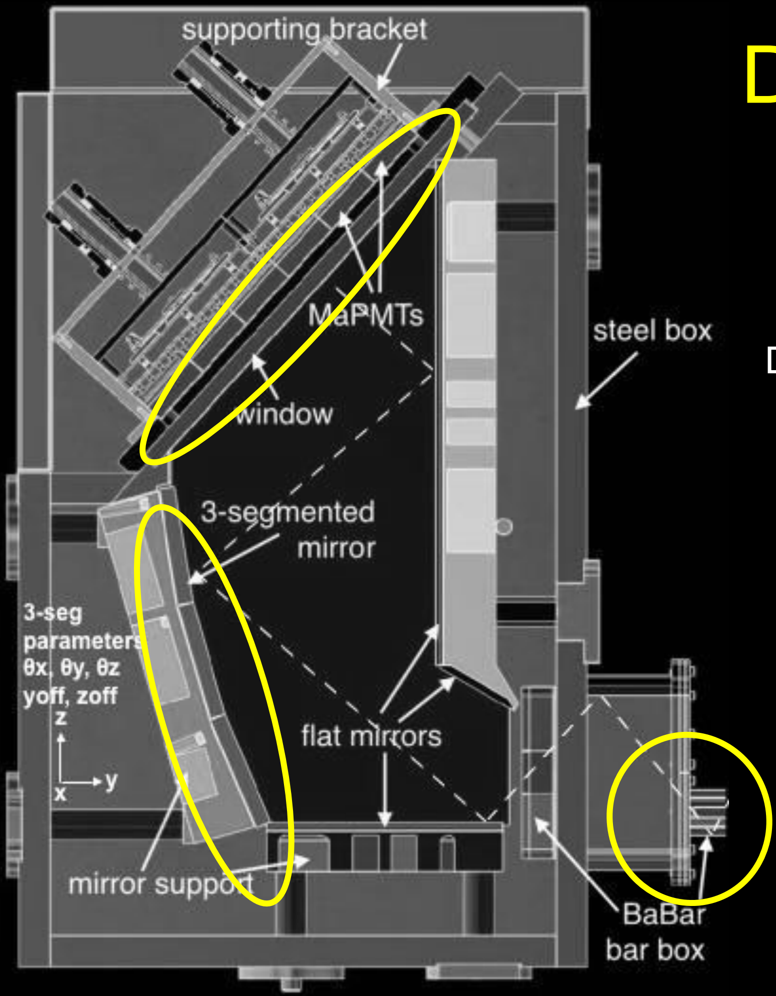


see J. Stevens' talk

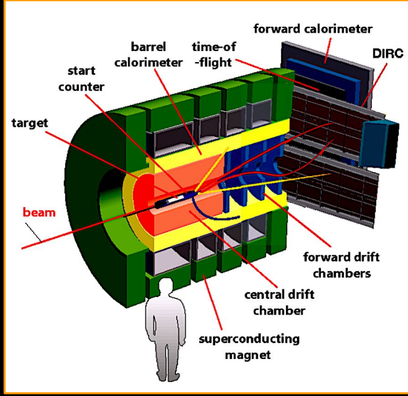
DIRC will improve GlueX PID capabilities

(current π/K separation limited to 2 GeV/c)



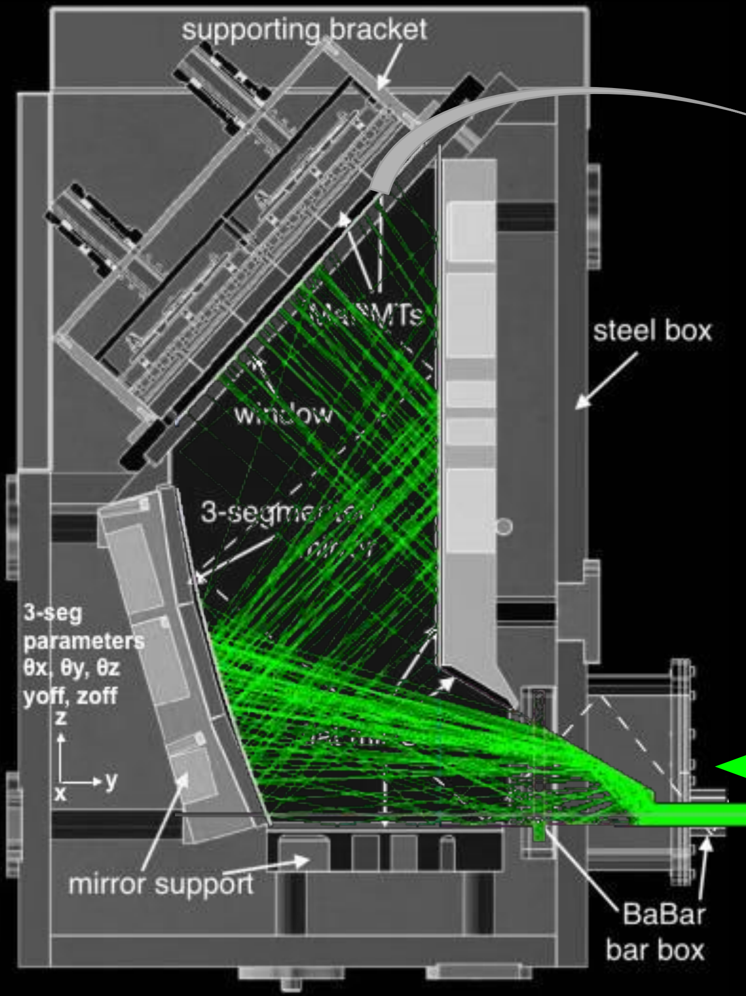


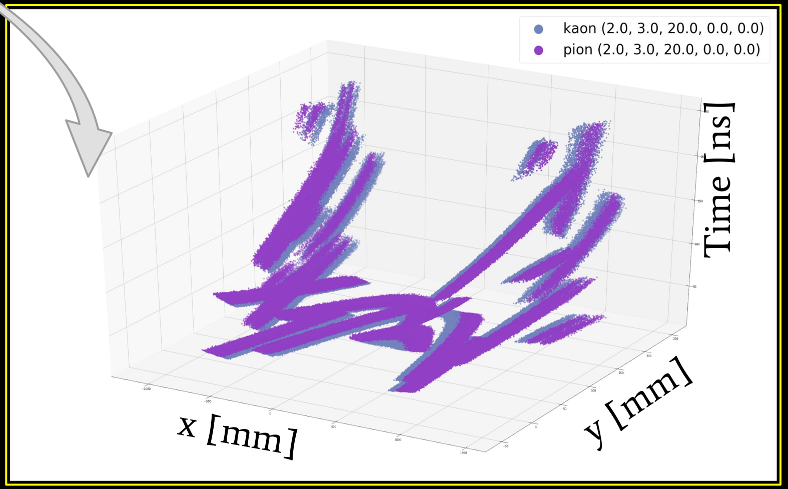
Detector Alignment

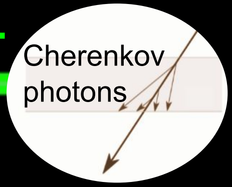


DIRC @ GlueX/JLab

- Optical box made by several components and filled by water.
- During data-taking this becomes a noisy black-box problem with many non-differentiable terms.
  - relative alignment of the tracking system with the location and angle of the bars
  - mirrors shifts cause parts of the image change
  - other offsets
- These aspects make seemingly impossible to analytically understand the change in PMT pattern
- Requires dedicated system for calibration.



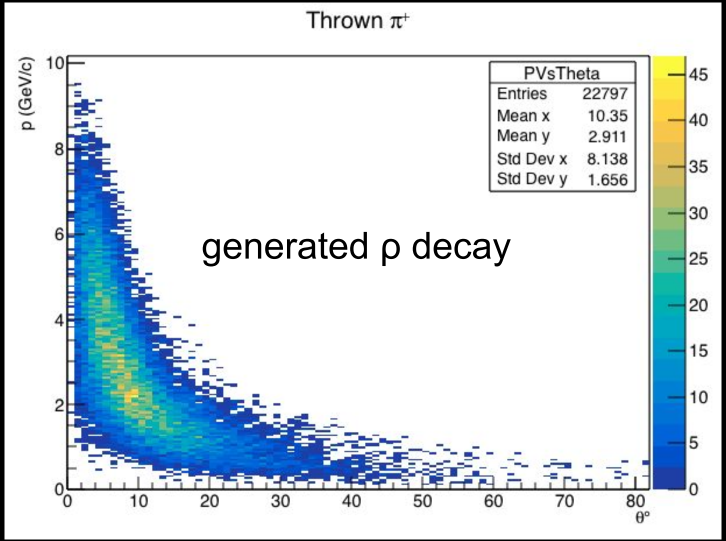




particle track

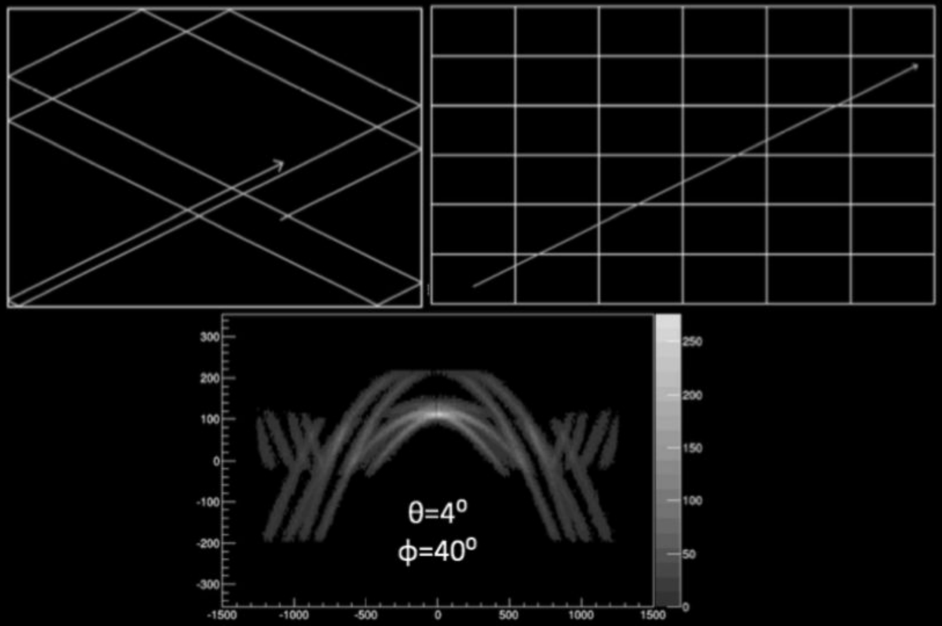
Pure sample of particles for alignment

- The idea is to use pure sample of pions produced by abundant channels like ρ decays
- At low momentum they are well identified by current GlueX PID capabilities.
- Use these pions as candles for alignment.
- Test alignment with one bar first and for a subrange of kinematics (momentum, angles, and position in the bar) - *proof of principle*
- Generalize technique (to kaons, other bars, etc.)



FastDIRC

J. Hardin and M. Williams, JINST 11.10 (2016)



Fast tracing, mapping straight lines through a tiled plane

1. Generation
2. Traces through bars
3. Traces through expansion volume

KDE-based

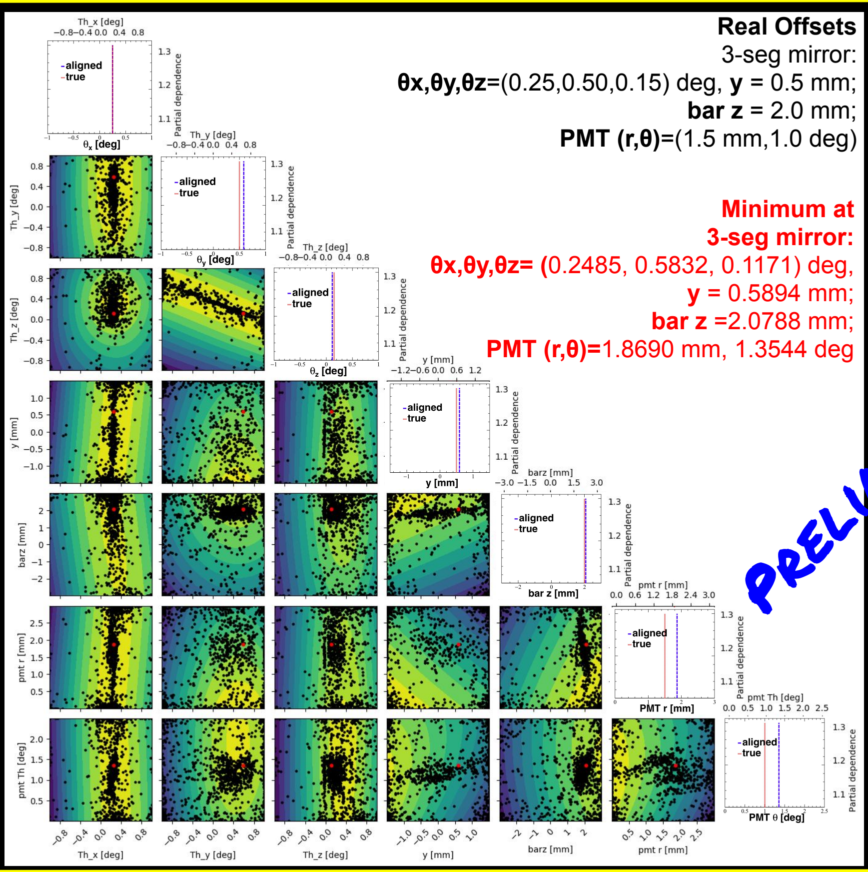
$$P(x) \approx \sum_{i}^{n} K(x - s_i)$$

open source
https://github.com/jmhardin/FasDIRC

better resolution in regions with high overlap

Toy model with main offsets

see C. Fanelli, EIC ML seminar



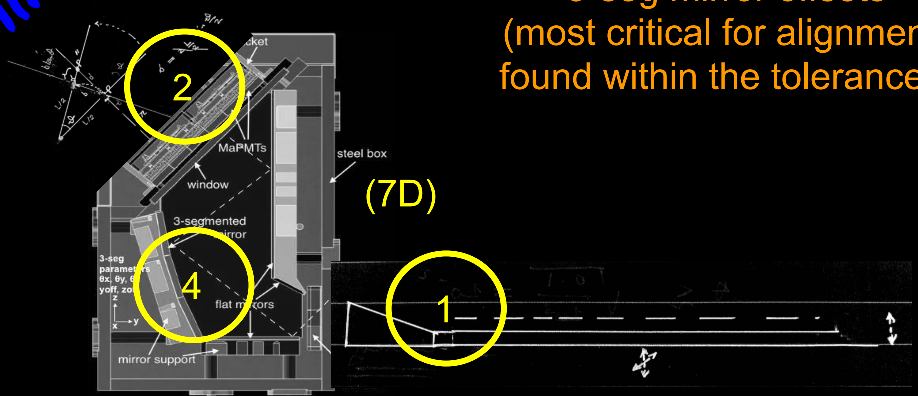
Real Offsets
3-seg mirror:
θx,θy,θz=(0.25,0.50,0.15) deg, **y** = 0.5 mm;
bar z = 2.0 mm;
PMT (r,θ)=(1.5 mm,1.0 deg)

Minimum at
3-seg mirror:
θx,θy,θz= (0.2485, 0.5832, 0.1171) deg,
y = 0.5894 mm;
bar z =2.0788 mm;
PMT (r,θ)=1.8690 mm, 1.3544 deg

PRELIMINARY

Particles used = 15000
Points explored = 1200

FoM = LogL normalized to a default alignment



(7D)

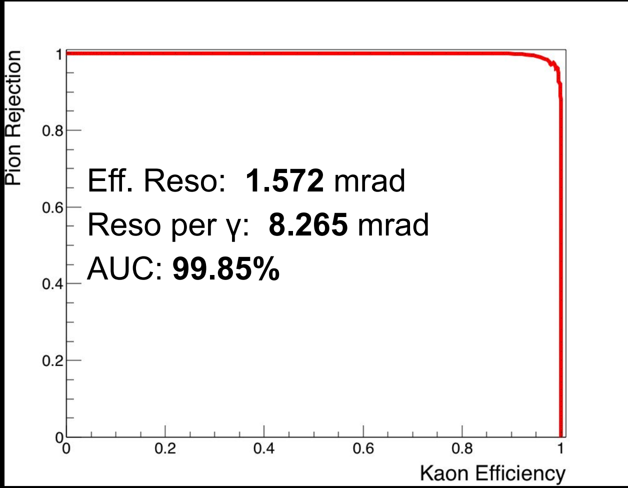
3-seg mirror offsets (most critical for alignment) found within the tolerances.

Toy model with main offsets

see C. Fanelli, EIC ML seminar

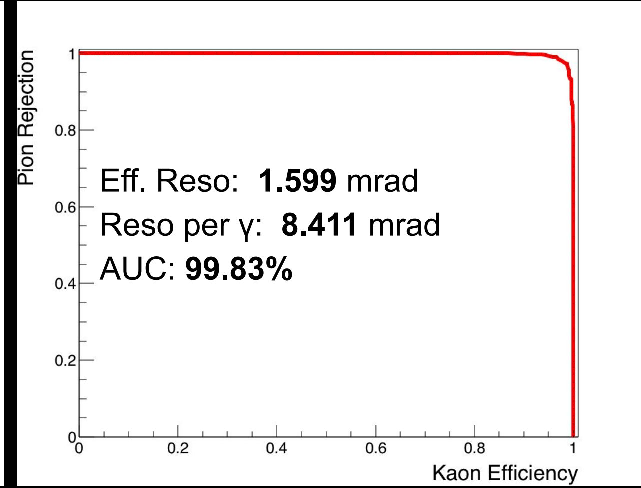
correct

3-seg mirror:
θx,θy,θz=(0.25,0.50,0.15) deg,
y = 0.5 mm;
bar z = 2.0 mm;
PMT (r,θ)=(1.5 mm,1.0 deg)



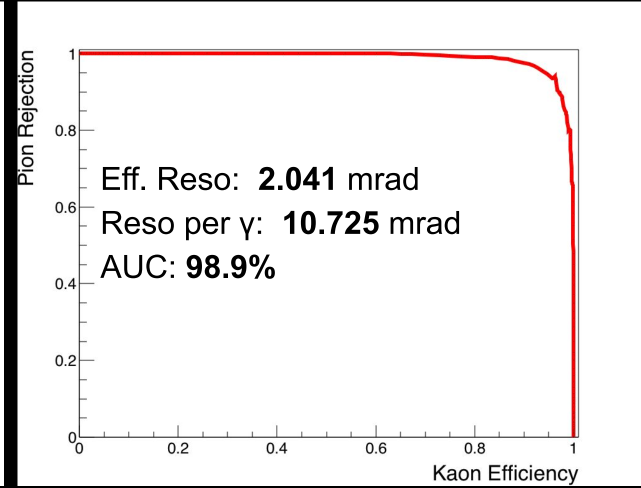
calibrated

3-seg mirror:
θx,θy,θz=(0.2485, 0.5832, 0.1171) deg,
y = 0.5894 mm;
bar z = 2.0788 mm;
PMT (r,θ)=(1.8690 mm, 1.3544 deg)



non-corrected

3-seg mirror:
θx,θy,θz=(0., 0., 0.) deg,
y = 0. mm;
bar z = 0. mm;
PMT (r,θ)=(0. mm, 0. deg)

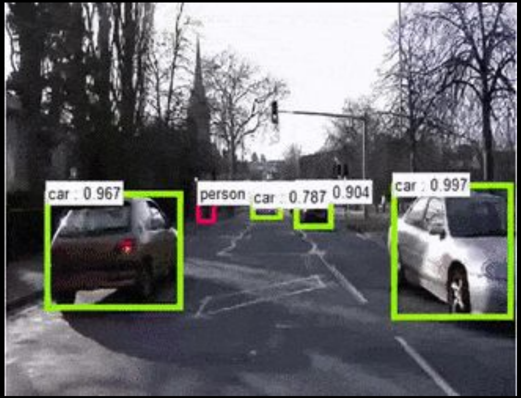


Kinematics: (E , θ, φ): (4 GeV, 4 deg, 40 deg)

Deep Learning

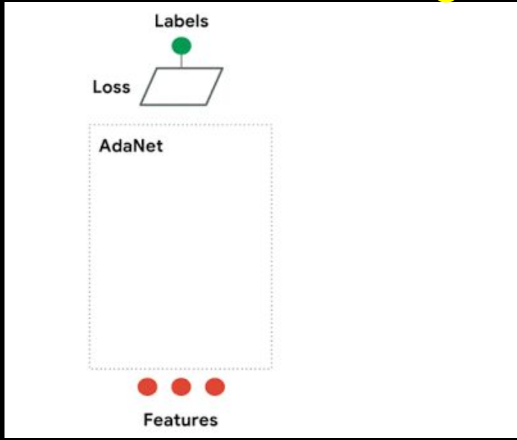
we stand at the height of some of the greatest accomplishments that happened in DL

Autopilot [2]

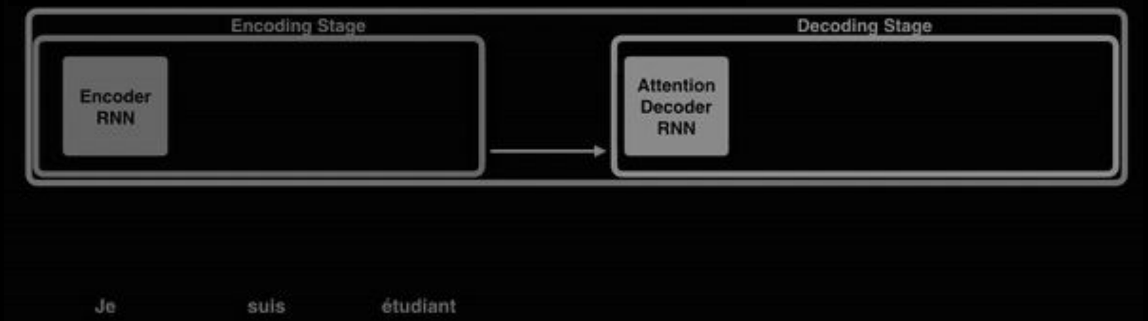


Ref [1] [2] [3] [4]

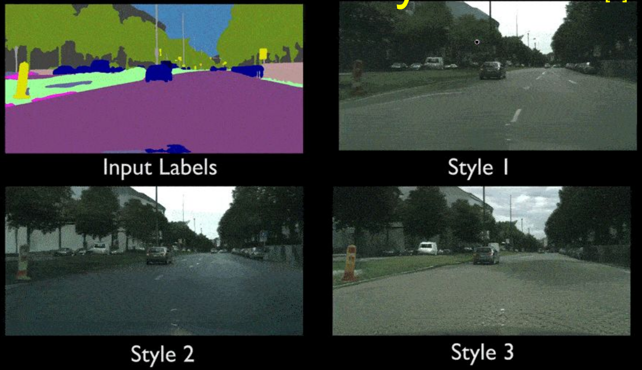
Meta-learning [3]



Natural Language Processing [1]

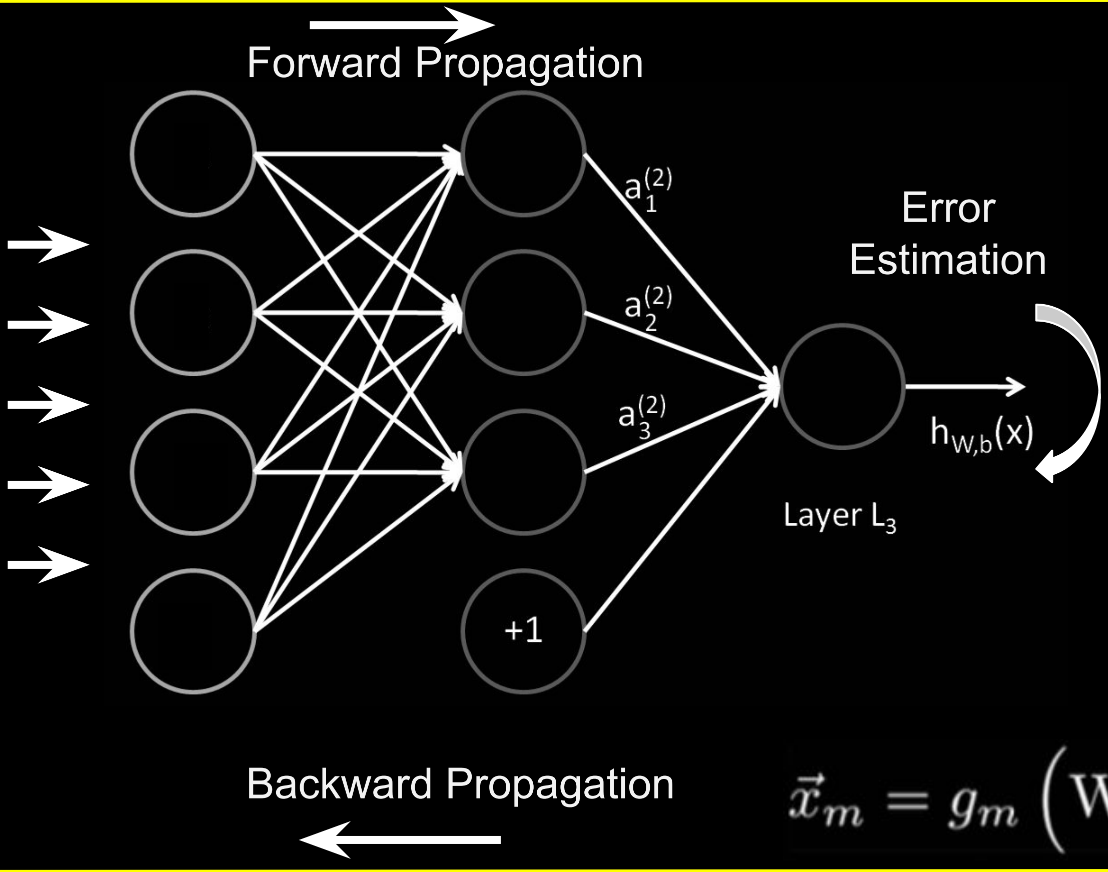


Video to video synthesis [4]



...but this is also the beginning of this incredible data-driven technology, in particular in our field

NN: How does it work?

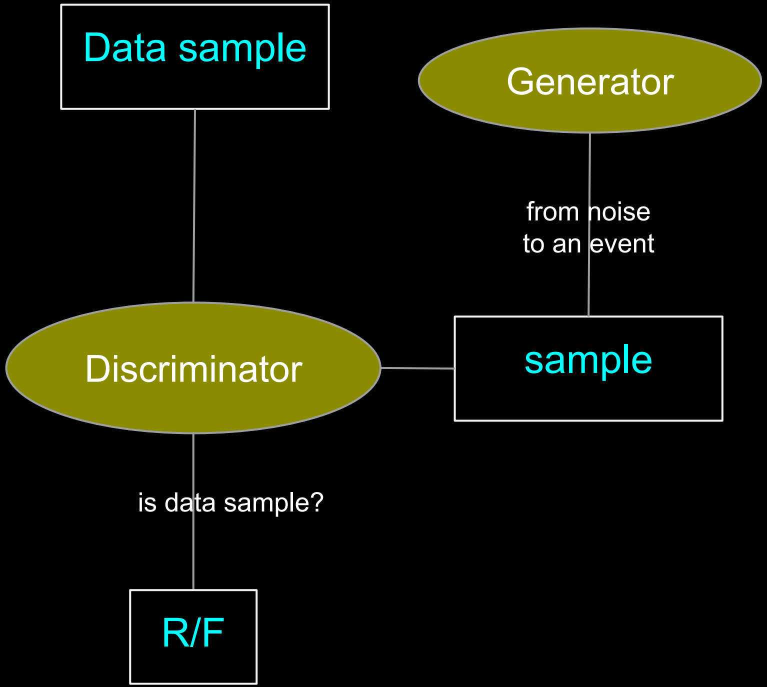


- The real magic about NN is the result of an optimization technique: back-propagation (how a NN works to improve its output over time)
- DL (more hidden) nets are good in learning non-linear functions (heavy processing tasks)
- Based on old school NN revitalized by augmented capabilities (e.g. GPU) and a plethora of new architectures (RNN, CNN, autoencoders, GAN, etc.)

$$\vec{x}_m = g_m\left(W_{m,m-1}\vec{x}_{m-1} + \vec{b}_m\right)$$

Generative Adversarial Network

arXiv:1406.2661



Fast Simulations

- Detailed simulation of detector response is provided by amazing tools like Geant, which is slow and often prohibitive for generating large enough samples.
- Cutting-edge application of deep learning uses GAN for fast simulation.
- 2-NN game, one model maps noise to images, the other classifies the images if real or fake.
- The goal is to confuse the discriminator.

- CALOGAN: Paganini, de Oliveira, Nachman 1705.02355
- jet images production: 1701.05927

CALOGAN can generate the reconstructed CALO image using random noise, skipping the GEANT and RECO steps

ML/DL for DIRC

Cherenkov detectors fast simulation using neural networks
D. Derkach et al, NIM (in press)

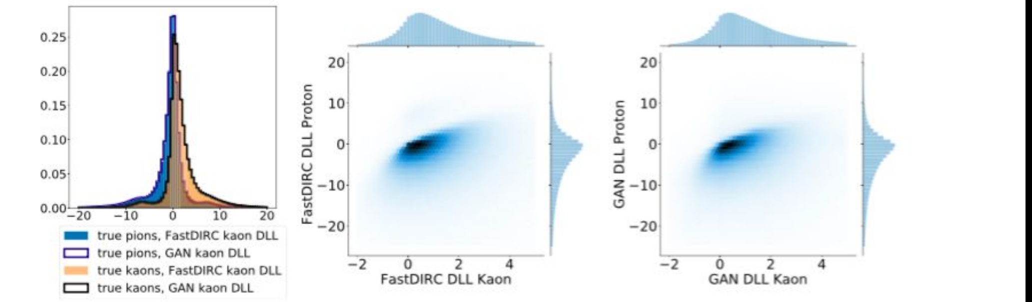


Fig. 1. Left: An example of 1D projection to kaon delta log-likelihood observables for FastDIRC and GAN simulation for samples consisting true pions (blue) and true kaons (brown). Centre and Right: An example of 2D projection to kaon and proton delta log-likelihood observables for FastDIRC (left) and GAN (right) simulation. The sample made of true pion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

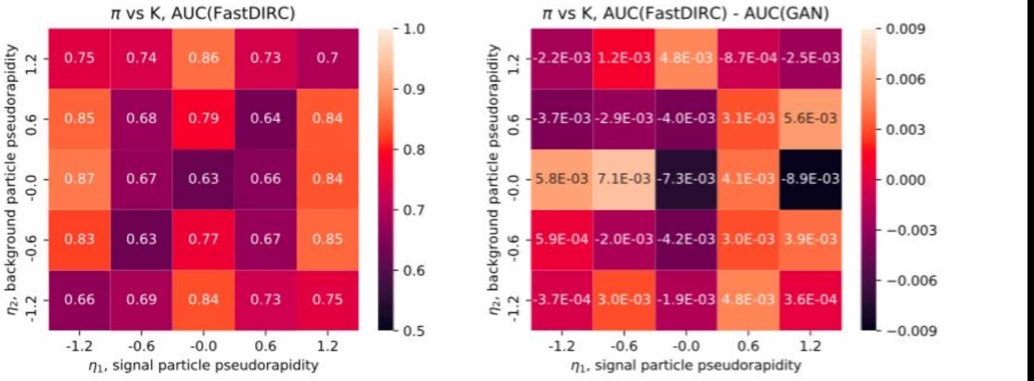


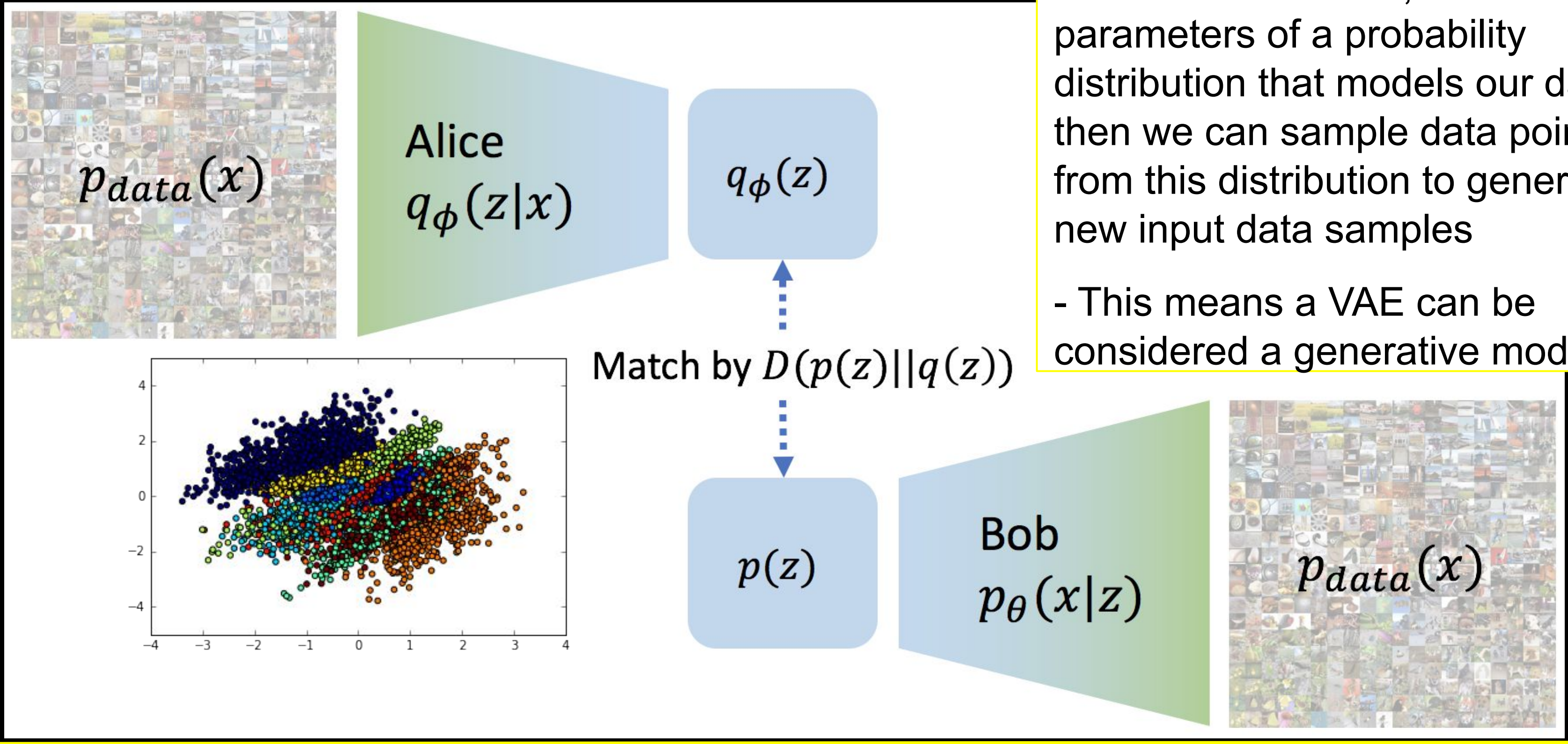
Fig. 2. Separation power between kaons and pions measured in area under receiver operating characteristic curve score (AUC score). Left is the FastDIRC simulation, right is the difference between GAN and FastDIRC AUC scores. The statistical uncertainty is around 0.005.

Variational autoencoder

- It learns a latent variable model of its input data

- Instead of letting the network to learn some function, we learn the parameters of a probability distribution that models our data, then we can sample data points from this distribution to generate new input data samples

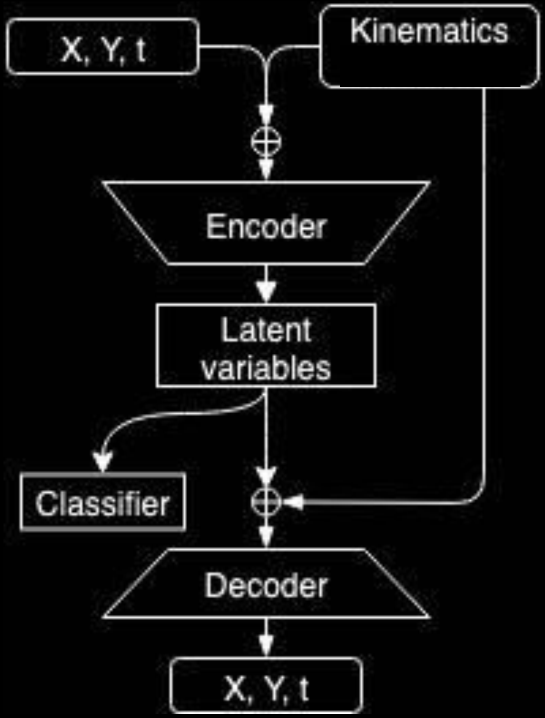
- This means a VAE can be considered a generative model



DeepRICH:

CF, J. Pomponi

(preliminary)



The model is trained minimizing a total loss function, consisting of:

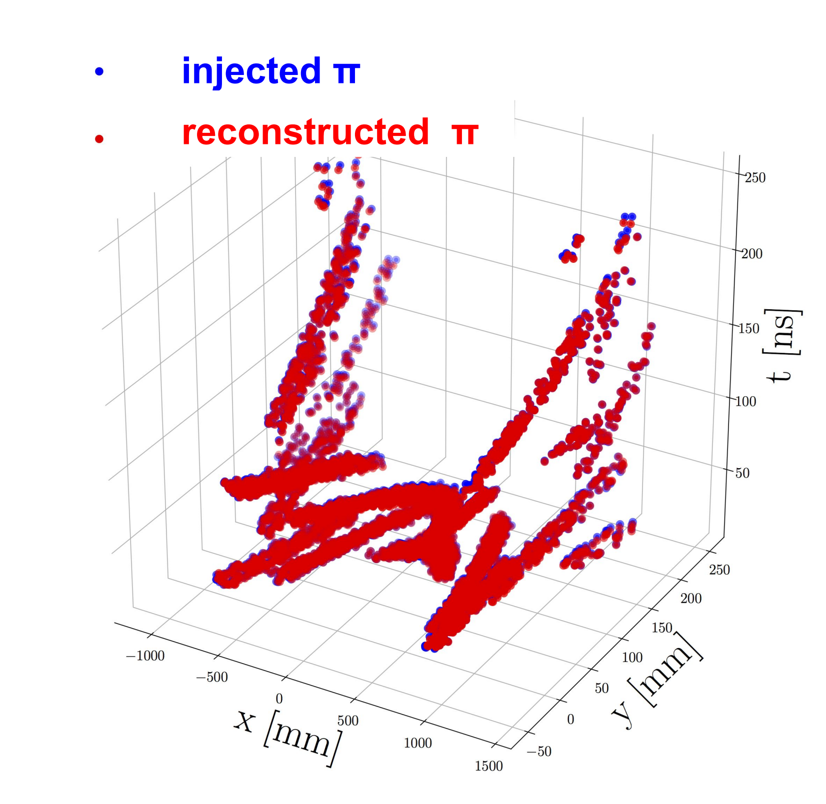
- average reconstruction loss
- cross-entropy for classification accuracy
- MMD between the distributions p(z) and q(z)

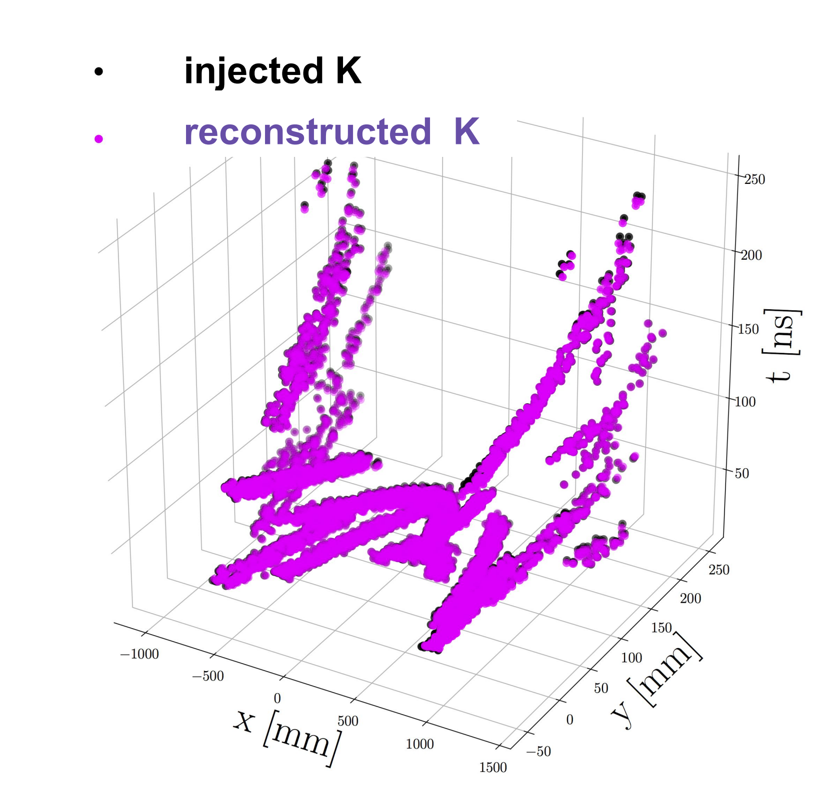
DeepRICH

CF, J. Pomponi

P, ϴ, φ = 5.0 GeV/c, 3.0 deg, 20.0 deg

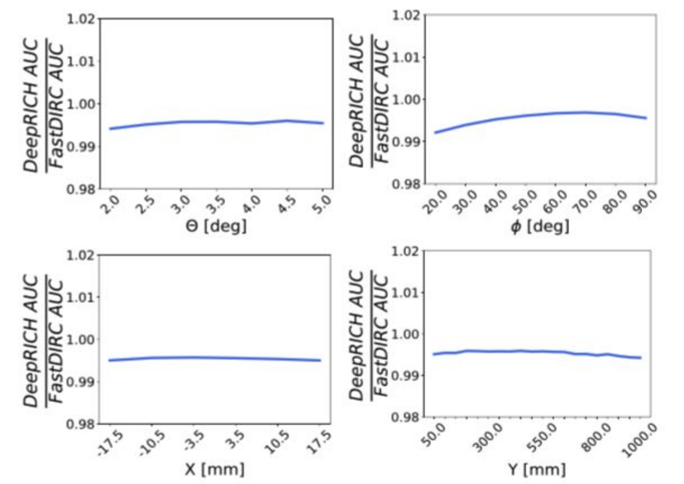
(preliminary)





DeepRICH

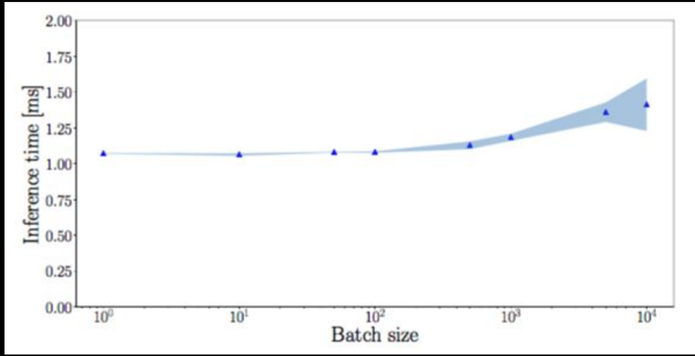
CF, J. Pomponi



- We proved that deepRICH can reach the PID performance of established algorithms. This depends only on the available resources for training.

	DeepRICH			FastDIRC		
Kinematics	AUC	ϵ_S	ϵ_B	AUC	ϵ_S	ϵ_B
4 GeV/c	99.74	98.18	98.16	99.88	98.98	98.85
4.5 GeV/c	98.78	95.21	95.21	99.22	96.33	96.32
5 GeV/c	96.64	91.13	91.23	97.41	92.40	92.47

- Remarkable reconstruction time ~1ms for a batch of 10^4 particles



More details in ArXiv 1911.11717

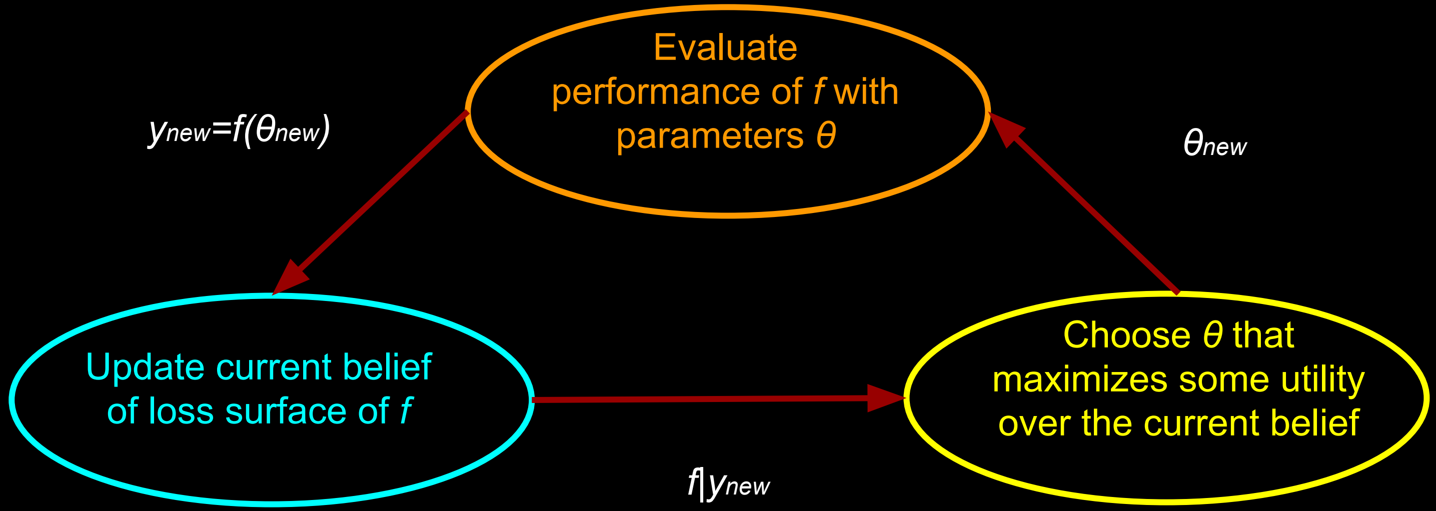
Summary

- d-RICH: we demonstrated the design is optimizable. When more realistic constraints will be available they will be implemented in BO. This can be useful in prototyping of dRICH design and any other detector.
- Global optimization techniques can be used for the GlueX DIRC expansion volume calibration with real data.
- Applied deep learning to PID for DIRC. Shown feasibility with a variational autoencoder. Potential for high performance (both in terms of reconstruction and time). Possibility to extend the architecture to fast simulation.

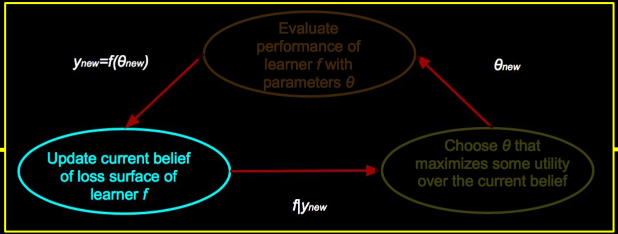
BACKUP

Bayesian Optimization

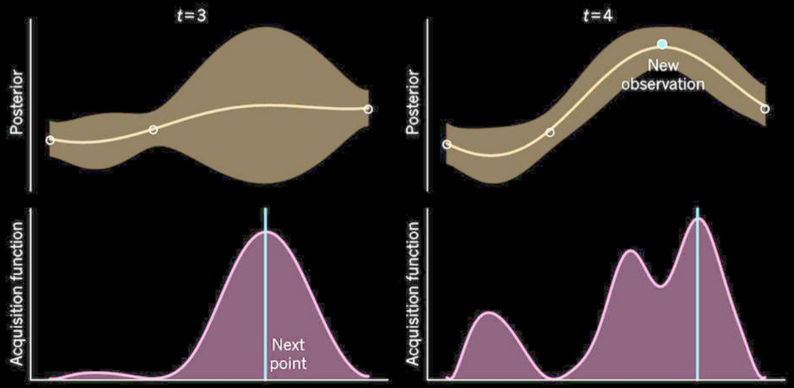
It basically consists of three steps



Update

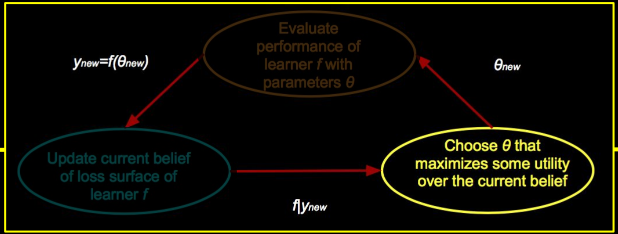


- GPs are the generalization of a Gaussian distribution to a distribution over functions, instead of random variables.
- GP is completely specified by its mean function and covariance function.

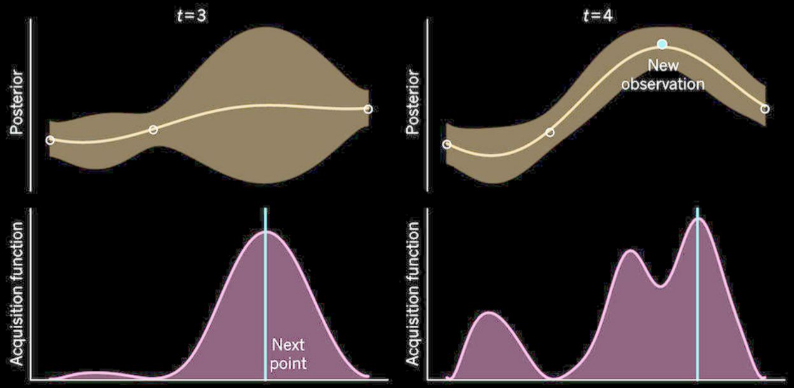


- How should I read this?
  - **Solid line**: function we are trying to min/max
  - **Shaded region:** probability model (we know the actual points already evaluated but we are more uncertain in regions where we haven't).
  - In every point a normal distribution of the potential performance function is built.

Next points



- Where am I going to sample next?
- We use utility functions called acquisition functions (formalize what is the best guess)
- Expected improvements is one example: find next point that improves the performance the most.



want to maximize

EXPLOITATION
Sample a θ with higher value than current one

EXPLORATION
Sample a point where uncertainty is high

CDF

PDF

best value we found so far

$$EI(\theta) = \begin{cases} \left(\mu(\theta) - f(\hat{\theta})\right)\Phi(Z) + \sigma(\theta)\phi(Z), & \sigma(\theta) > 0 \\ 0, & \sigma(\theta) = 0 \end{cases},$$

$$Z = \frac{\mu(\theta) - f(\hat{\theta})}{\sigma(\theta)}$$

BO

http://ash-aldujaili.github.io/blog/2018/02/01/ei/

Modelled with a Gaussian Process, the function value at a given point x can be considered as a normal random variable with mean μ and variance σ^2. Given the best (minimum in a minimization setup) function value obtained so far-let's denote it by f^*:

we are interested in quantifying the improvement over f^* we will have if we sample a point x. Mathematically, the improvement at x can be expressed as follows

$$I(x) = \max(f^* - Y, 0)$$

where Y is the random variable $\sim \mathcal{N}(\mu, \sigma^2)$ that corresponds to the function value at x. Since I is a random variable, one can consider the average (expected) improvement (EI) to assess x:

$$EI(x) = E_{Y\sim\mathcal{N}(\mu,\sigma^2)}[I(x)]$$

With the reparameterization trick, $Y = \mu + \sigma\epsilon$ where $\epsilon \sim \mathcal{N}(0, 1)$, we have:

$$EI(x) = E_{\epsilon\sim\mathcal{N}(0,1)}[I(x)]$$

which can be written as (from linearity of integral, and the definition of $\frac{d}{d\epsilon}e^{-\epsilon^2/2}$ derivative)

$$EI(x) = \int_{-\infty}^{\infty} I(x)\phi(\epsilon)d\epsilon$$

$$EI(x) = \int_{-\infty}^{(f^*-\mu)/\sigma} (f^* - \mu - \sigma\epsilon)\phi(\epsilon)d\epsilon$$

$$EI(x) = (f^* - \mu)\Phi(\frac{f^* - \mu}{\sigma}) - \sigma \int_{-\infty}^{(f^*-\mu)/\sigma} \epsilon\phi(\epsilon)d\epsilon$$

$$EI(x) = (f^* - \mu)\Phi(\frac{f^* - \mu}{\sigma}) + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{(f^*-\mu)/\sigma} (-\epsilon)e^{-\epsilon^2/2}d\epsilon$$

$$EI(x) = (f^* - \mu)\Phi(\frac{f^* - \mu}{\sigma}) + \frac{\sigma}{\sqrt{2\pi}} e^{-\epsilon^2/2}\Big|_{-\infty}^{(f^*-\mu)/\sigma}$$

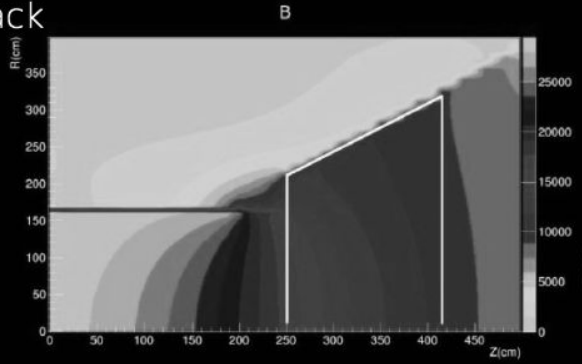
$$EI(x) = (f^* - \mu)\Phi(\frac{f^* - \mu}{\sigma}) + \sigma\left(\phi(\frac{f^* - \mu}{\sigma}) - 0\right)$$

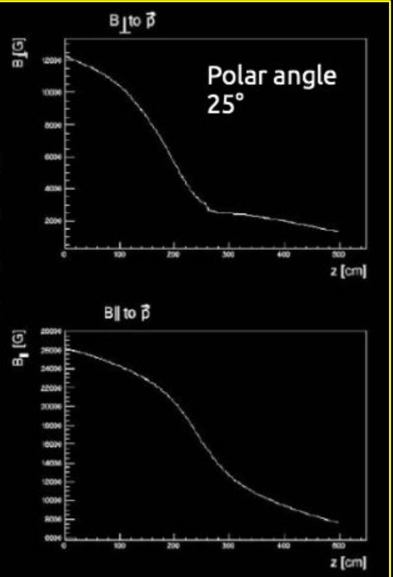
$$EI(x) = (f^* - \mu)\Phi(\frac{f^* - \mu}{\sigma}) + \sigma\phi(\frac{f^* - \mu}{\sigma})$$

Field Effects

Smearing from field perpendicular to the track affects the Cherenkov angle (Ring) resolution

- **Can be suppressed by active shaping of the field**





Related issues:

- Cost and space for adapting the magnet
- Effect of the field on the photo-detector

  Indeed the choice of the photo-sensors will be driven by magnetic field and cost effectiveness!

A. Del Dotto, EICUG2017

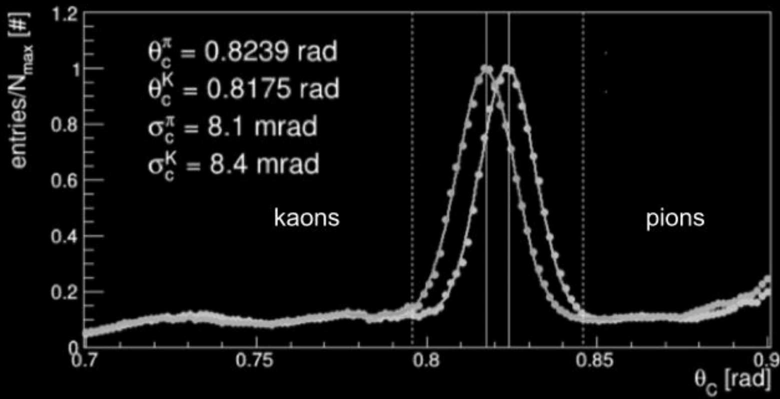
(GlueX) DIRC Reconstruction Algorithms

LUT-based geometrical

R. Dzhygadlo et al. Nucl. Instr. And Meth. A, 766:263 (2014)

1. Creation of the LUT: store directions at the end of the radiator for each hit pixel

2. Direction from the LUT for the hit pixels are combined with the track directions (from tracking)

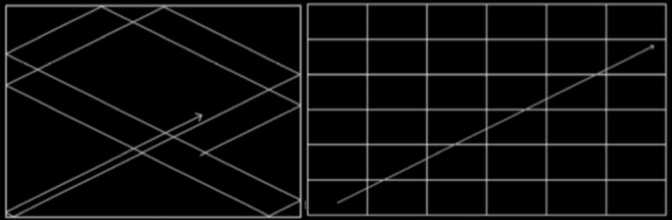


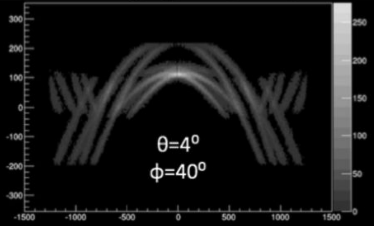
KDE-based

J. Hardin and M. Williams, JINST 11.10 (2016)

Fast tracing mapping straight lines through a tiled plane

1. Generation - 2. Traces through bars - 3. Traces through expansion volume





$$P(x) \approx \sum_{i}^{n} K(x - s_i)$$

https://github.com/jmhardin/FasDIRC

basically a trade-off memory/CPU usage

faster reconstruction/hit pattern

better resolution in regions with high overlap

Hit Patterns

DIRC rings for π^+ plotted with time on the z-axis.

Credits:
J. Hardin, PhD thesis

- 3D (x,y,t) readout and this allows to separate spatial overlaps.
- Patterns take up significant fractions of the PMT in x,y and are read out over 50-100 ns due to propagation time in bars.
- H12700 PMTs have a time resolution of O(500 ps) and read-out electronics giving time information in 1 ns buckets.

J. Hardin and M. Williams, JINST 11.10 (2016)