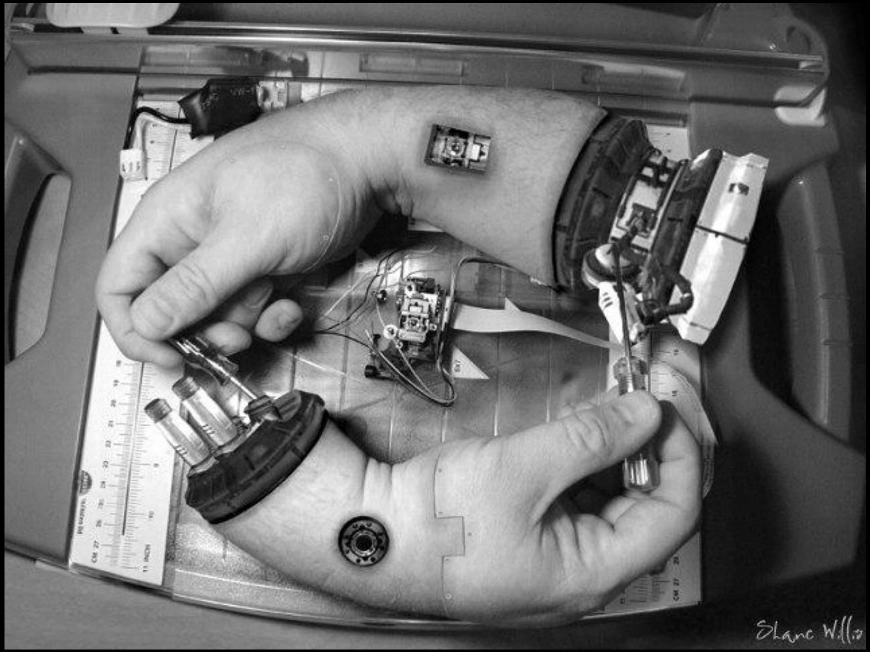
Subdetector Parameter Optimization

EIC²
EIC Center at Jefferson Lab

GlueX

Jefferson Lab



C. Fanelli

Outline

Subdetector Parameter Optimization

Detector Design

- Motivation
- Practical Example
  - Theoretical Framework
  - Implementation
    - Toy model and code snippet

→ https://repl.it/@cfanelli2/driver#main.py

(Q/A)

(Q/A)

Calibration/Alignment

- Applications

Vision

Q/A

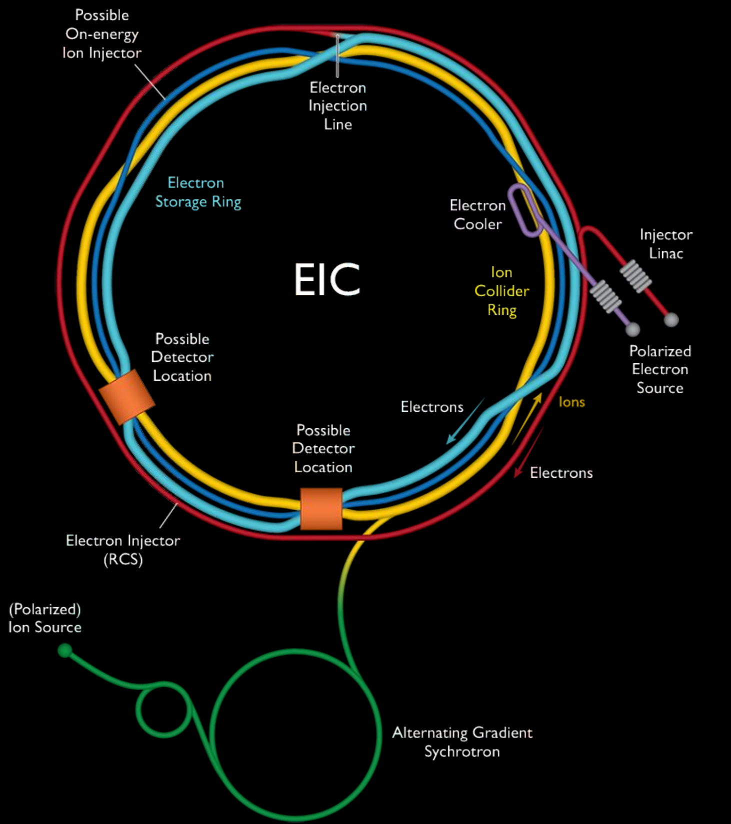
Focus: Methods/ Procedures/ Applications

Experiments: EIC, GlueX

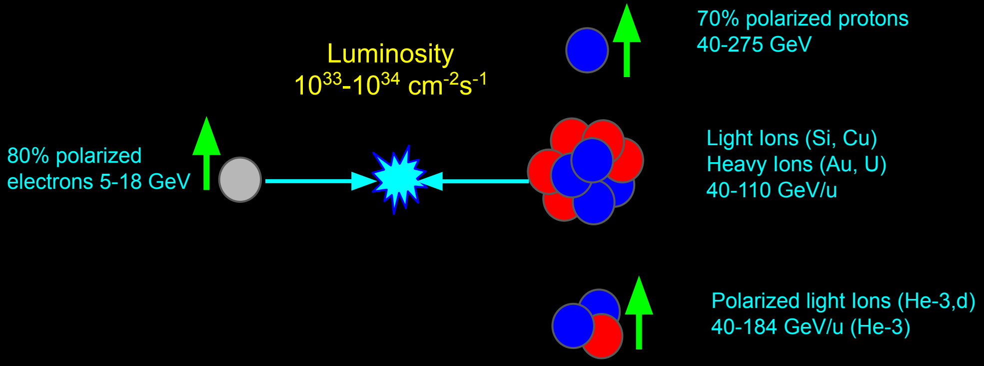
Detector Design: Intro

- Particle detectors are essential tools to understand the universe and they play a crucial function in our society, from medical imaging to drug development, from dating ancient findings to testing materials properties.
- Fundamental nuclear and particle physics research often requires realizing expensive large-scale experiments combining multiple sub-detectors to investigate the building blocks of nature. According to the DOE, "*AI techniques that can optimize the design of complex, large-scale experiments have the potential to revolutionize the way experimental nuclear and particle physics is currently done*".
- Surprisingly at present few AI-based approaches (and generally procedural methods) for designing particle detectors have been explored/developed.
- More than 50 years have passed since Charpak (nobel prize in 1992) revolutionised particle detectors with the construction of a MWPC. Nowadays we have the unique opportunity to design complex detection systems with the support of AI.
- The Electron Ion Collider will be a flagship nuclear physics facility in the US that will be constructed over the next 10 years and it is currently at its design phase. Its R&D program can be one of the first to systematically leverage on AI.

EIC@BNL



Will be constructed over ten years at an estimated cost between \$1.6 and \$2.6 billion



- Two intersecting accelerators, one producing a beam of electrons, the other a high-energy beam of protons or heavier atomic nuclei
- Wide coverage of CoM energy $\sqrt{s_{e\text{-}p}}$ ~ (20-140) GeV
- Two large acceptance detectors

A machine for delving deeper than ever before into the building blocks of matter

EIC scientific program in a nutshell

Emergence of Mass

- Nucleons: 99% of the mass of the visible universe
- How does the proton mass emerge from QCD, and why is it so heavy?
- What is the mechanical structure of the proton?

Nucleon Spin and Imaging

- Full map of nucleon spin structure and dynamics in momentum and position space
- Towards a comprehensive 5D picture of the quantum structure of the proton

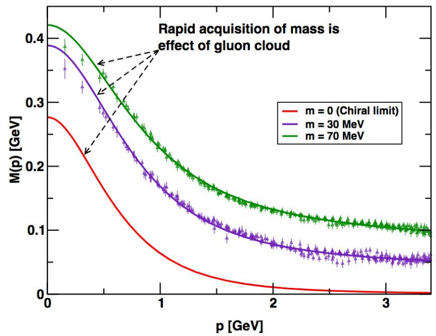
Physics with high-energy nuclear beams

- Saturation: Is the high-energy/low-x limit governed by a universal dense saturated gluon matter?

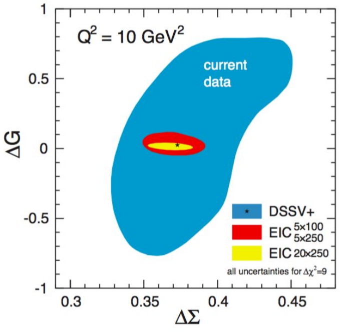
Other Topics

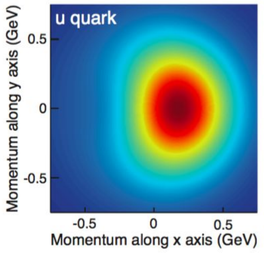
- Confinement, Hadronization
- Passage of color charge through cold QCD matter
- Jet Physics in ep/eA collisions
- BSM
- etc

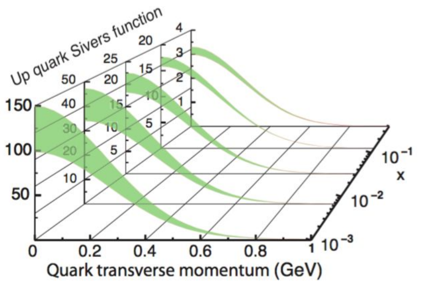
C.D. Roberts. Prog.Part.Nucl.Phys..61:50–65. 2008

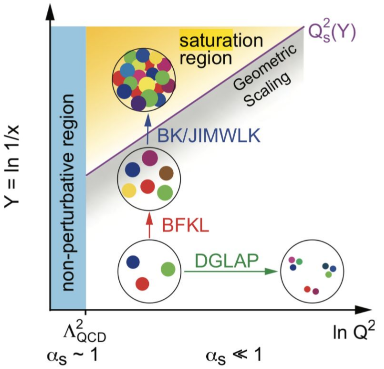


Accardi, A. *et al* - BNL-98815-2012- JAJLAB-PHY-12-1652arXiv:1212.1701



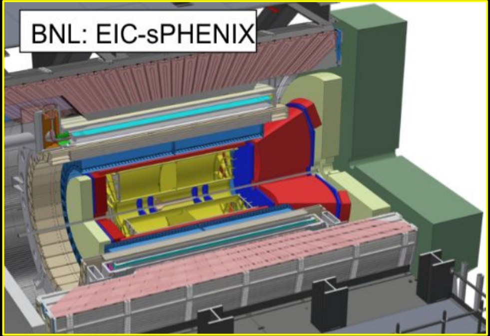


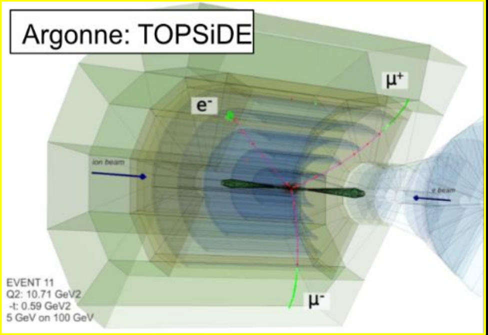


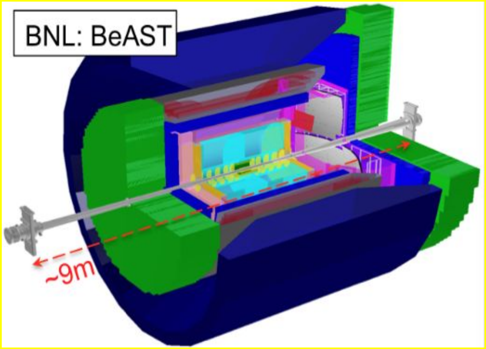


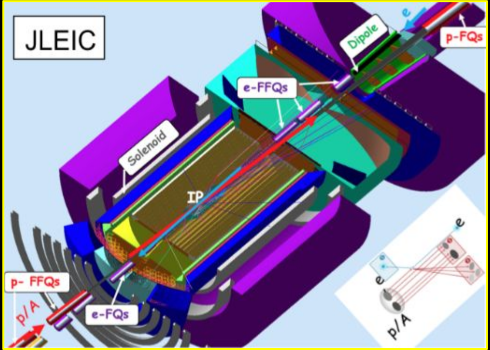
Towards a “Handbook Detector” for EIC

At the beginning of the 2020 we had different concepts...

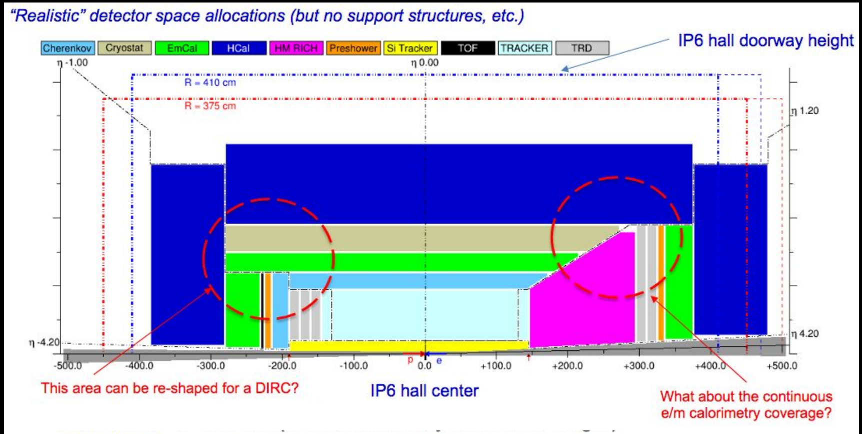








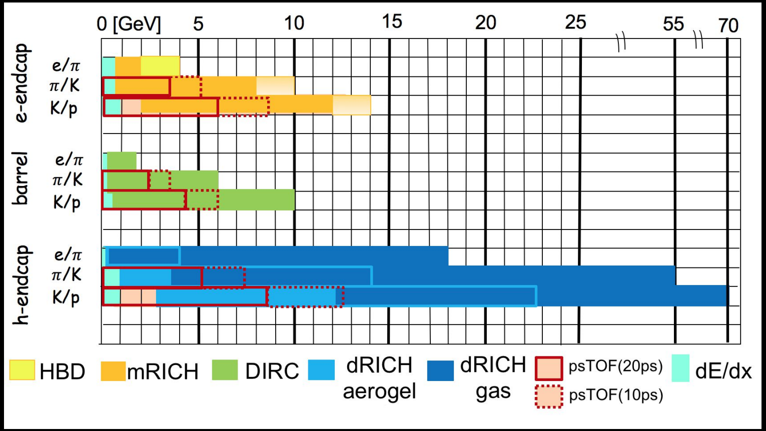
- After CD-0, the project rapidly evolved as the EIC community is working on the EIC Yellow Book to have a general purpose "Reference Detector” for BNL.
- See recent talks by A. Kiselev and W. Akers at the 3rd Yellow Report workshop at CUA

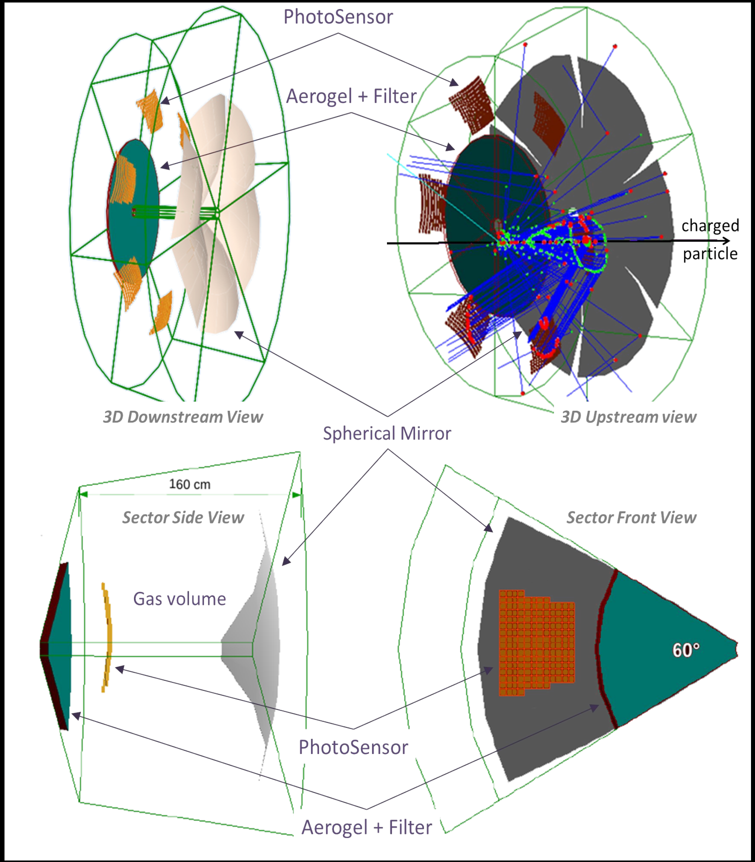


Dual RICH: case study

E. Cisbani, A. Del Dotto, CF*, M. Williams et al.
"AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case."
Journal of Instrumentation 15.05 (2020): P05009.

aerogel (4 cm, n(400 nm): 1.02)
+ 3 mm acrylic filter
+ gas (1.6 m, n(C_2F_6): 1.0008)





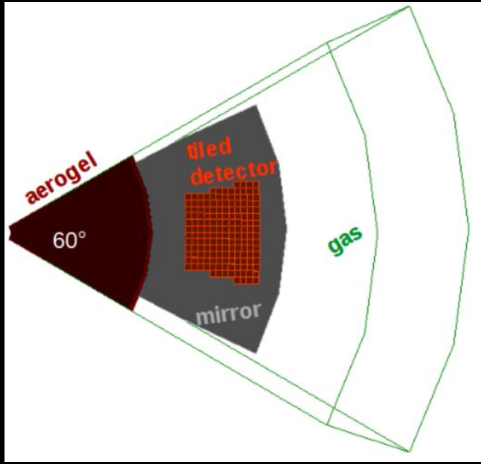
- Continuous momentum coverage.
- Simple geometry and optics, cost effective.
- Legacy design from INFN, see EICUG2017

- 6 Identical open sectors (petals)
- Optical sensor elements: 8500 cm^2/sector, 3 mm pixel
- Large focusing mirror

Construction Constraints on Design Parameters

The idea is that we have a bunch of parameters to optimize that characterize the detector design.
We know from previous studies their ranges and the construction tolerances.

parameter	**description**	**range** [units]	**tolerance** [units]
R	mirror radius	[290,300] [cm]	100 [μm]
pos r	radial position of mirror center	[125,140] [cm]	100 [μm]
pos l	longitudinal position of mirror center	[-305,-295] [cm]	100 [μm]
tiles x	shift along x of tiles center	[-5,5] [cm]	100 [μm]
tiles y	shift along y of tiles center	[-5,5] [cm]	100 [μm]
tiles z	shift along z of tiles center	[-105,-95] [cm]	100 [μm]
$n_{aerogel}$	aerogel refractive index	[1.015,1.030]	0.2%
$t_{aerogel}$	aerogel thickness	[3.0,6.0] [cm]	1 [mm]



Ranges depend mainly on mechanical constraints and optics requirements.
These requirements can change in the next future based on inputs from prototyping.

Choice of Figure of Merit

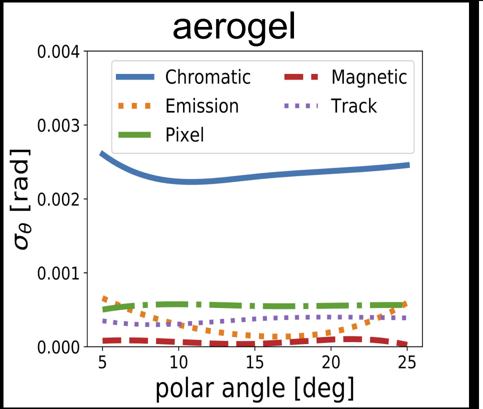
Goal is improve the distinguishing power of pions/kaons, hence:

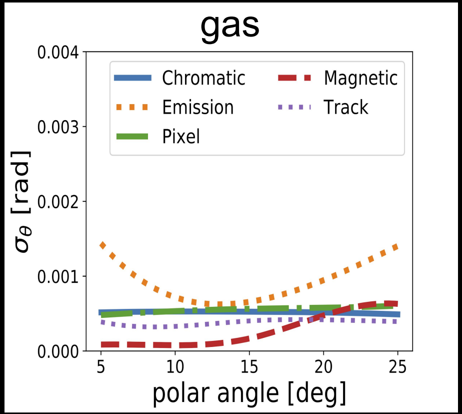
$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

$$N_\gamma = (N_\gamma^\pi + N_\gamma^K)/2$$

$$h = 2 \cdot \left[\frac{1}{(N\sigma)|_1} + \frac{1}{(N\sigma)|_2} \right]^{-1}$$

Main contributions to resolution





Remember that we do not have an explicit form of the FOM we are trying to optimize as a function of the design parameters

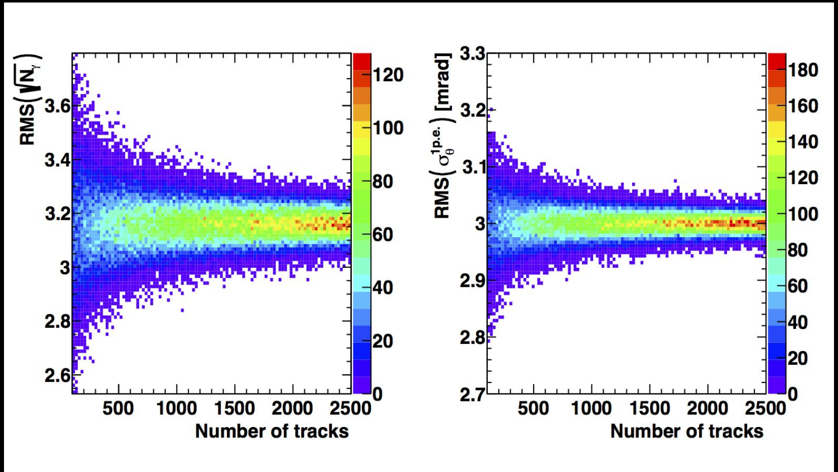
@ p_1 = 14 GeV/c (aerogel) and p_2 = 60 GeV/c (gas) considering the two parts disentangled

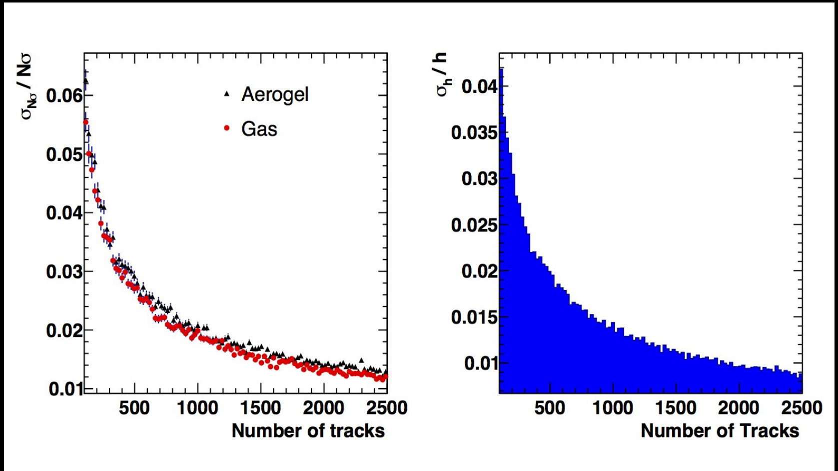
Noise Studies

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

- Dedicated studies to characterize the noise as this is an optimization of a noisy function

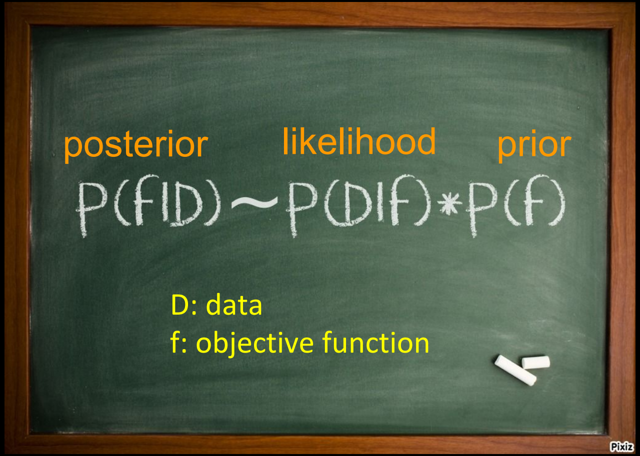
- We choose N tracks = 400 based on the studies on noise to minimize as much as possible computing time during simulation.





Bayesian Optimization

- Objective *f* is a black-box function and can be noisy.
- Evaluations are expensive making grid or exhaustive search impractical.
- *f* lacks of special structure (e.g. convex), and it has no gradient information.



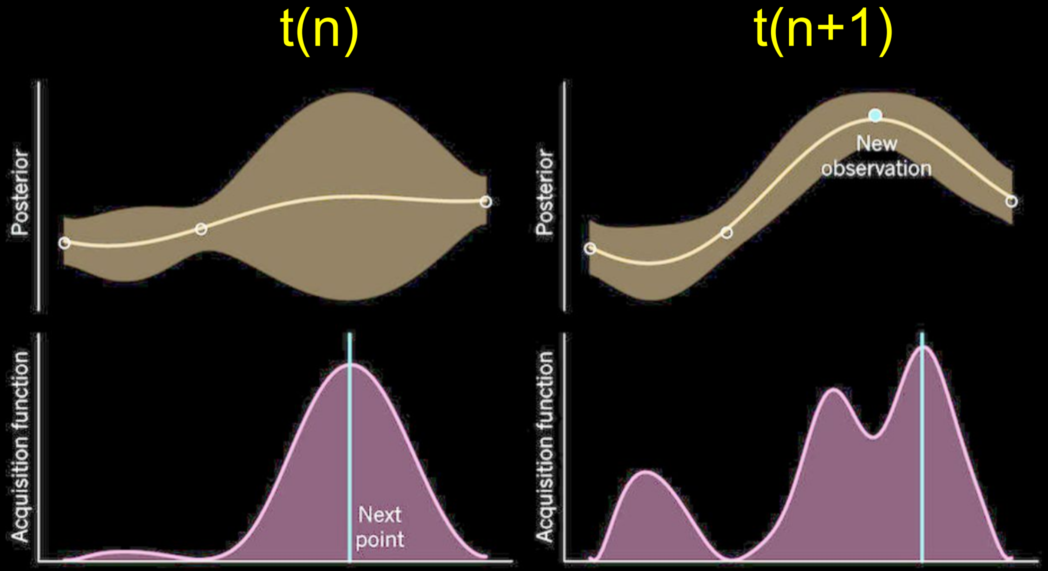
If you don't have the above constraints,
do not use Bayesian Optimization

- We want to determine the optimum of *f*, no need to improve estimates of regions where *f* is not optimal. The idea is to build a surrogate function:
  - With a Prior over the space of objective functions, to model our black-box function.
  - Likelihood ~ probability of observing the data given the function *f*.
  - The Posterior probability is the surrogate objective function. It captures the updated beliefs about the unknown objective.

https://machinelearningmastery.com/what-is-bayesian-optimization/
http://krasserm.github.io/2018/03/21/bayesian-optimization/
http://krasserm.github.io/2018/03/19/gaussian-processes/

Bayesian Optimization

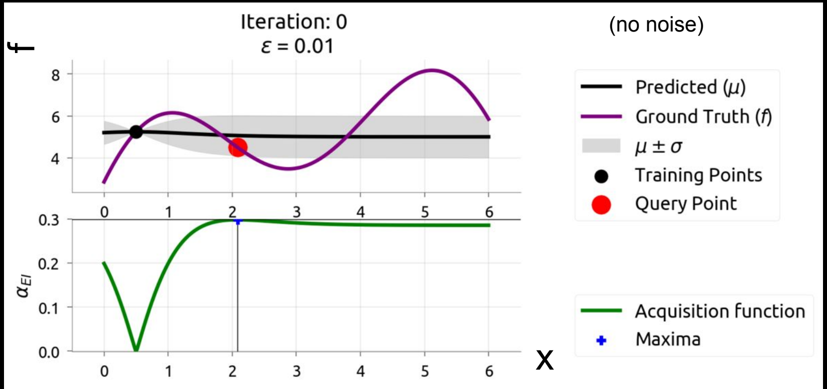
- BO is a sequential strategy developed for global optimization.
- After gathering evaluations we builds a posterior distribution used to construct an **acquisition function.**
- This cheap function determines what is **next query point.**



1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

http://krasserm.github.io/2018/03/21/bayesian-optimization/
http://krasserm.github.io/2018/03/19/gaussian-processes/

Acquisition Functions



Exploitation **Exploration**

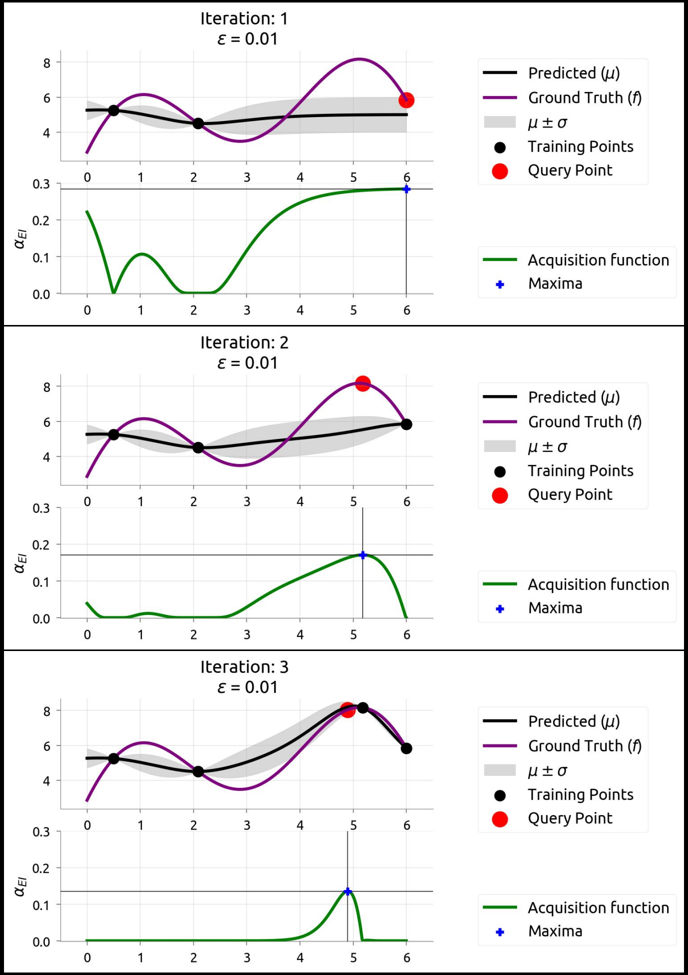
$$EI(x) = \begin{cases} (\mu_t(x) - f(x^+) - \epsilon)\Phi(Z) + \sigma_t(x)\phi(Z), & \text{if } \sigma_t(x) > 0 \\ 0, & \text{if } \sigma_t(x) = 0 \end{cases}$$

Best found so far

$$Z = \frac{\mu_t(x) - f(x^+) - \epsilon}{\sigma_t(x)}$$

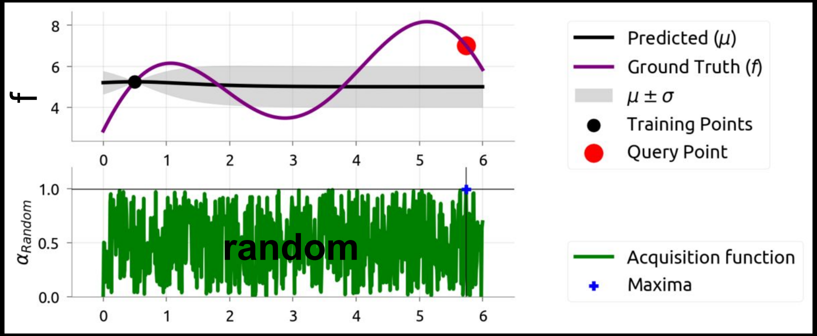
We are sampling x

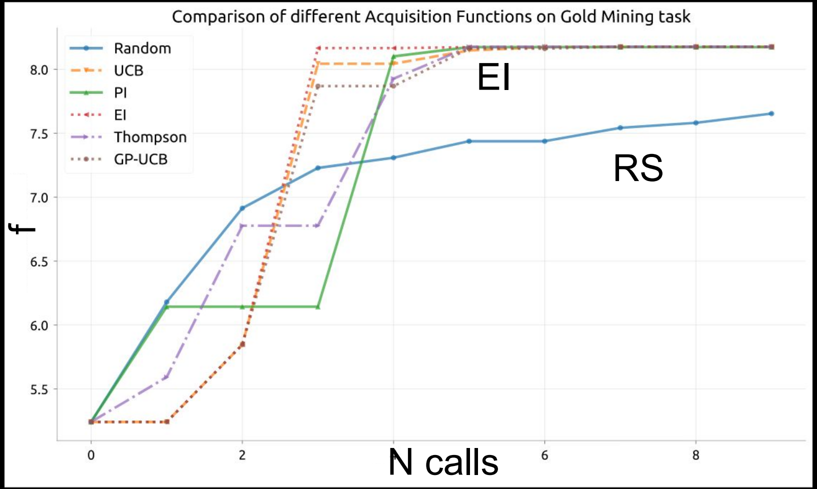
- **Exploitation**: search where μ is high
- **Exploration**: search where σ is high



Acquisition Functions

- Many acquisition functions, e.g., Probability of Improvement, Expected Improvement, Upper (Lower) confidence bound, etc
- In most cases, acquisition functions provide knobs for controlling the exploration-exploitation tradeoff
- When optimization is more complex (more dimensions), then a random acquisition might perform poorly





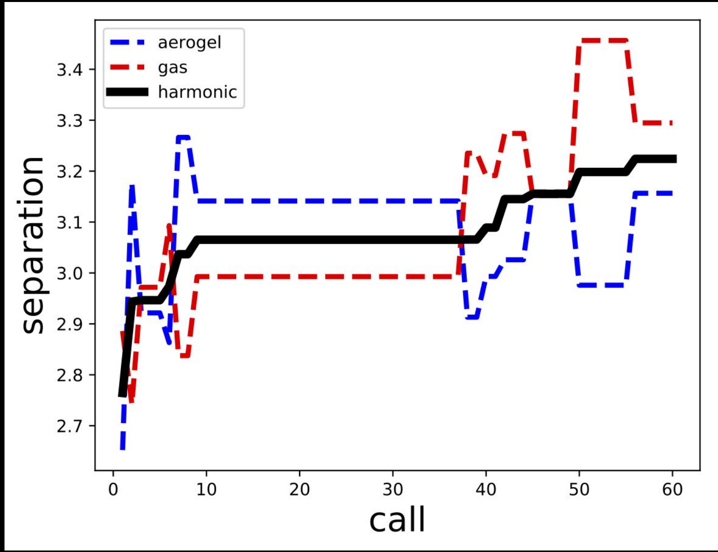
E. Brochu, Eric, V. M. Cora, and N. De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." *arXiv:1012.2599* (2010).

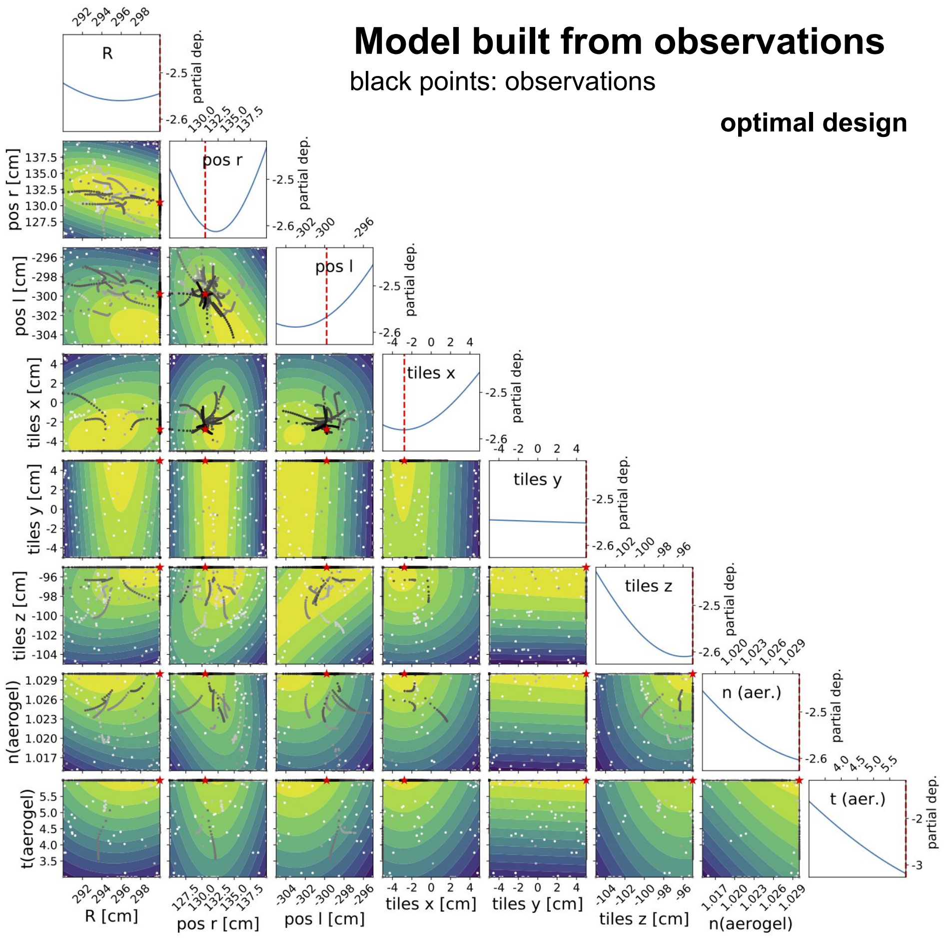
https://distill.pub/2020/bayesian-optimization/
https://distill.pub/2019/visual-exploration-gaussian-processes/

Question(s)?

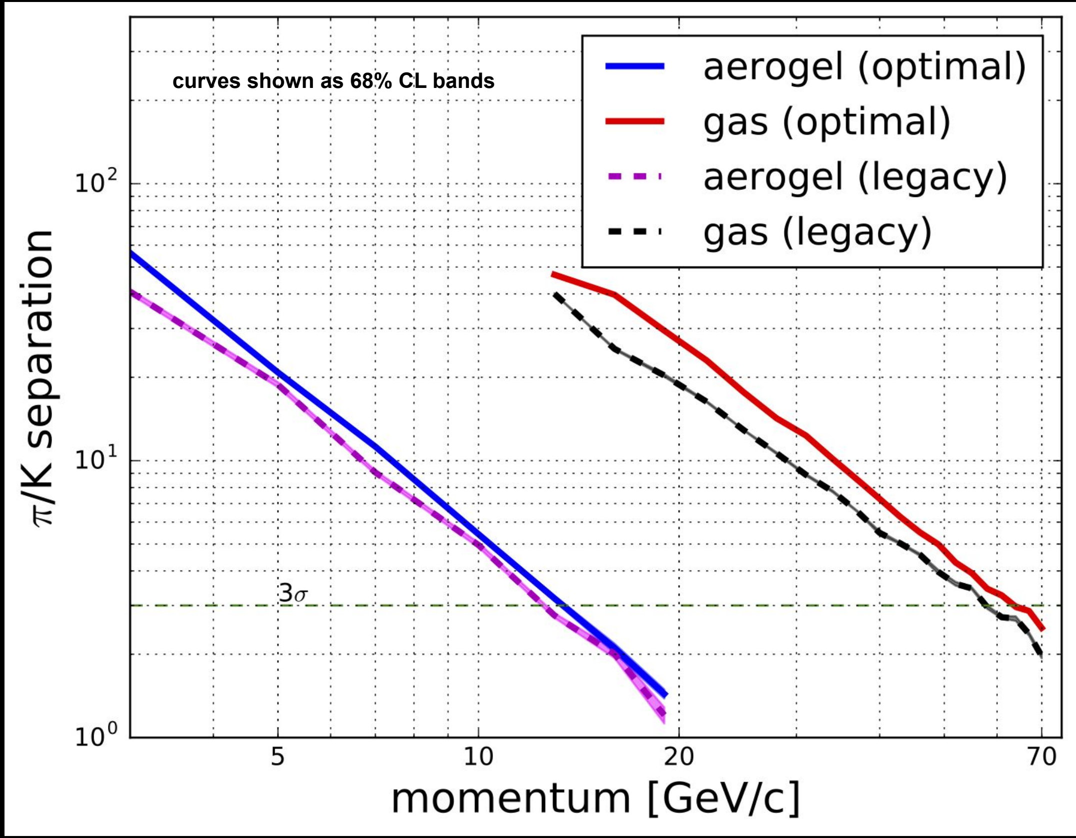
The Model and the Optimized FoM

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





dRICH Performance @ the optimal design point

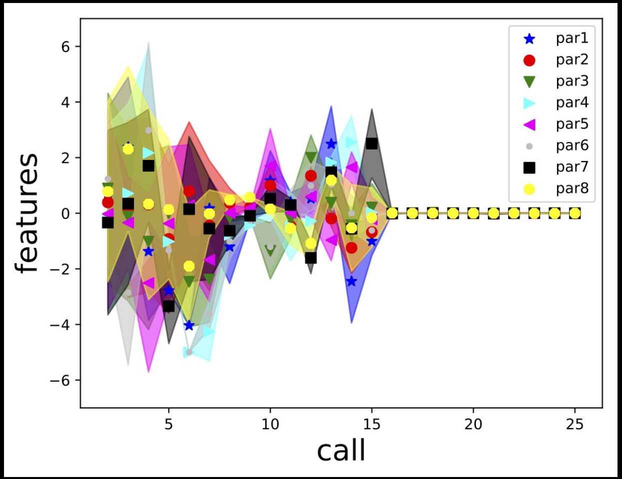


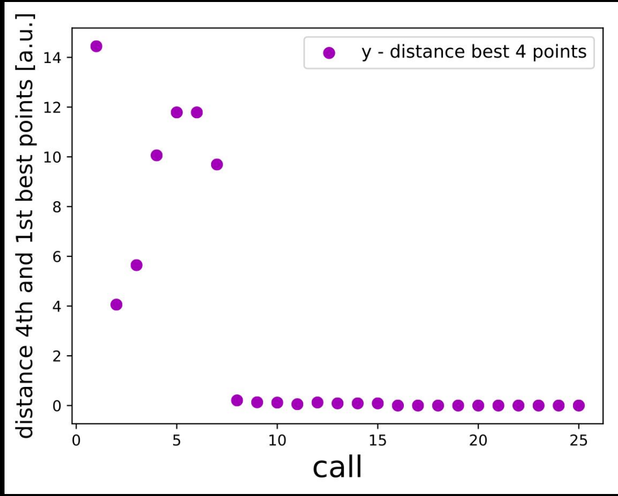
- Statistically significant Improvement in both parts.
- In particular in the gas region where the 5σ threshold shifted from 43 to 50 GeV/c and the 3σ one extended up to
- Notice that before this study we did not know "how well" the legacy design was performing.

E. Cisbani, A. Del Dotto, CF*, M. Williams et al.
JINST 15.05 (2020): P05009.

Convergence Criteria

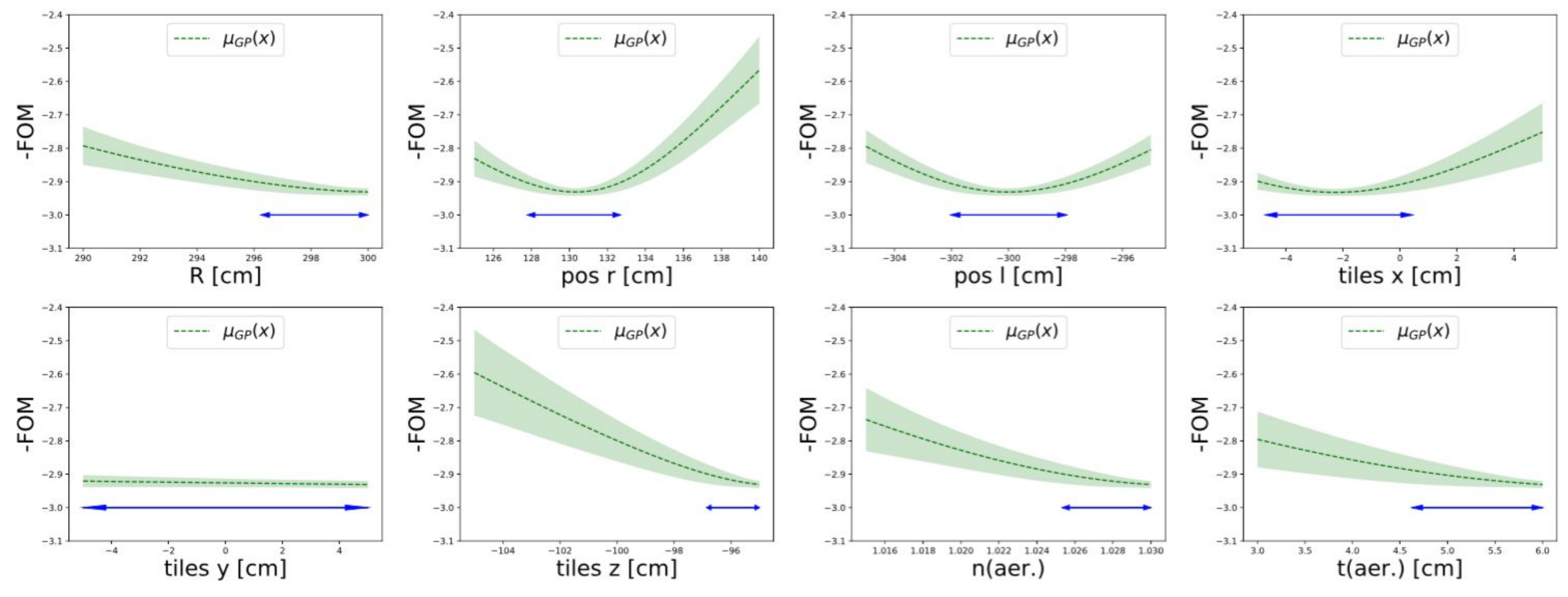
- We defined a set of conditions to ensure convergence.
- These correspond to the logic AND of booleans on each feature and on the variation of the figure of merit.
- They are built on standardized Z and Fisher statistics.
- Pre-processing of data required to remove outliers.





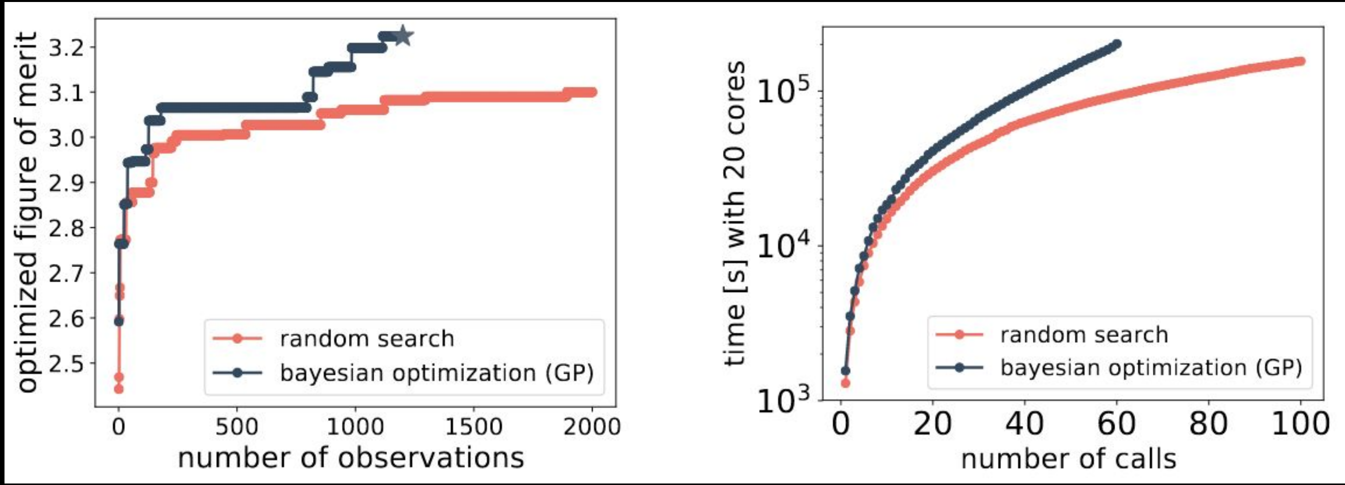
Tolerance Regions

- BO provides a model of how the FoM depends on the parameters, hence it is possible to use the posterior to define a tolerance on the parameters (regions ensuring improved PID, see previous slide).



- Larger than the construction tolerances on each parameter.
Notice a small lateral shift of the tiles has negligible impact on the PID capability.

Comparison with Random Search



Each call:
400 tracks generated/core
20 cores

1 design point ~ 10 mins/CPU

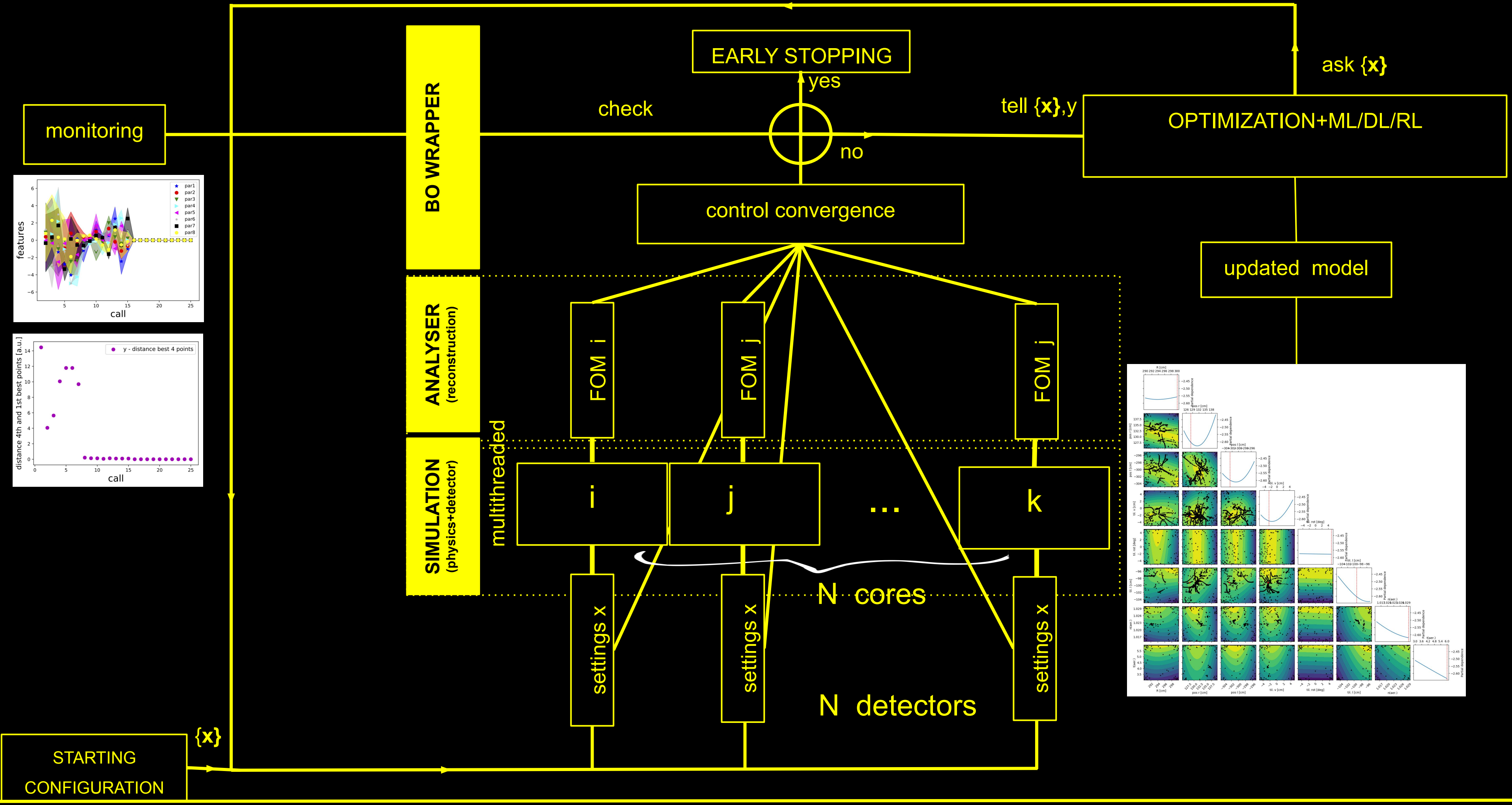
Budget: 100 calls

- BO with GP scales cubically with number of observations.

- Bayesian optimization methods are more promising because they offer principled approaches to weighting the importance of each dimension.

- For this 8D problem - even with 50 cores, RS looks unfeasible due to the curse of dimensionality.
  - Recall that the probability of finding the target with RS is $1-(1-v/V)^T$, where T is trials, v/V is the volume of target relative to the unit hypercube

Bergstra, Bengio, “Random search for hyper-parameter optimization”, J. Mach. Learn. Res.13 (Feb) (2012) 281–305.



EARLY STOPPING
yes
check
tell {x},y
ask {x}
OPTIMIZATION+ML/DL/RL
monitoring
BO WRAPPER
no
control convergence
updated model
ANALYSER
(reconstruction)
FOM i
FOM j
FOM j
SIMULATION
(physics+detector)
multithreaded
i
j
...
k
N cores
settings x
settings x
settings x
N detectors
{x}
STARTING
CONFIGURATION
features
call
par1
par2
par3
par4
par5
par6
par7
par8
y - distance best 4 points
distance 4th and 1st best points [a.u.]
call

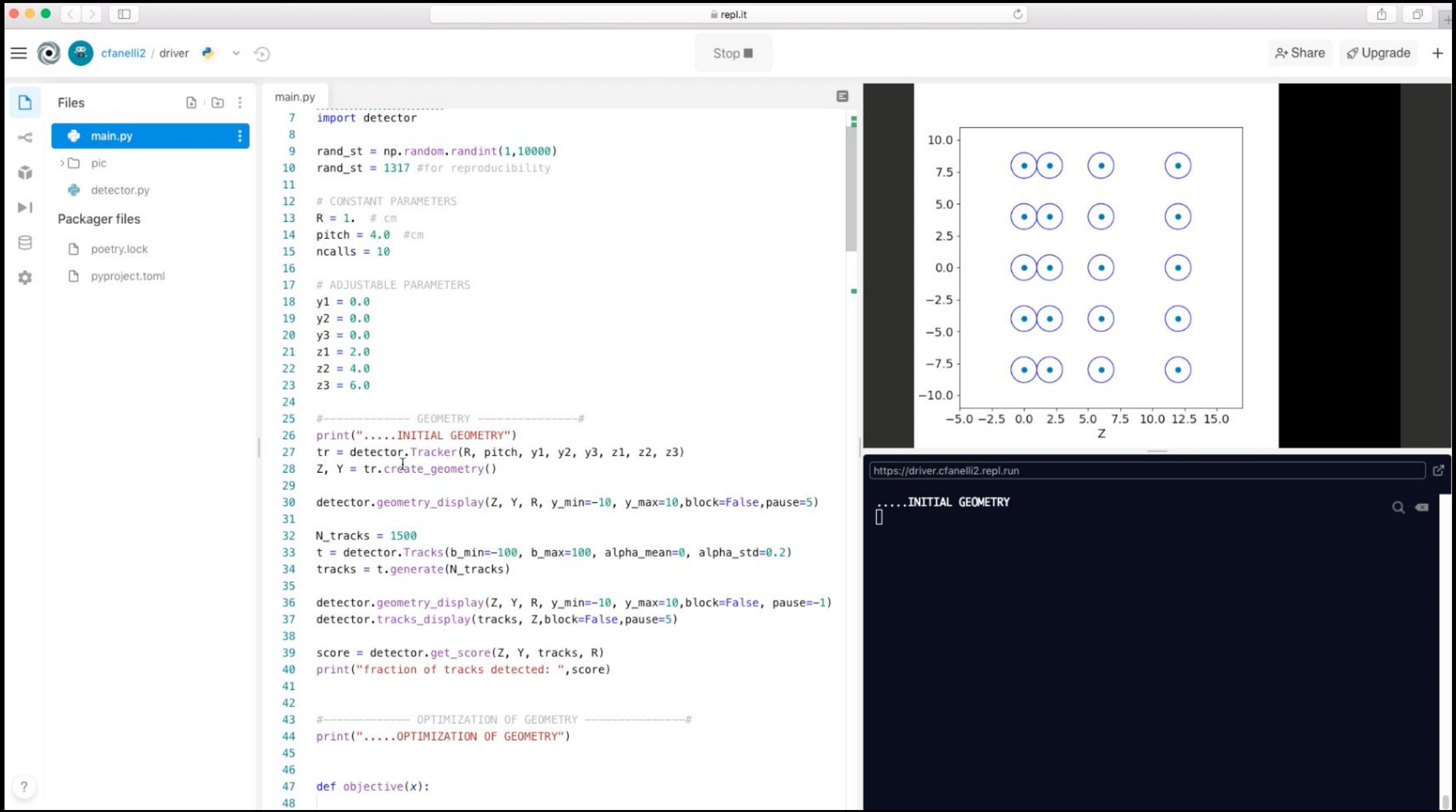
Frameworks and Deployment in the Industry

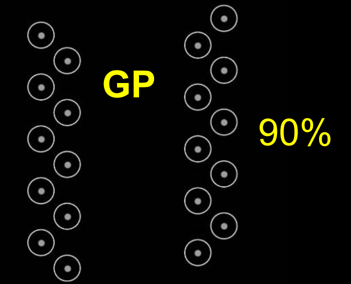
- scikit-optimize
- sigopt
- hyperopt
- spearmint
- MOE
- BOTorch
- GPFlowOpt
- GPyOpt
- DragonFly
- Hyperband
- Smac
- etc

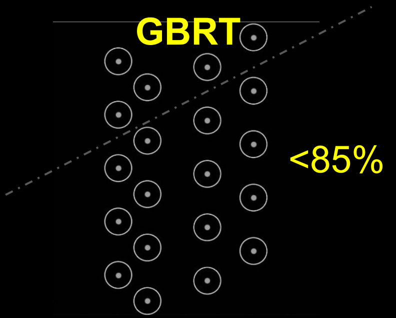
- Bayesian Optimization has been applied to Optimal Sensor Set selection for predictive accuracy.
- Uber uses Bayesian Optimization for tuning algorithms via backtesting.
- Facebook uses Bayesian Optimization for A/B testing.
- Netflix and Yelp use Metrics Optimization software like Metrics Optimization Engine (MOE) which take advantage of Parallel Bayesian Optimization.

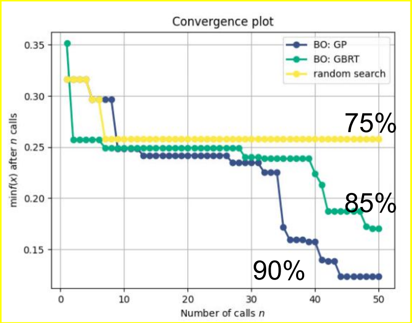
Detector Toy Model

https://repl.it/@cfanelli2/driver#main.py



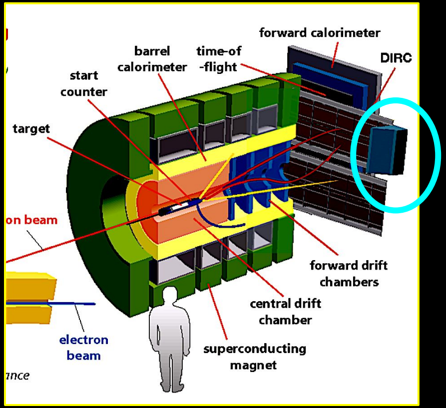




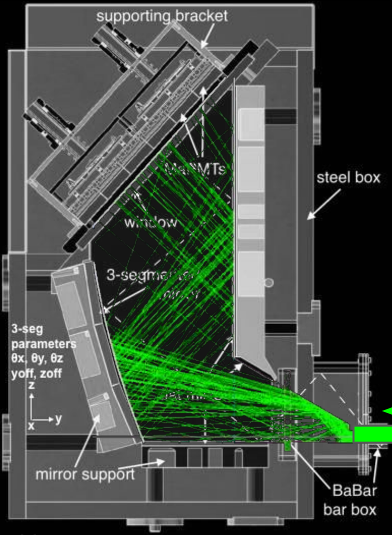


Question(s)?

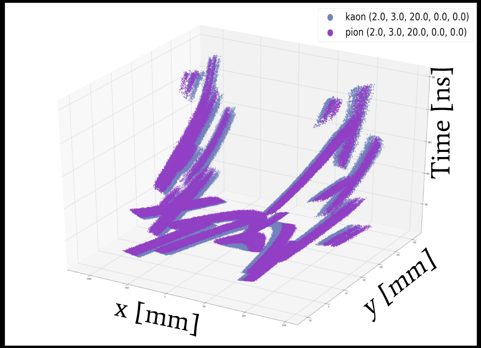
GlueX DIRC Alignment



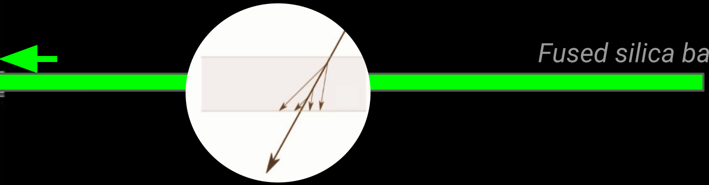
GlueX View



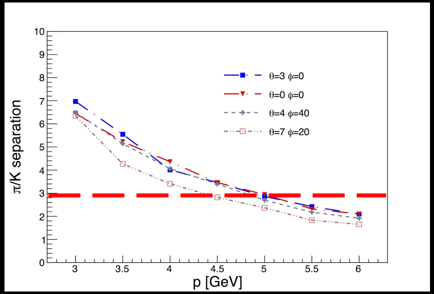
Optical box



3D Readout



Fused silica bars



π/K separation with DIRC

3D (x,y,t) readout allows to separate spatial overlaps.

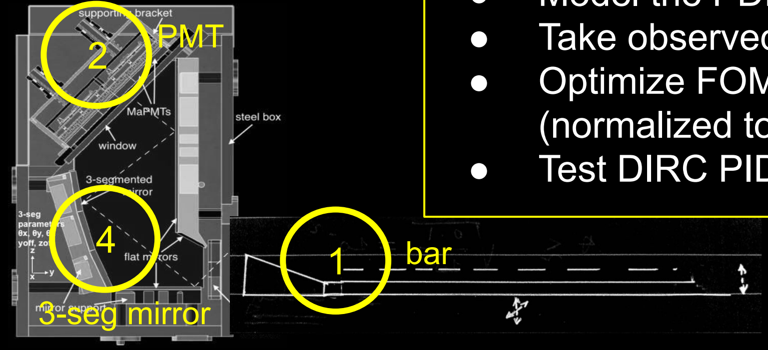
Patterns take up significant fractions of the PMT in x,y and are read out over 50-100 ns due to propagation time in bars.

H12700 PMTs have a time resolution of O(200 ps) and read-out electronics giving time information in 1 ns buckets.

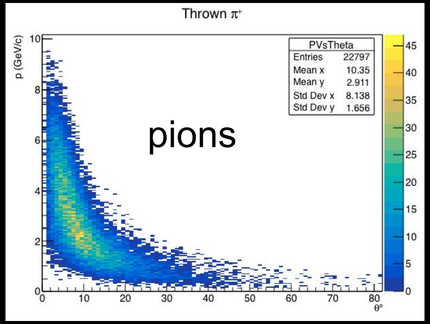
Approach

CF with DIRC group @ MIT

Main alignment parameters



- Select high purity sample of particles at low P (well identified by GlueX PID w/o DIRC)
- Model the PDF as a function of the offsets
- Take observed hits to build Likelihood
- Optimize FOM = logL (normalized to a default alignment)
- Test DIRC PID on larger momentum P



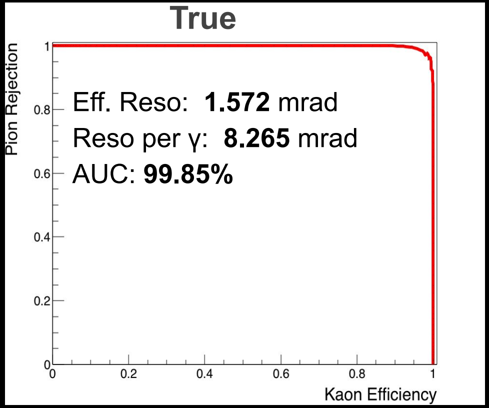
True:

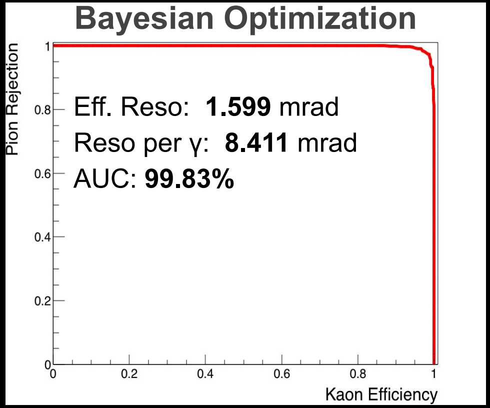
3-seg mirror:
θx,θy,θz=(0.25,0.50,0.15) deg,
y = 0.50 mm;
bar: z = 2.00 mm;
PMT: (r,θ)=(1.50 mm,1.00 deg)

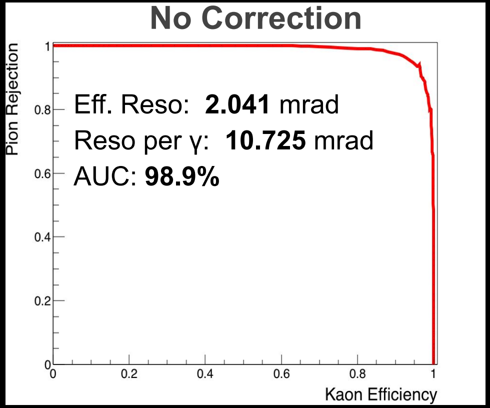
BO-reversed engineered:

3-seg mirror:
θx,θy,θz=(0.25, 0.58, 0.12) deg,
y = 0.59 mm;
bar: z = 2.08 mm;
PMT: (r,θ)=(1.87 mm, 1.35 deg)

Pion rejection vs Kaon efficiency at large P



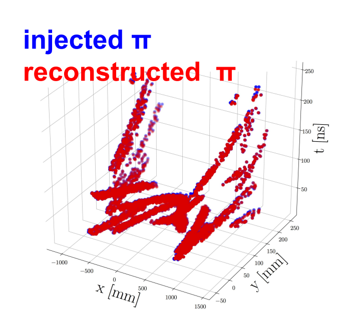


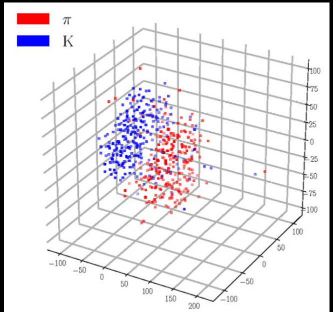


Hyperparameters tuning: DeepRICH

CF and J. Pomponi

DeepRICH

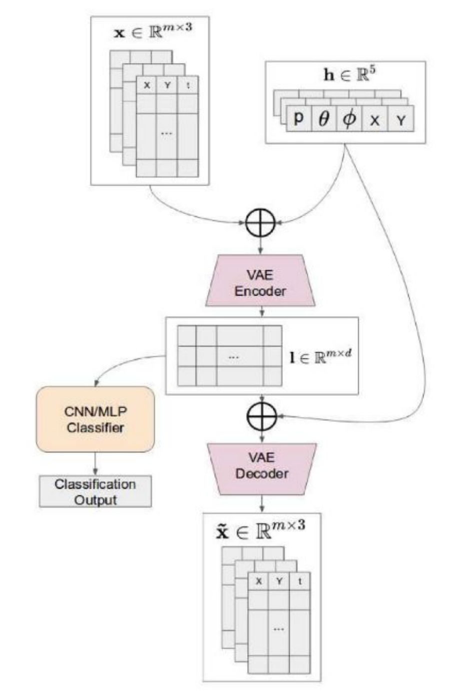




latent space

t-SNE used for 3D visualization

injected



reconstructed

CF and J. Pomponi. *Machine Learning: Science and Technology* 1.1 (2020): 015010.

Hyperparameters

Table 2. List of hyperparameters tuned by the BO. The tuned values are shown in the outermost right column. The optimized test score is about 92%.

symbol	description	range	optimal value
NLL	λ_r	$[10^{-1},10^2]$	0.784
CE	λ_c	$[10^{-1},10]$	1.403
MMD	λ_v	$[1,10^3]$	1.009
LATENT_DIM	latent variables dimension	[10,200]	16
var_MMD	σ in $\mathcal{N}(0,\sigma)$	[0.01,2]	0.646
Learning Rate	learning rate	[0.000 1, 1]	$6.6\cdot10^{-4}$

DeepRICH Performance

Table 3. The area under curve (%), the signal efficiency to detect pions ε_S and the background rejection of kaons ε_B corresponding to the point of the ROC that maximizes the product ε_S*ε_B. The corresponding momenta at which these values have been calculated are also reported. This table is obtained by integrating over all the other kinematic parameters (i.e. a total of ~6k points with different θ, ϕ, X, Y for each momentum).

	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

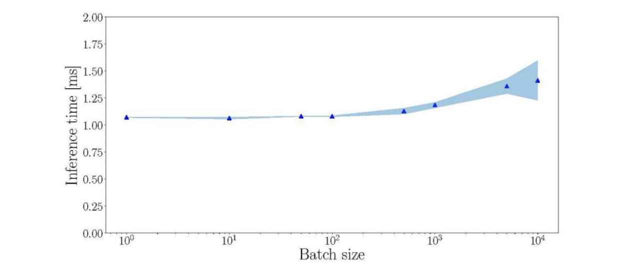
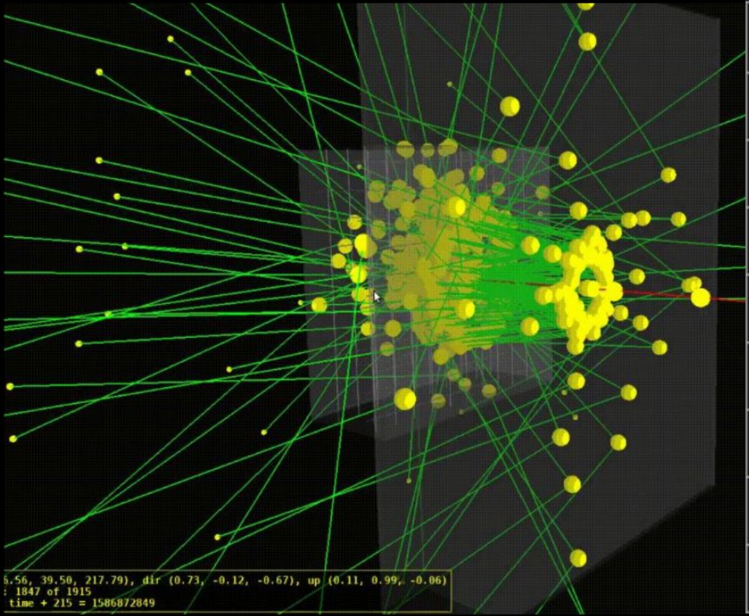


Figure 9. After training, the inference time is almost constant as a function of the batch size, meaning that the effective inference time—*i.e.*, the reconstruction time per particle—can be lower than a μs, the architecture being able to handle 10^4 particles in about 1.4 *ms* in the inference phase. Notice that the corresponding memory size in the inference phase is approximately equal to the value reported in table 4.

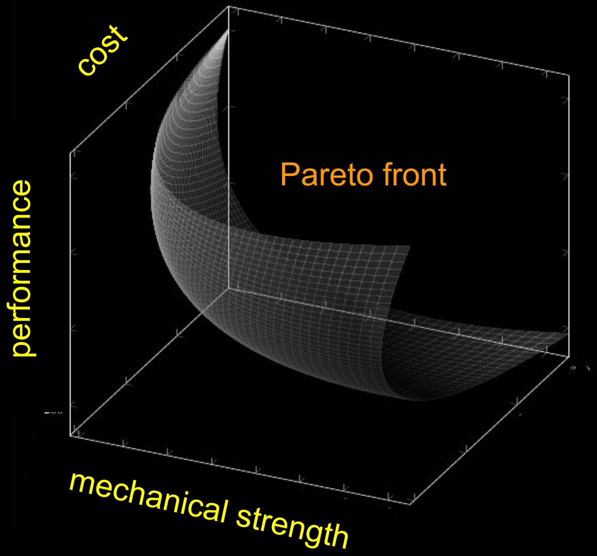
Towards Multi-objective Optimization

V. Berdnikov, E. Cisbani, J. Crafts, CF,
T. Horn, I. Pegg, R. Trotta

- Ongoing EIC project: aerogel endowed with planes of fibers for mechanical stability.
- BO-assisted preliminary design will be extended to a multi-objective (resolution, mechanical strength, cost) design optimization problem with constraints.
- Several approaches on the market, genetic algorithms, bayesian, reinforcement learning .



Detector Part	Parameter
aerogel	thickness
aerogel	width
aerogel	refraction index
fiber	diameter
fiber	pitch
fiber	gap
sensor plane	distance
sensor plane	size



“Taking the Human out of the Loop”…

B. Shahriari, et al. *Proceedings of the IEEE* 104.1 (2015): 148-175.

Bayesian Optimization

DL-enhanced

Multi-objective

Evolutionary autoML

Meta-learning

Reinforcement Learning

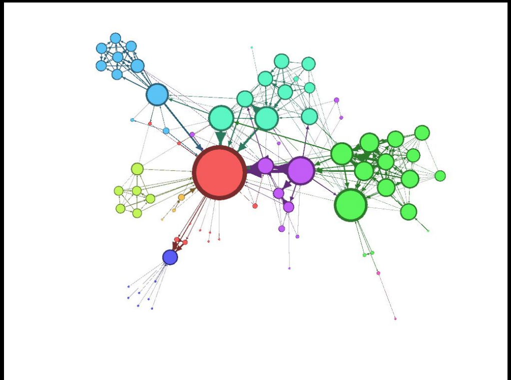
Accelerated Discovery

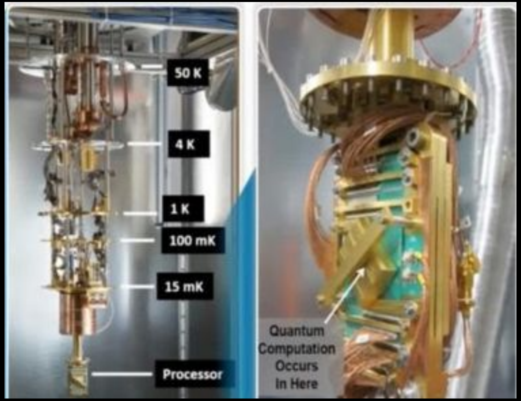
- Optimal Design
- Inverse Design
- Self-Design
- Calibration/ Alignment
- Self-Calibration
- etc.

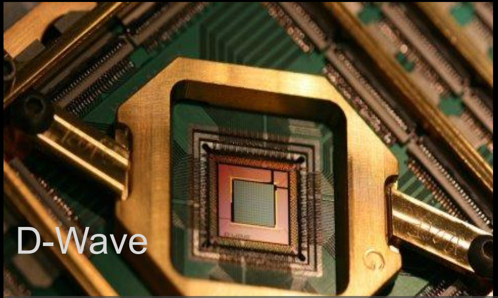
All discussed previously with BO regards mostly “offline” applications and expensive functions to evaluate...

Looking at the future: near real time optimizations

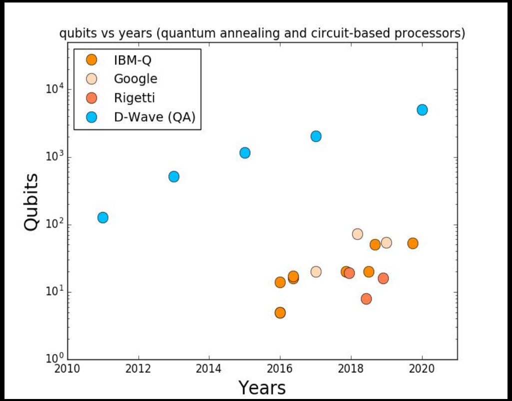
- The EIC will be built in ten years from now, and this allows to look at new emerging technologies.
- New innovative approaches like Streaming Readout will further the convergence of online and offline analysis leading to better data quality control during data taking and near real time calibrations.
- The EIC “Software” will likely include AI and Quantum Computing (and perhaps a combination of the two).
- QC in particular can revolutionize the way we deal with optimization problems.







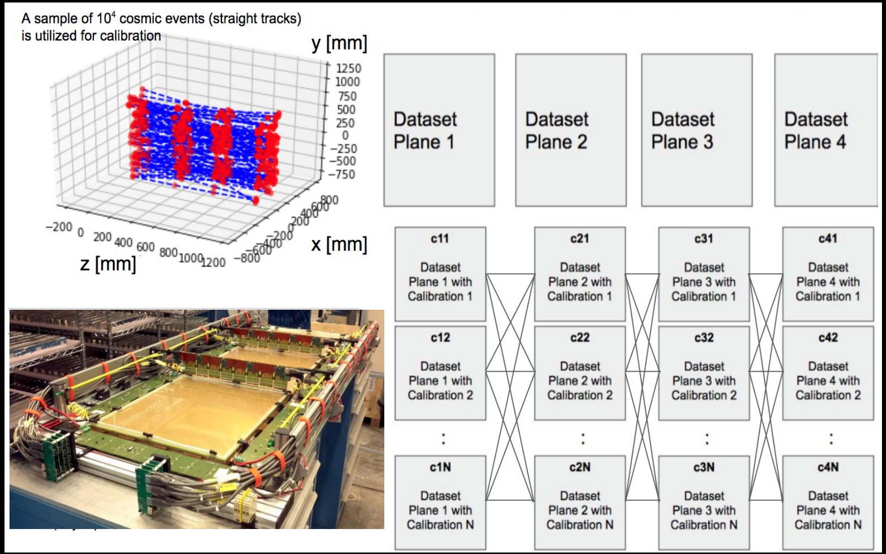
5[th] generation 5000-qubit system will be made available between 2020/2021



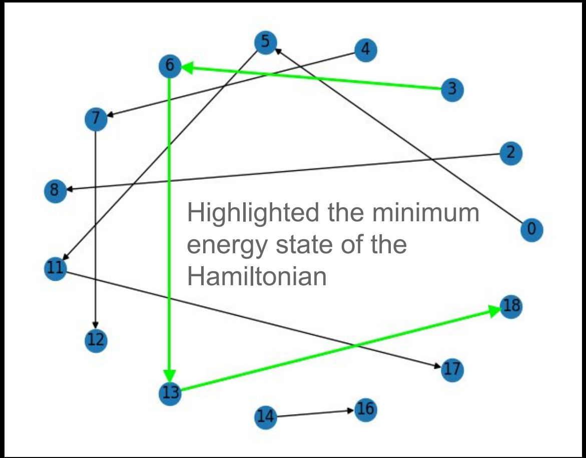
Alignment with QC+AI

TraQCking and Fast Alignment,
CF*, E. Cisbani, A. Del Dotto, W. Phelps, A. Quiroga

Detector: tracker system with 4 planes of GEM chambers; each plane has spatial and angular offsets



$$E(x_1,\ldots,x_N) = \sum_{i=1}^{N} h_i\, x_i + \sum_{i<j=1}^{N} J_{ij}\, x_i x_j$$



- Hopfield Networks with calibrations as units.
- With the current limits on available qubits we can process (study as a “whole”) ~ thousand of different sets of calibration constants simultaneously…
- Ideally more qubits = more configurations explored.

Summary

"Civilization advances by extending the number of important operations which we can perform without thinking of them" (Alfred North Whitehead)

- Supported by unprecedented computing resources and novel learning algorithms we can "optimize" detectors in a more efficient way ever done in past experiments.

- This can be extended to a system of subdetectors as in large scale experiment like EIC.

- <u>Detector design</u> assisted by AI is a very hot area of research and is just at its onset. It is possible that in the near future we will construct detectors entirely designed by AI.

- In the next decade we will likely use "Intelligent Detectors" able to self optimize <u>calibration/alignment</u> parameters on real time.

- AI/QIS will revolutionize the way experimental nuclear and particle physics is currently done.

Questions?