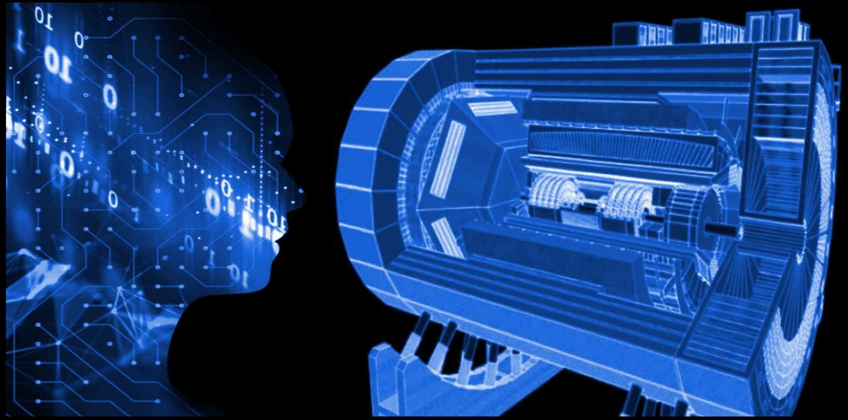


AI for Detector Design

National Nuclear Physics Summer School
MIT, 2022

Cristiano Fanelli

Lecture 1

Outline

- Lecture 1 (1.5h)
  - What this is not (and is) about
  - Complexity of modern detectors
  - How do we design and optimize detectors?
  - Examples
  - Toy example 1
- Lecture 2 (1.5h)
  - Multiple competing objectives
  - The ECCE example
  - Examples
  - Toy example 2
- Lecture 3 (1.5 h)
  - MOO in HEP/NP
  - Improving the workflow
    - Learning interactions of simulated particles with matter
    - Learning event reconstruction, pattern recognition
  - End-to-end optimization pipelines
  - Conclusions
  - Toy example 3

Useful References

[1] AI4NP Winter School, 2020 https://github.org/cfteach

[3] AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case https://arxiv.org/abs/1911.05797

[4] AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider https://arxiv.org/abs/2205.09185

[5] MODE: White Paper, https://arxiv.org/pdf/2203.13818.pdf

[6] Machine Learning in Nuclear Physics, https://arxiv.org/abs/2112.02309

What this is (and is not) about

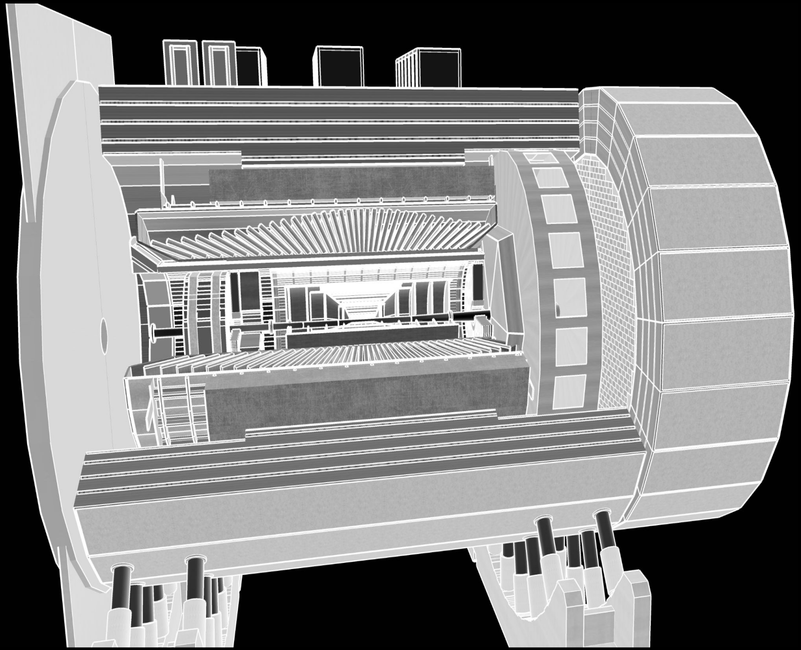
✅ What you will learn:

These lectures provide an overview of SOTA approaches for detector design with AI in NP/HEP. I will discuss what the challenges are and which techniques are used in this emerging area of research and why all this is beneficial for modern complex detector design.
Given the multidisciplinarity, it may be also of inspiration for other applications (actually embraces a wealth of use cases)

❌ What you will not learn:

- These are not lectures on Particle Detectors per se
  - For that, a great classic is Particle Detectors, C. Grupen and B. Schwartz
- These are not lectures on simulation toolkit like Geant to simulate detectors
  - https://geant4.web.cern.ch/
- These lectures in general assume some knowledge of MC event generators, detector simulation, event reconstruction and particle identification
  - I will only explain how they contribute but I won't go into details
- These lectures are definitely more focused on AI/ML but you won't learn AI/ML in 4.5h
  - I will try to provide some concrete examples and leave code snippets for optimization!

Complexity of Modern Detectors in Nuclear Physics



Detector Design with AI

- Do we need AI to design detectors?
    - Naively, no. We have done this in the past without AI.

Detector Design with AI

- Do we need AI to design detectors?
    - Naively, no. We have done this in the past without AI.

The design of measuring instruments can be quite complex, still the optimization can be tractable, i.e. a parametrized model can allow to define a likelihood $L = p(x|\theta)$, where θ are the modeling parameters and x the simulated data. [1]

For instrument whose functionality is based on quantum phenomena — interaction of radiation with matter — the optimization problem is intractable. Access to the generating function of observed data through forward simulation (setting referred to as likelihood-free or simulation-based inference [2])

Over the course of the past eighty eighty years, the intractability of the design optimization problems commonly encountered in particle physics has not prevented physicists from successfully conceiving, commissioning, and operating detectors of huge complexity. [1]
=> Long-standing "paradigms"

[1] A. Baydin,, et al. "Toward machine learning optimization of experimental design." *Nuclear Physics News* 31.1 (2021): 25-28.
[2] K. Cranmer, J. Brehme, G. Louppe, The frontier of simulation-based inference, PNAS Vol 117, No. 48

Detector Design with AI

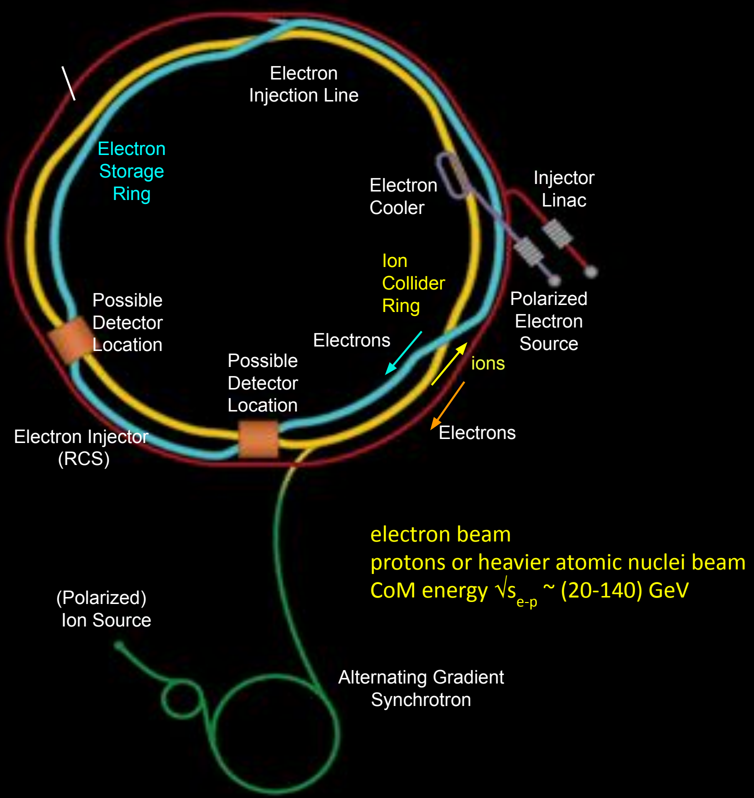
- Why these lectures then?
  - Accurate simulations are computationally expensive
  - Given the increasing complexity of modern experiments we seek to decrease the computational burden to optimally design detectors
  - Improving the detector design involves often optimizing simultaneous “tasks” in a multidimensional design space
  - Unprecedented opportunity to rethink the design strategy in terms of geometry, material, performance, costs…
    - still leveraging on existing paradigms (e.g., validation);
  - (bonus: are complex detectors designed in the past sub-optimal?)
- In this context, AI-assisted approaches able to outperform manual, brute-force, approaches.
  - Designing detectors with AI is a multidisciplinary effort that combines multiple domains of expertise

Detector Design with AI

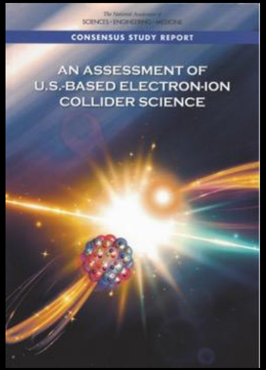
- Fundamental nuclear and particle physics research often requires realizing expensive large-scale experiments combining multiple sub-detectors to investigate the building blocks of nature.
    - "*AI techniques that can optimize the design of complex, large-scale experiments have the potential to revolutionize the way experimental nuclear and particle physics is currently done*". [1]
- More than 50 years have passed since Charpak (Nobel Prize in 1992) revolutionised particle detectors with the construction of a MWPC. Nowadays we can 3D print scintillation detectors and complex detection elements with thin layers of AC-coupled resistive silicon sensors. [2,3]
- Thanks to the fast progress in CS in the past two decades, along with optimization software and the development of DNN, we now have the unique opportunity to integrate these new tools during the design of complex detection systems.
- Using AI will allow to optimize large detectors in NP experiments like the Electron Ion Collider. EIC 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.
- In the following I will often utilize the Electron Ion Collider detector as a reference for our discussion.

[1] R. Stevens et al., AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science
[2] Y. Mishnayot et al., 3-dimensional printing of scintillating materials, Rev. Sci. Instrum., 85:085102, 2014
[3] G. Giacomini et al., Fabrication and performance of AC-coupled LGADs.

EIC Science



National Academy of Sciences



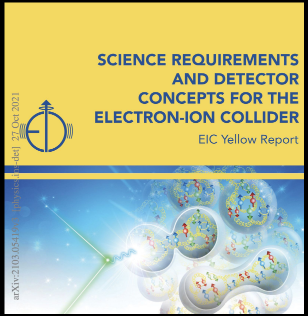
- Finding 1: An EIC can uniquely address three profound questions about nucleons — neutrons and protons — and how they are assembled to form the nuclei of atoms:
  - How does the mass of the nucleon arise?
  - How does the spin of the nucleon arise?
  - What are the emergent properties of dense systems of gluons?
- Finding 2: These three high-priority science questions can be answered by an EIC with highly polarized beams of electrons and ions, with sufficiently high luminosity and variable center of mass energy.
- Finding 3: An EIC would be a unique facility in the world and would maintain U.S. leadership in nuclear physics
- Finding 4: An EIC would maintain U.S. leadership in the accelerator science and technology colliders and help to maintain scientific leadership more broadly.

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

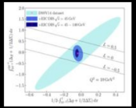
EIC Science

EIC Yellow Report (2021)

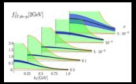
arXiv:2103.05419



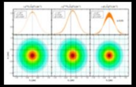
- Origin of Nucleon Spin



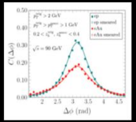
- Confined motion of partons



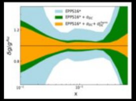
- 3D imaging quarks and gluons



- Nucleon mass

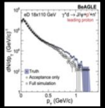


- High gluon densities in nuclei

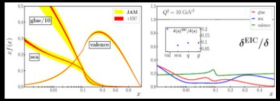


- Quarks and gluons in the nucleus

- Light-ion tagging



- Pion/Kaon structure

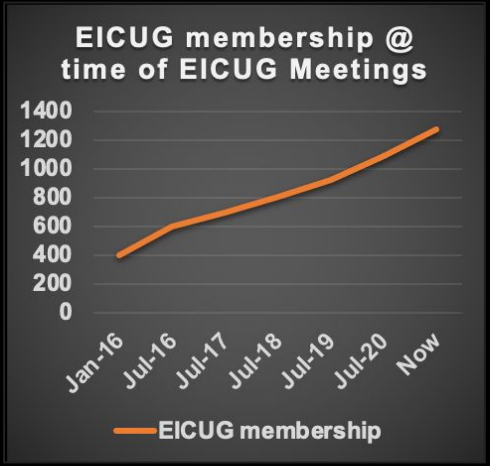


- Diffractive jets?
- Nuclear modifications and in-medium evolution
  - D/D* reconstruction and heavy-flavor in jets

Khalek, R. Abdul, et al. "Science requirements and detector concepts for the electron-ion collider: EIC yellow report." arXiv:2103.05419, 2021

Slide taken from J. Lajoie, The ECCE Experiment, Workshop VIII Streaming Readout, 2021

World-wide interest



Typical EIC experimental measurements

Inclusive Reactions in ep/eA

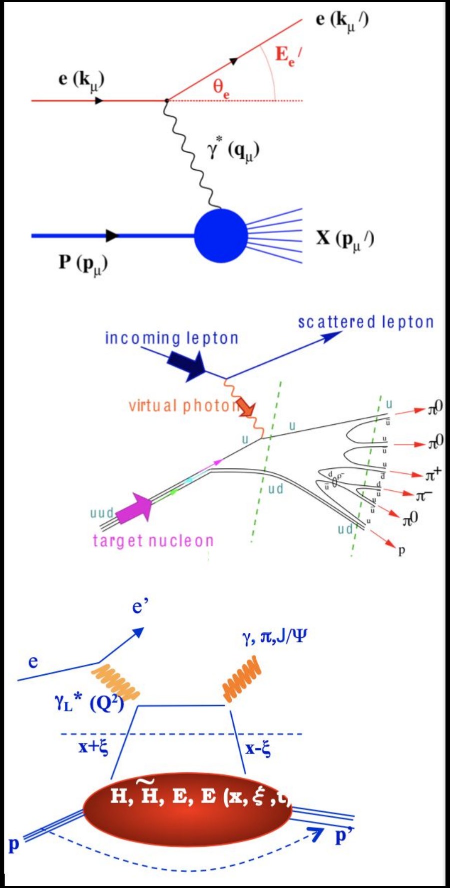
- Structure Functions: g_1, F_2, F_L

Semi-Inclusive Reactions in ep/eA

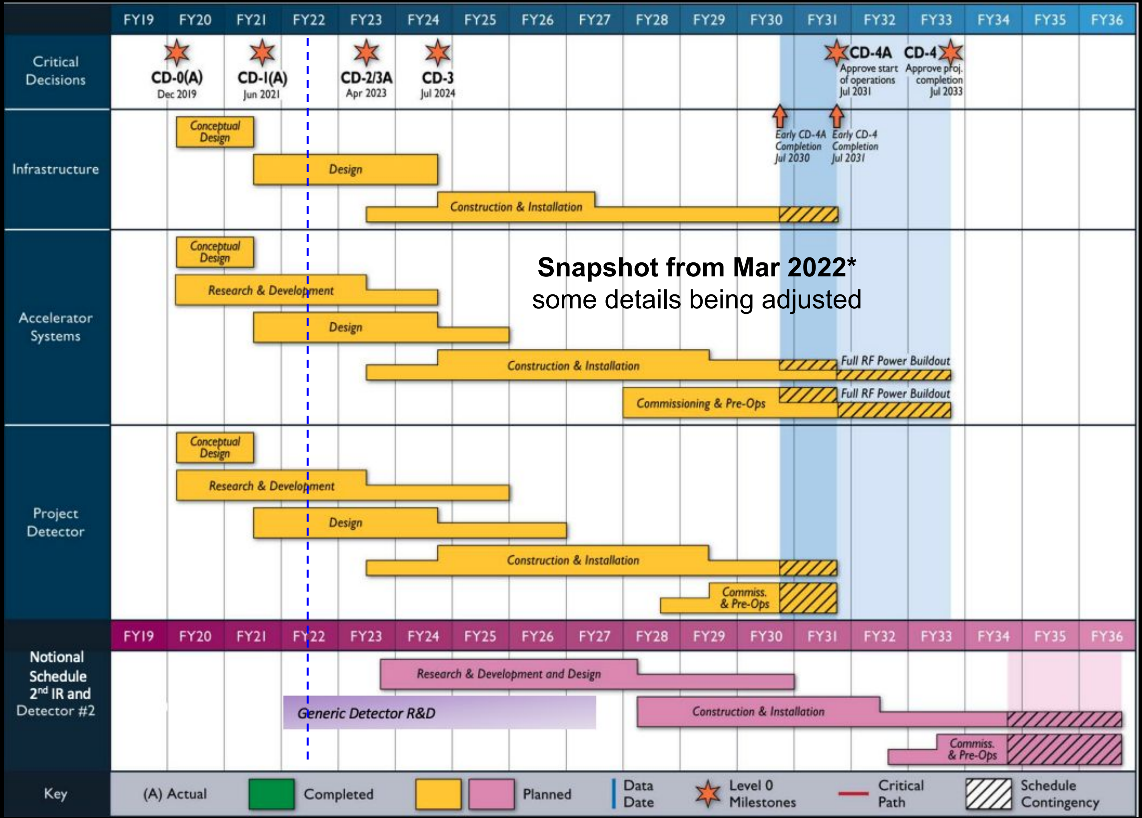
- TMDs, Helicity PDFs, FFs; di-hadron correlations; kaon asymmetries, cross-sections, …

Exclusive Reactions in ep/eA

- DVCS, exclusive VM production (GPDs; parton imaging)

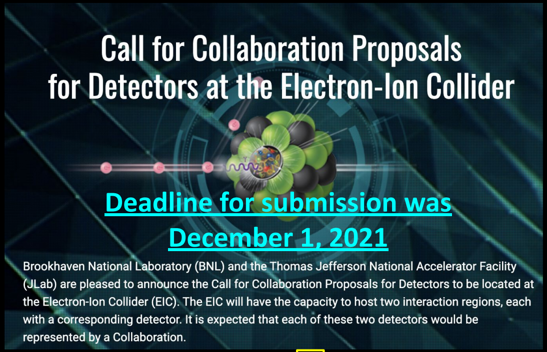


EIC Schedule and Milestones



*EIC Schedule, J. Yeck, Mar 2022

proto-collaborations



Process completed on March 21, 2022

Panel Report

6. Recommendations: ECCE Reference Detector

The panel unanimously recommends ECCE as Detector 1. The proto-collaboration is urged to openly accept additional collaborators and quickly consolidate its design so that the Project Detector can advance to CD2/3a in a timely way.

EIC DETECTOR 1 GENERAL MEETING

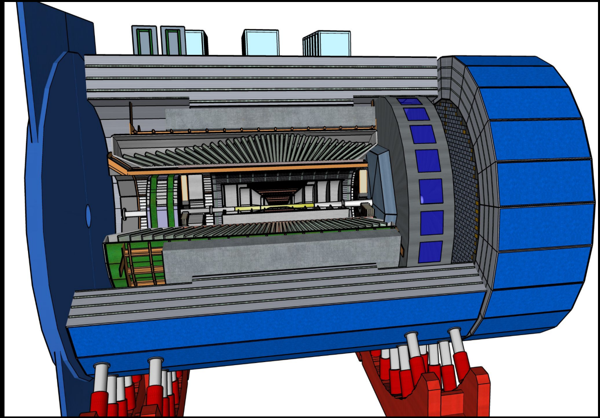
Following the DPAP process , the EIC Community is moving towards the formation of a scientific collaboration to support the realization of the EIC project detector - temporarily referred to as "Detector-1".

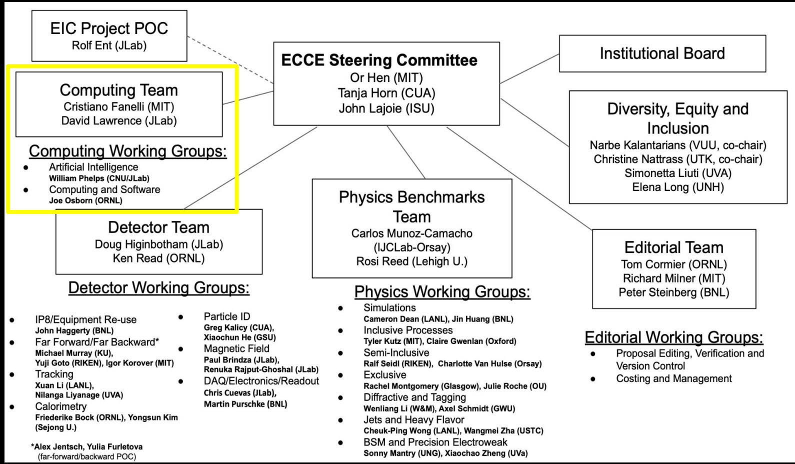
Towards Collaboration

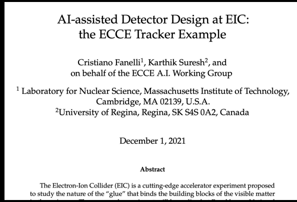
EIC Comprehensive Chromodynamics Experiment

- Proto-collaboration that comprised scientists from 98 institutions
- Develop low-risk, cost-effective, flexible and optimized EIC detector
- Detector concept based on a 1.5 T solenoidal magnet

https://www.ecce-eic.org






AI-assisted Detector Design at EIC: the ECCE Tracker Example

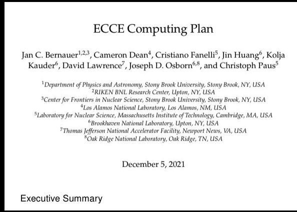
Cristiano Fanelli[1], Karthik Suresh[2], and on behalf of the ECCE A.I. Working Group

[1] Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
[2] University of Regina, Regina, SK S4S 0A2, Canada

December 1, 2021

Abstract

The Electron-Ion Collider (EIC) is a cutting-edge accelerator experiment proposed to study the nature of the "glue" that binds the building blocks of the visible matter

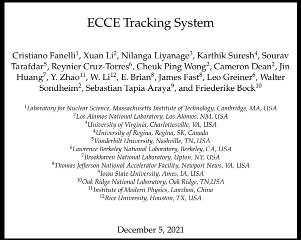

ECCE Computing Plan

Jan C. Bernauer[1,2,3], Cameron Dean[4], Cristiano Fanelli[5], Jin Huang[6], Kolja Kauder[6], David Lawrence[7], Joseph D. Osborn[6,8], and Christoph Paus[5]

[1] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA
[2] RIKEN BNL Research Center, Upton, NY, USA
[3] Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, NY, USA
[4] Los Alamos National Laboratory, Los Alamos, NM, USA
[5] Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[6] Brookhaven National Laboratory, Upton, NY, USA
[7] Thomas Jefferson National Accelerator Facility, Newport News, VA, USA
[8] Oak Ridge National Laboratory, Oak Ridge, TN, USA

December 5, 2021

Executive Summary

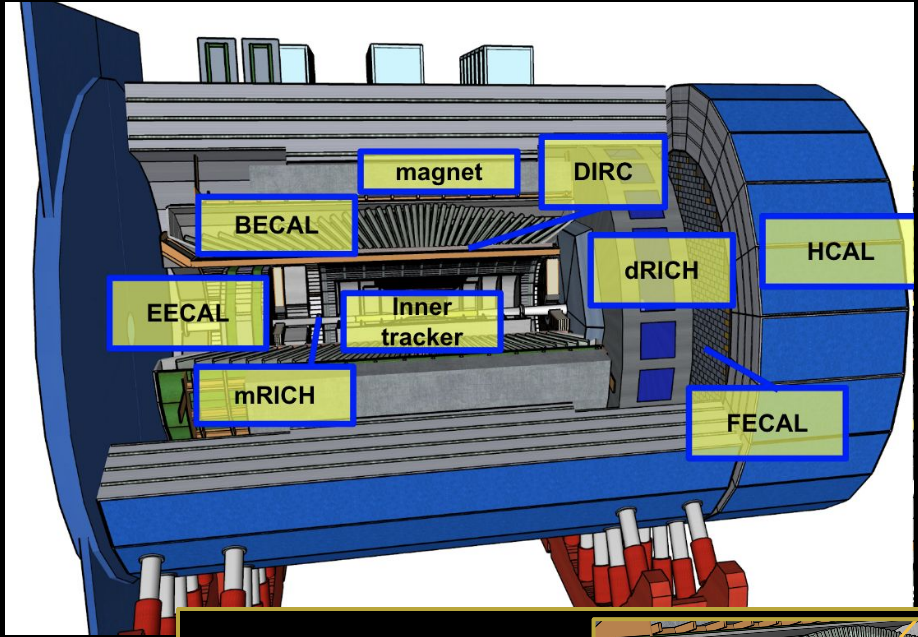

ECCE Tracking System

Cristiano Fanelli[1], Xuan Li[2], Nilanga Liyanage[3], Karthik Suresh[4], Sourav Tarafdar[5], Reynier Cruz-Torres[6], Cheuk Ping Wong[2], Cameron Dean[2], Jin Huang[7], Y. Zhao[11], W. Li[12], E. Brian[8], James Fast[8], Leo Greiner[6], Walter Sondheim[2], Sebastian Tapia Araya[9], and Friederike Bock[10]

[1] Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[2] Los Alamos National Laboratory, Los Alamos, NM, USA
[3] University of Virginia, Charlottesville, VA, USA
[4] University of Regina, Regina, SK, Canada
[5] Vanderbilt University, Nashville, TN, USA
[6] Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[7] Brookhaven National Laboratory, Upton, NY, USA
[8] Thomas Jefferson National Accelerator Facility, Newport News, VA, USA
[9] Iowa State University, Ames, IA, USA
[10] Oak Ridge National Laboratory, Oak Ridge, TN, USA
[11] Institute of Modern Physics, Lanzhou, China
[12] Rice University, Houston, TX, USA

December 5, 2021

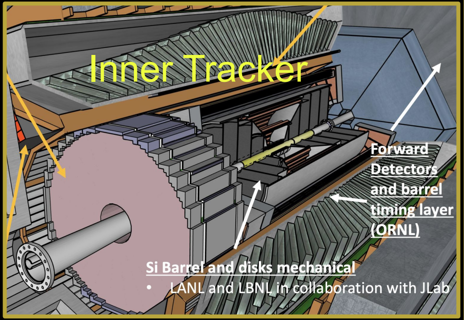
Reference Detector



Tracker

Combines:

- ITS-3 Si technology
- Gaseous detectors
- AC-LGAD ToFs



PID with Cherenkov detectors

- dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction
- DIRC (detection of internally reflected Cherenkov light) in the barrel
- modular RICH in the electron direction.

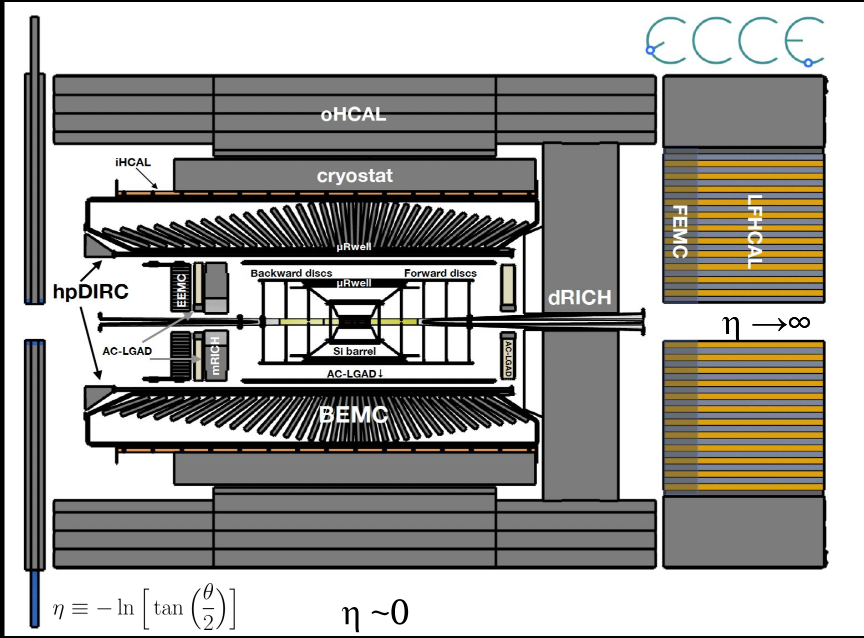
Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.

All three rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!

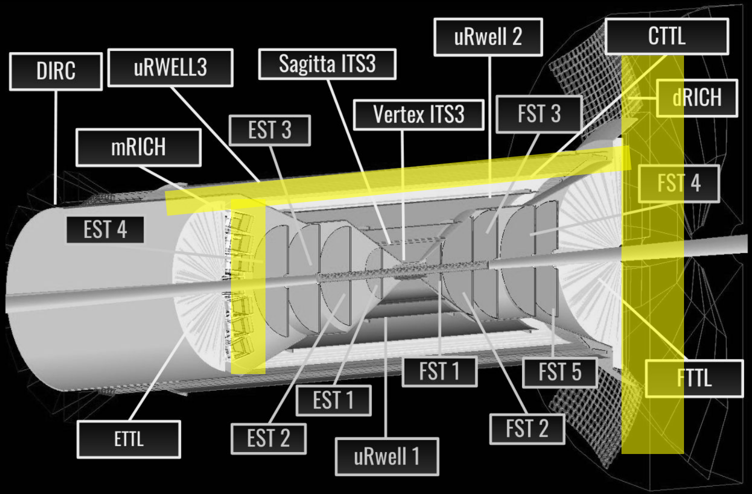
Reference Detector

In these lectures I will often refer to these subsystems

We have a reference (ECCE) detector.
Possible updates are currently being investigated (detector-1).



Tracker System + PID



- The tracking system reconstructs charged particle tracks. It combines different technologies.
- Imaging Cherenkov detectors are the backbone of PID in EIC. Compute intensive to simulate / reconstruct.

Ideal vs Real Detectors

Ideal Case:

- Given a process,
  - detect all final state particles with 100% efficiency,
  - determining the particles types with certainty and
  - reconstructing their “true” 4-momenta

Real Case:

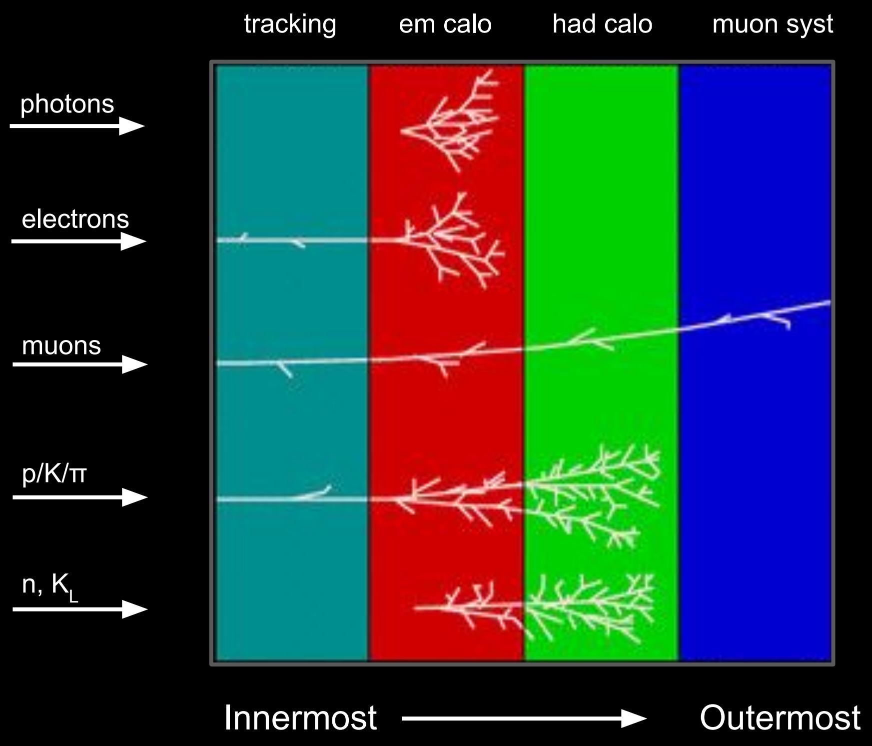
- For most particle species we deal with
- decay products, secondary vertices, invariant mass peaks;
- one really never gets 100% acceptance (due to support system, beam pipe, sub-detector frames, cracks, …)
- as well as a 100% detection efficiency (due to detector imperfections, the limitations of our reconstruction algorithms, DAQ, etc.)
- PID is never 100% accurate and we deal with finite detector resolutions (detector size and technology limitations, costs…)
- Background processes make the overall picture more complicated

How do we detect particles?

- Long-lived: through their interaction with the detector material
  - Tracking
  - Calorimetry
  - PID detectors
- Short-lived: through measuring their decay products

neutrinos	none	Missing energy
electrons	Ionisation, electromagnetic	Track and EM shower
muons	Ionisation	Penetrating track
p, K, π	Ionisation, hadronic	Track and hadron shower
photons	electromagnetic	EM shower
neutrons, K^0_L	hadronic	hadron shower
B, D	Weak decay	Secondary vertex
J/ψ, Υ, W, Z, H, t	prompt decay	Invariant mass

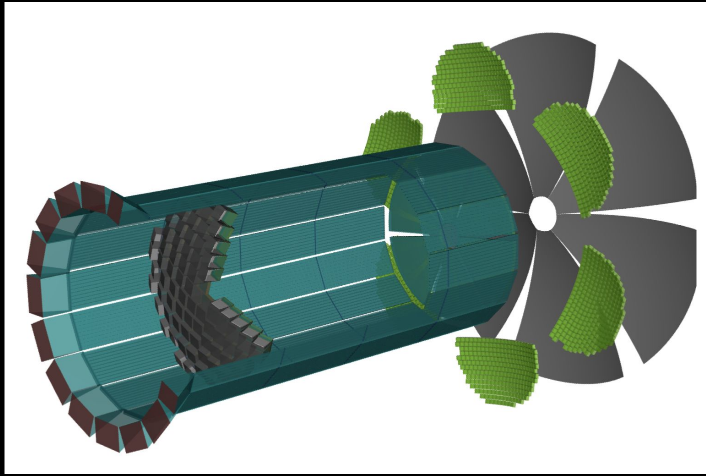
Particle Reconstruction



- The tracking system reconstructs charged particle tracks. It combines different technologies.
- Calorimetry measurement is “destructive” →components belonging to the tracking system are the closest to the IP

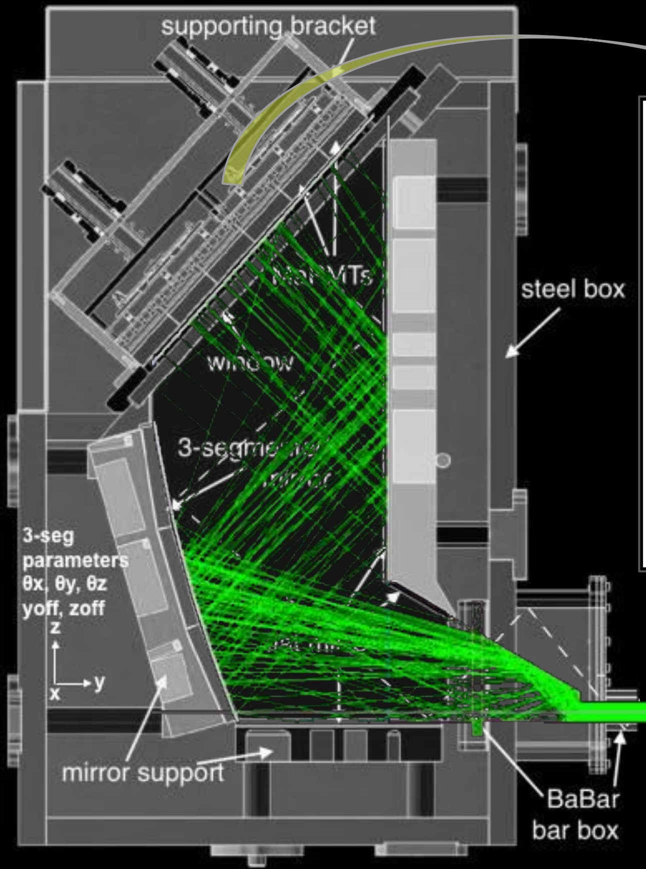
Particle Identification with Cherenkov

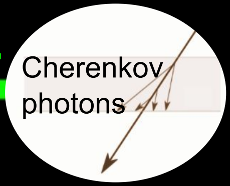
		electrons/photons		π/K/p	
eta	Nomenclature	PID	Min E Photon	P-range [GeV/c]	Separation
-3.5 to -2.0	Backward	π suppression up to 1:1E-4	20 MeV	≤ 10 GeV/c	≤ 3σ
-2.0 to -1.0	Backward	π suppression up to 1:1E-3 - 1:1E-2	50 MeV		
-1.0 to 1.0	Barrel	π suppression up to 1:1E-2	100 MeV	≤ 6 GeV/c	
1.0 to 3.5	Forward	3σ e/π up to 15 GeV/c	50 MeV	≤ 50 GeV/c	



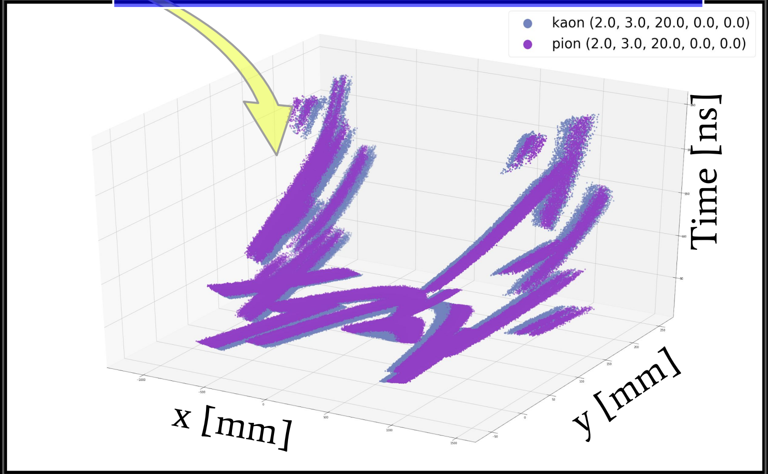
- Cherenkov detectors form the backbone of PID at EIC
  - Currently, all EIC detector designs use a dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction, a DIRC (detection of internally reflected Cherenkov light) in the barrel, and a modular RICH in the electron direction.
  - Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.
  - All three rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!

Particle Identification: DIRC





Hit pattern defined in (x,y,t)



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.

1PMT made by 64 pixels, each pixel is 6mm x 6mm size

Displayed PDF. Patterns are sparse with variable photon yield

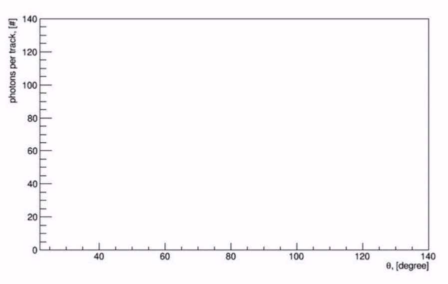
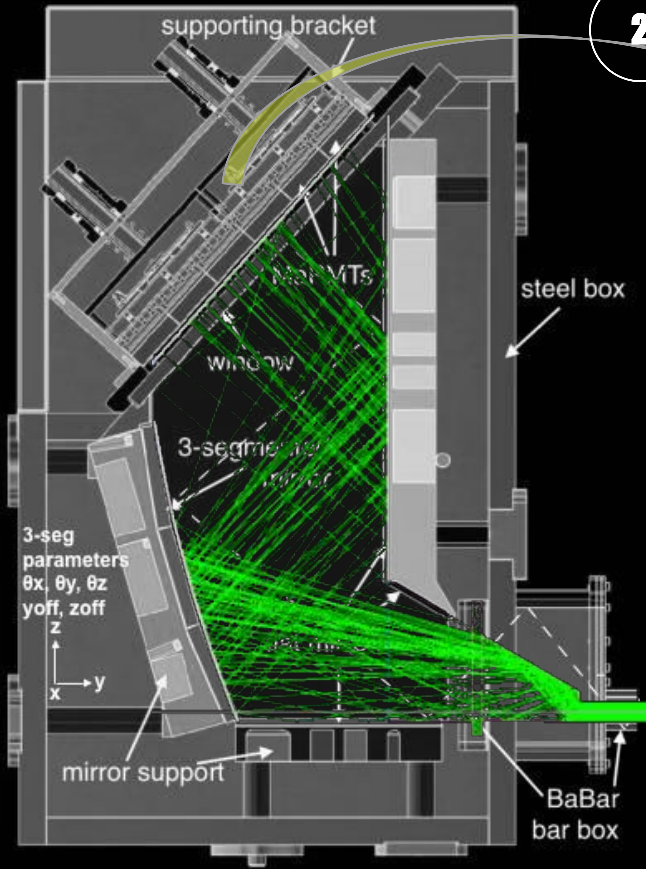
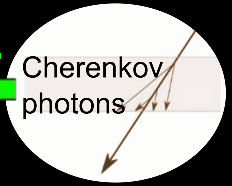


Image of expansion volume taken from GlueX DIRC, Ali et al., JINST 15 (2020) 04, C04054

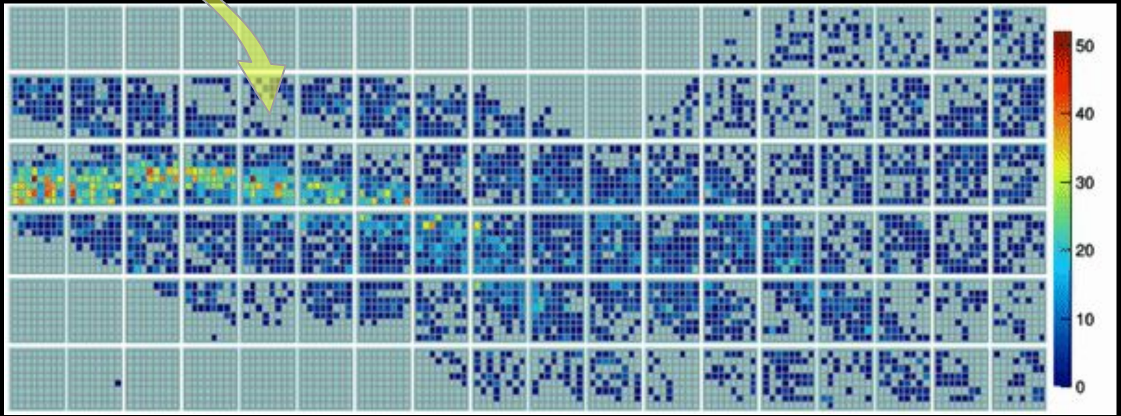
Particle Identification: DIRC





2

Kaons @ 4 GeV/c for different polar and azimuthal angle



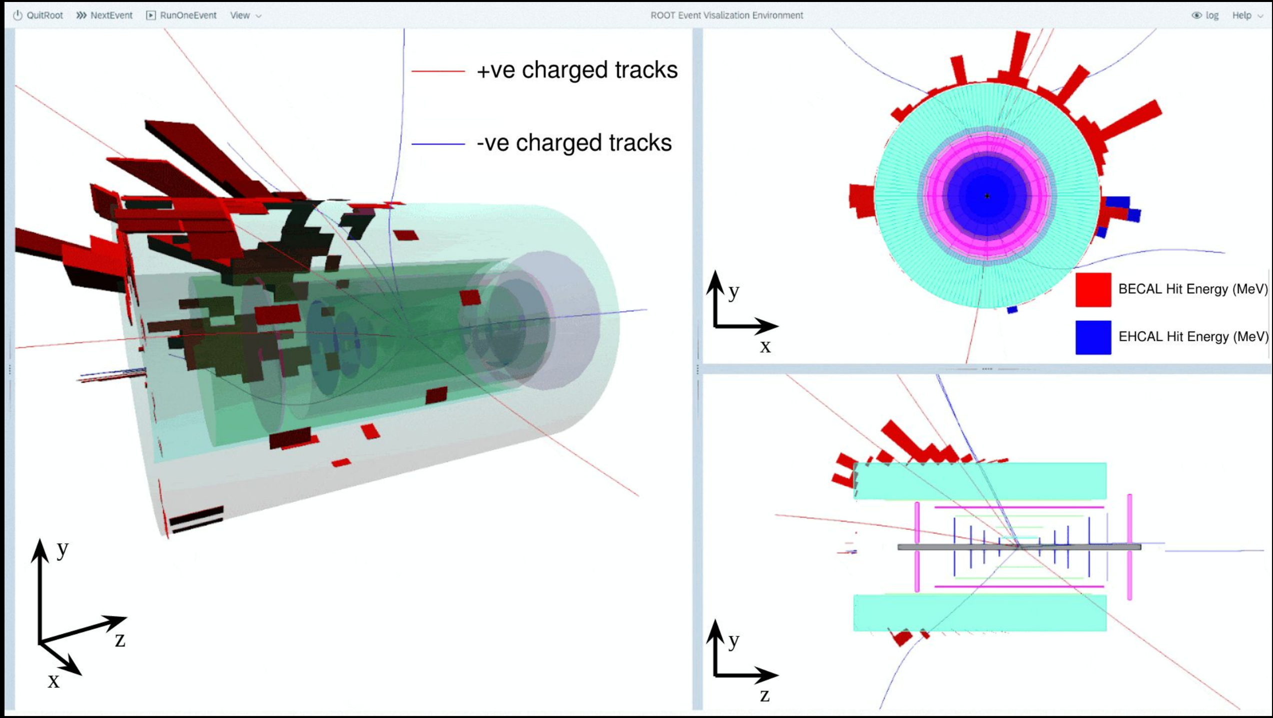
A. Ali et al., The GlueX DIRC Program, 2020 JINST 15 C04054.

Dependence on charged particle kinematics
(p,(θ,φ)*,X,Y)

1

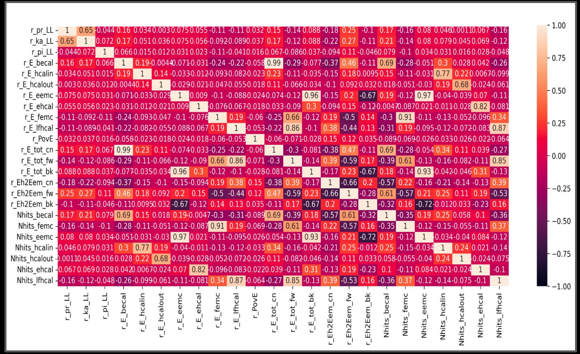
Image of expansion volume taken from GlueX DIRC, Ali et al., JINST 15 (2020) 04, C04054

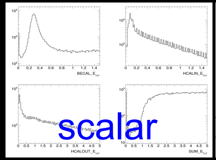
Event Display and Reconstructed Features

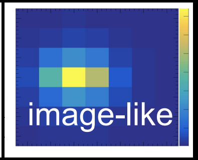


Reconstruction typically deals with relatively large feature space (low and high-level features) combining sub-detectors

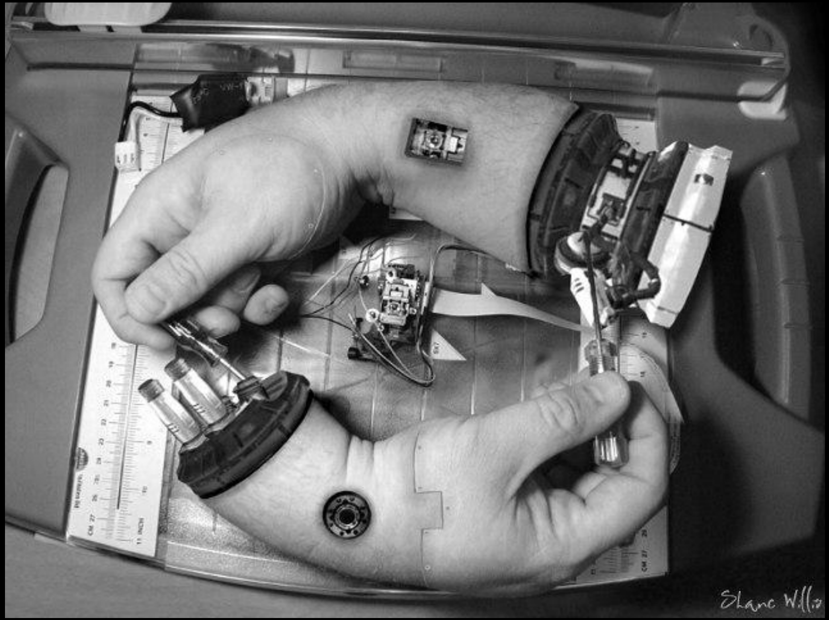
For illustrative purposes, showing example of calorimetry (outer layers)







How do we design and optimize Detectors?



AI for Design

It is a relatively new but active area of research.
Many applications in, e.g., industrial material, molecular and drug design.

Guo, Kai, et al. *Materials Horizons* 8.4 (2021): 1153-1172.

Table 1 Popular ML methods in design of mechanical materials

ML method	Characteristics	Example applications in mechanical materials design
Linear regression; polynomial regression	Model the linear or polynomial relationship between input and output variables	Modulus[112] or strength[123] prediction
Support vector machine; SVR	Separate high-dimensional data space with one or a set of hyperplanes	Strength[123] or hardness[125] prediction; structural topology optimization[159]
Random forest	Construct multiple decision trees for classification or prediction	Modulus[112] or toughness[130] prediction
Feedforward neural network (FFNN); MLP	Connect nodes (neurons) with information flowing in one direction	Prediction of modulus,[97,112] strength,[93] toughness[130] or hardness;[97] prediction of hyperelastic or plastic behaviors;[143,145] identification of collision load conditions;[147] design of spinodoid metamaterials[163]
CNNs	Capture features at different hierarchical levels by calculating convolutions; operate on pixel-based or voxel-based data	Prediction of strain fields[104,105] or elastic properties[102,103] of high-contrast composites, modulus of unidirectional composites,[136] stress fields in cantilevered structures,[137] or yield strength of additive-manufactured metals;[121] prediction of fatigue crack propagation in polycrystalline alloys;[140] prediction of crystal plasticity;[120] design of tessellate composites;[107–109] design of stretchable graphene kirigami;[155] structural topology optimization[156–158]
Recurrent neural network (RNN); LSTM; GRU	Connect nodes (neurons) forming a directed graph with history information stored in hidden states; operate on sequential data	Prediction of fracture patterns in crystalline solids;[114] prediction of plastic behaviors in heterogeneous materials;[142,144] multi-scale modeling of porous media[173]
Generative adversarial networks (GANs)	Train two opponent neural networks to generate and discriminate separately until the two networks reach equilibrium; generate new data according to the distribution of training set	Prediction of modulus distribution by solving inverse elasticity problems;[138] prediction of strain or stress fields in composites;[139] composite design;[164] structural topology optimization;[165–167] architected materials design[115]
Gaussian process regression (GPR); Bayesian learning	Treat parameters as random variables and calculate the probability distribution of these variables; quantify the uncertainty of model predictions	Modulus[122] or strength[123,124] prediction; design of supercompressible and recoverable metamaterials[110]
Active learning	Interacts with a user on the fly for labeling new data; augment training data with post-hoc experiments or simulations	Strength prediction[124]
Genetic or evolutionary algorithms	Mimic evolutionary rules for optimizing objective function	Hardness prediction;[126] designs of active materials;[160,161] design of modular metamaterials[162]
Reinforcement learning	Maximize cumulative awards with agents reacting to the environments.	Deriving microstructure-based traction-separation laws[174]
Graph neural networks (GNNs)	Operate on non-Euclidean data structures; applicable tasks include link prediction, node classification and graph classification	Hardness prediction;[127] architected materials design[168]

Z. Zhou et al., *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019

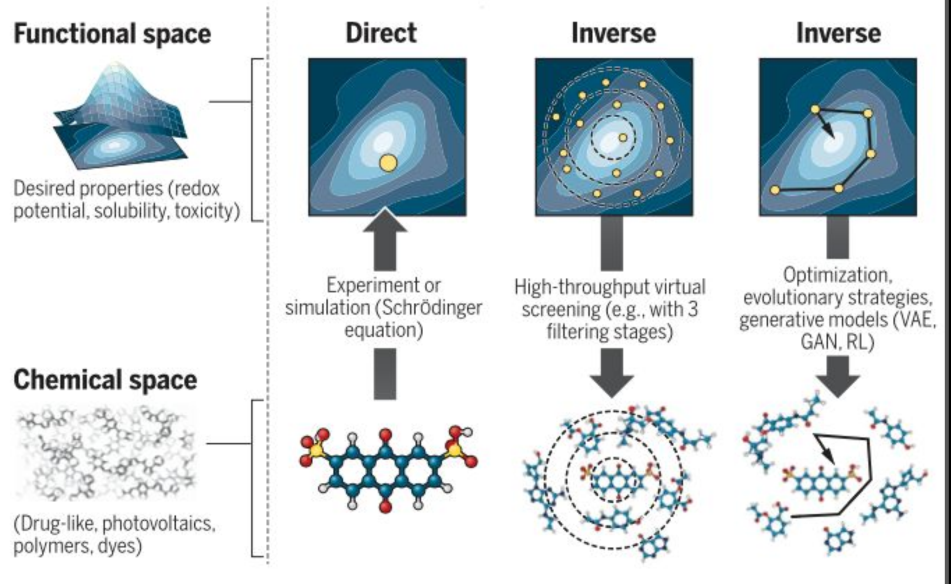


Fig. 2. Schematic of the different approaches toward molecular design. Inverse design starts from desired properties and ends in chemical space, unlike the direct approach that leads from chemical space to the properties.

B. Sanchez-Lengeling, A. Aspuru-Guzik. *Science* 361.6400 (2018): 360-365.

Full Optimization of Detectors/Accelerators

- When it comes to designing detectors and accelerators with AI this is a frontier topic with few examples in the literature.

S. Shirobokov, V. Belavin, M. Kagan, A. Ustyuzhanin, and A.G. Baydin. Black-Box Optimization with Local Generative Surrogates, 2020. arXiv: 2002.04632.

T. Dorigo. Geometry optimization of a muon-electron scattering detector. Physics Open, 4:100022, 2020.

F. Ratnikov. Using machine learning to speed up and improve calorimeter R&D. Journal of Instrumentation, 15(05):C05032, 2020.

E. Cisbani et al. AI-optimized detector design for the future Electron Ion Collider: the dual-radiator RICH case. JINST 15(05):P05009, 2020.

A. Edelen et al. Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems. Physical Review Accelerators and Beams, 23(4):044601, April 2020. Publisher: American Physical Society. doi:10.1103/PhysRevAccelBeams.23.044601.

D. Koser et al. Input beam matching and beam dynamics design optimization of the IsoDAR RFQ using statistical and machine learning techniques. arXiv:2112.02579 [physics], 2021. (Submitted to Frontiers in Physics). arXiv:2112.02579. [61] F. Van Der Veken et al. Machine learning in accelerator physics: applications at the CERN Large Hadron Collider. In Proceedings of Artificial Intelligence for Science, Industry and Society PoS(AISIS2019), volume 372, page 044. SISSA Medialab, July 2020.

S. Meyer et al. Optimization and performance study of a proton CT system for pre-clinical small animal imaging. Phys. Med. Biol., 65(15):155008, 2020. doi:10.1088/1361-6560/ab8afc.

C. Fanelli et al., AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider arXiv:2205.09185, 2022

Full Optimization of Detectors/Accelerators

- When it comes to designing detectors and accelerators with AI this is a frontier topic with few examples in the literature.
  - What follows uses “detector” as example but applies to both detector and accelerator.
- Typically full detector design is studied once the subsystem prototypes are ready (phase constraints from the full detector or outer layers are taken into consideration).
- Need to use advanced simulations which are computationally expensive (Geant).
- Many parameters (and multiple objective functions): curse of dimensionality [1].
- Entails establishing a procedural body of instructions [2]. The choice of a suitable algorithm is a challenge itself (no free lunch theorem [3]) and always requires some degree of customization.

[1] Bellman, Richard. *Dynamic programming*. Vol. 295. RAND CORP SANTA MONICA CA, 1956.
[2] CF et al. *JINST* 15.05 (2020): P05009.
[3] Wolpert, D.H., Macready, W.G., 1997. Trans. Evol. Comp 1, 67–82

Characterizing the Detector Design Problem

The detector design problem in NP physics (either collider or fixed target) experiments is typically characterized by:

1. *A number of sub-detectors* layers *starting from the interaction point;*

2. *A relational '*hierarchy*' (or* coupling*) among different components/sub-detectors (e.g., the presence of material in front of a sub-detector; calorimetry after tracking; etc);*

3. *Symmetry (e.g., hermetic detectors with large acceptance like EIC have a 'cylindrical' geometry);*

4. *Modularity (e.g., repeated sub-elements within a sub-detector);*

5. *Constraints (e.g., volumes cannot overlap);*

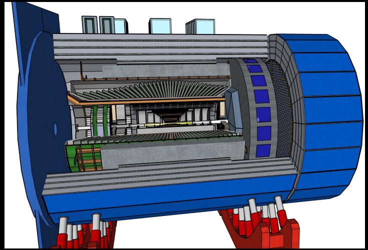
6. *"Heterogeneous" parts (e.g., certain processes like developing showers in calorimeters take longer than others to simulate; point 6 actually encompasses different aspects in the pipeline, see later).*

7. *The detector response is typically* noisy *and detailed simulations can be* compute expensive.

The Typical Workflow

See invited talk at IAEA Technical Meeting on AI

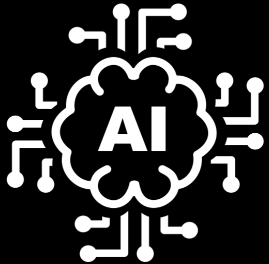
IAEA
International Atomic Energy Agency



- AI can assist in designing more efficiently detectors (performance, costs).
- It helps steering the design (and eventually fine-tune it).
- It can capture hidden correlations among design parameters.

A.I.
gathers observations and suggests new points

customization



Design parameters

Physics Events

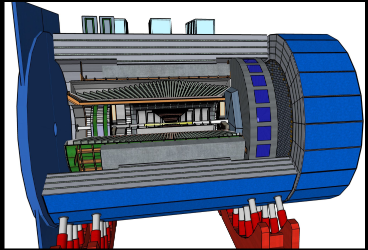
Detector Simulation

compute intensive (Geant4)

Analysis of High-level reconstruction of events

(AI/ML can also speed-up the simulation/reconstruction stack; cf. Amdahl's law)

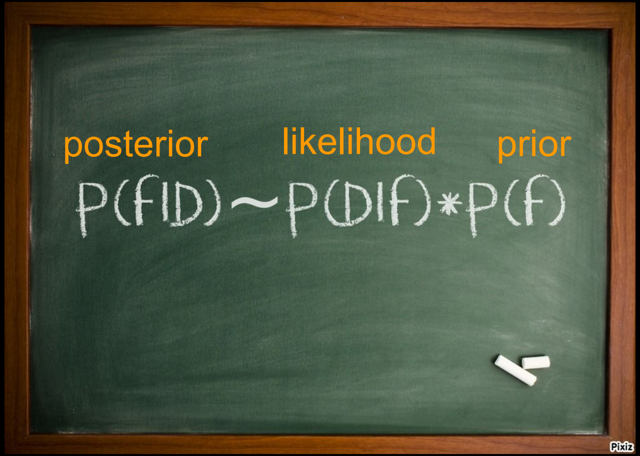
Design Optimization



- AI can assist in designing more efficiently detectors (performance, costs).
- It helps steering the design (and eventually fine-tune it).
- It can capture hidden correlations among design parameters.

- AI-assisted "Optimization" is not necessarily fine-tuning, this is a common misconception
- These techniques can be utilized in different phases of design and R&D
  - E.g., detector modeling can be done with optimization on a reduced set of parameters keeping frozen all the others, before a global optimization
  - Even the global optimization may be approximate but still steer the design towards the most interesting regions
  - Different technology choice/change can be made during the design phase and as part of the decision making, informed by AI — see later discussion on EIC tracking

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.

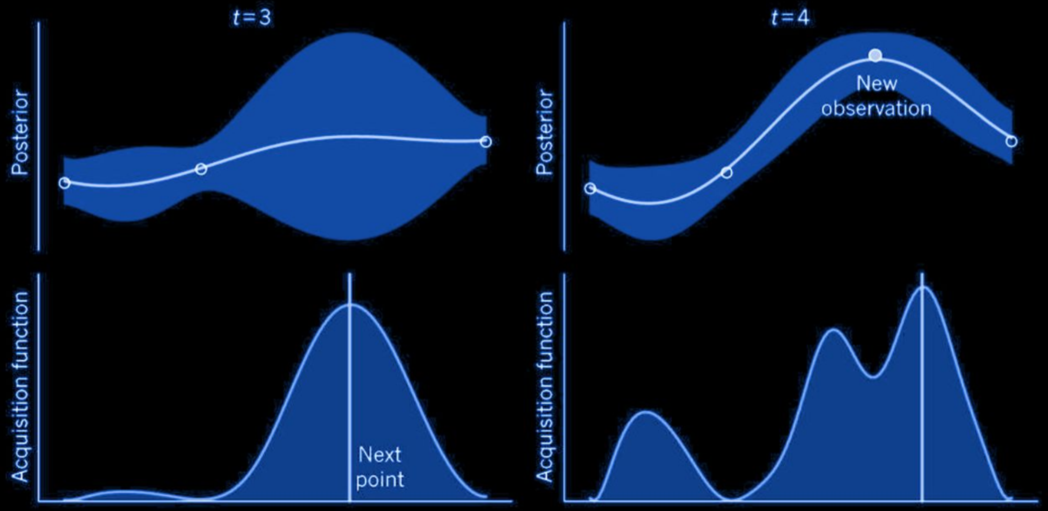
If you don't have the above constraints,
do not use Bayesian Optimization

- 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/

BO in a nutshell

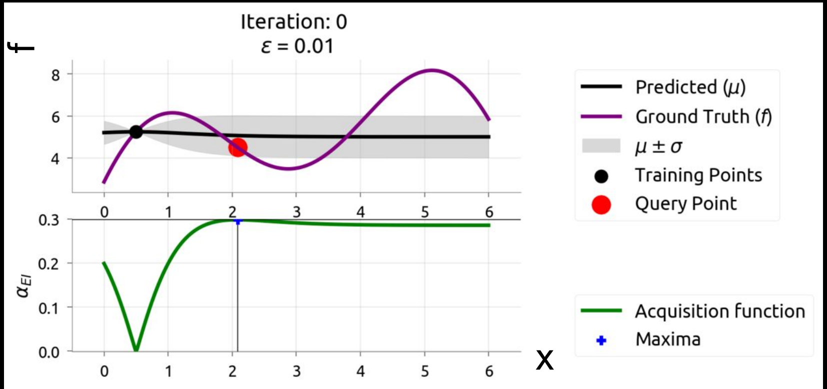
- BO is a sequential strategy developed for global optimization.
- After gathering evaluations we builds a posterior distribution used to construct an **acquisition function**.
- This cheap function determines what is **next query point**.



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.

Extension to multiple objectives

Acquisition Functions



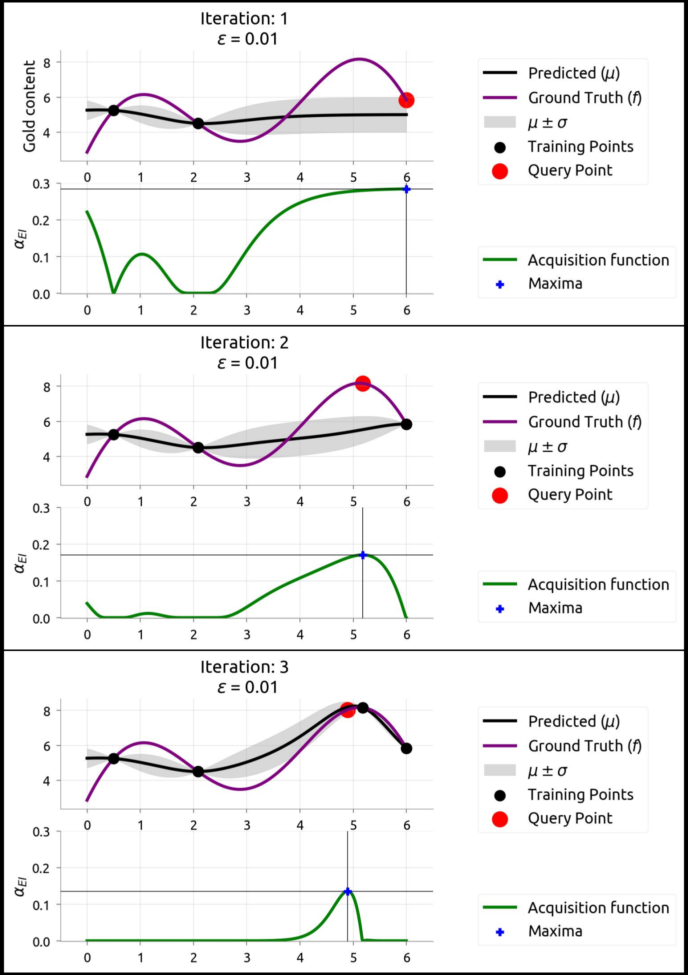
Exploitation **Exploration**

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

Best found so far

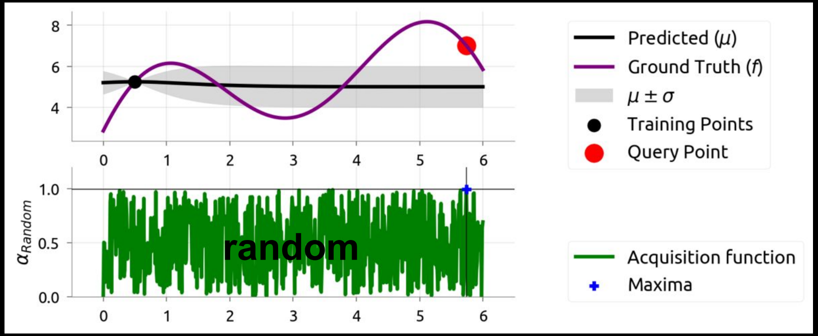
$$Z = \frac{\mu_t(x) - f(x^+) - \epsilon}{\sigma_t(x)}$$

We are sampling x



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

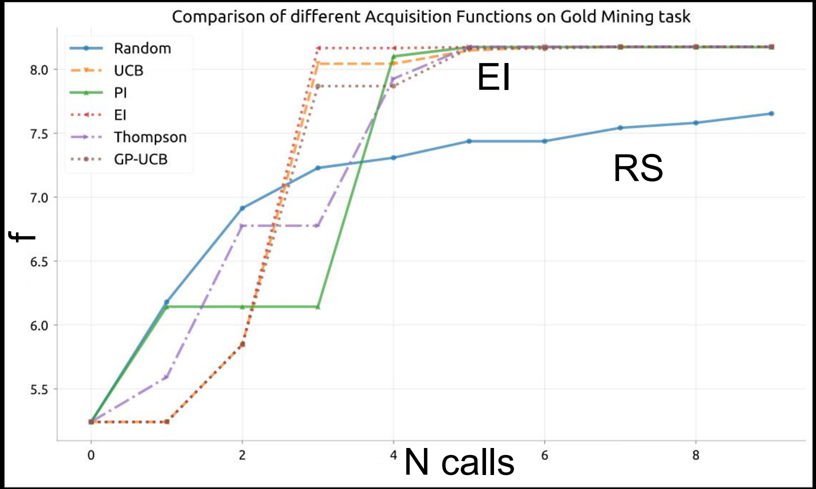
Acquisition Functions



- 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



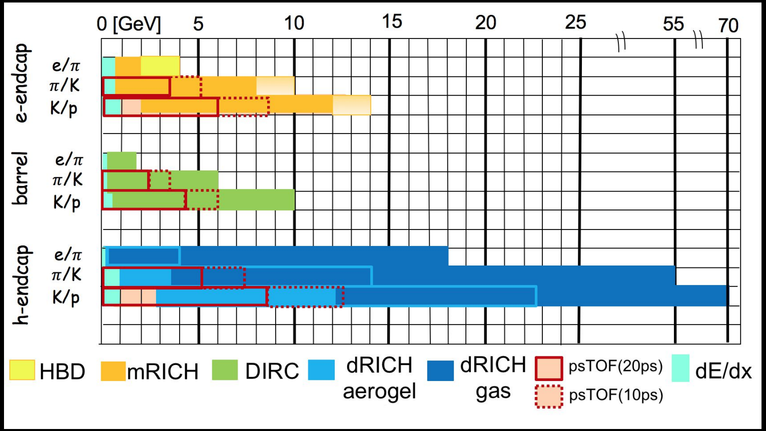
E. Brochu, Eric, V. M. Cora, and N. De Freitas. "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning." *arXiv:1012.2599* (2010).

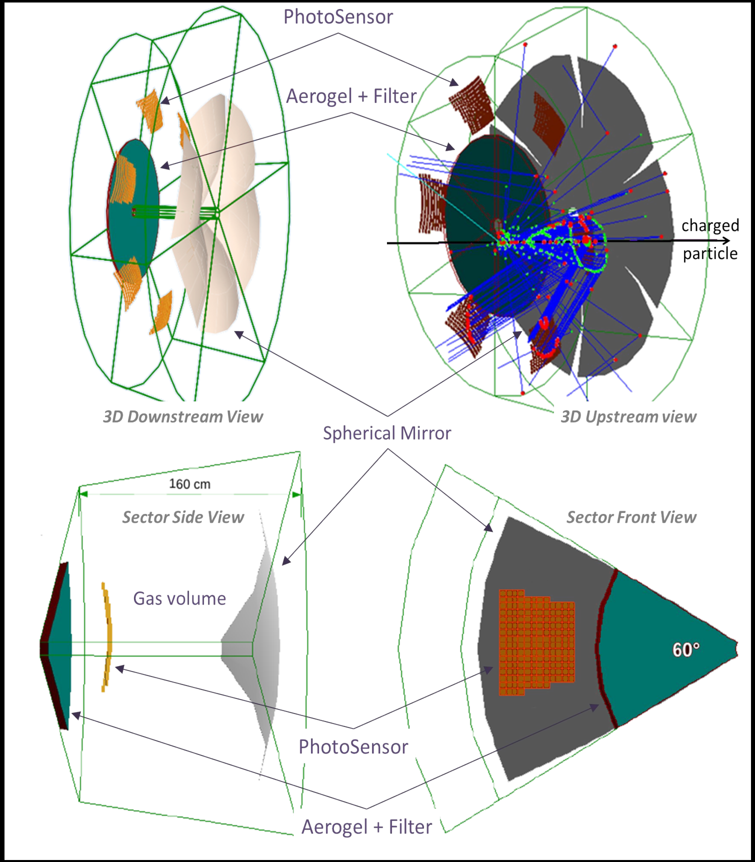
- https://distill.pub/2020/bayesian-optimization/
- https://distill.pub/2019/visual-exploration-gaussian-processes/
- https://www.borealisai.com/en/blog/tutorial-8-bayesian-optimization/

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)





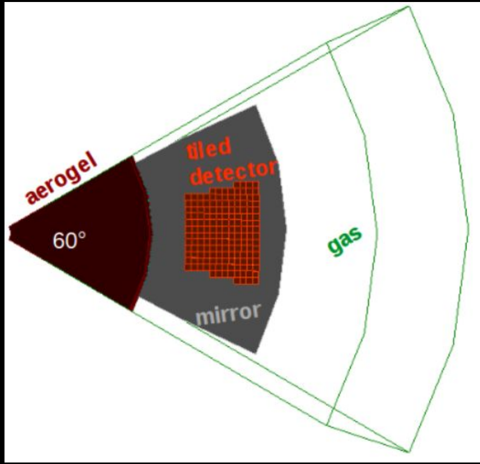
- Continuous momentum coverage.
- Simple geometry and optics, cost effective.
- Legacy design from INFN, see EICUG2017

- 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

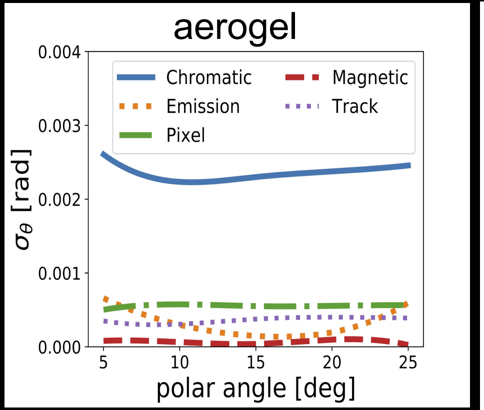
Goal is improve the distinguishing power of pions/kaons, hence:

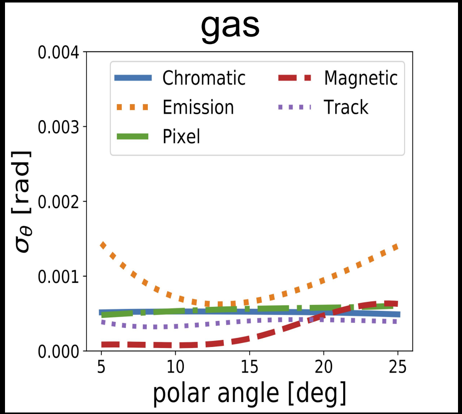
$$\boxed{N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}}$$

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

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

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

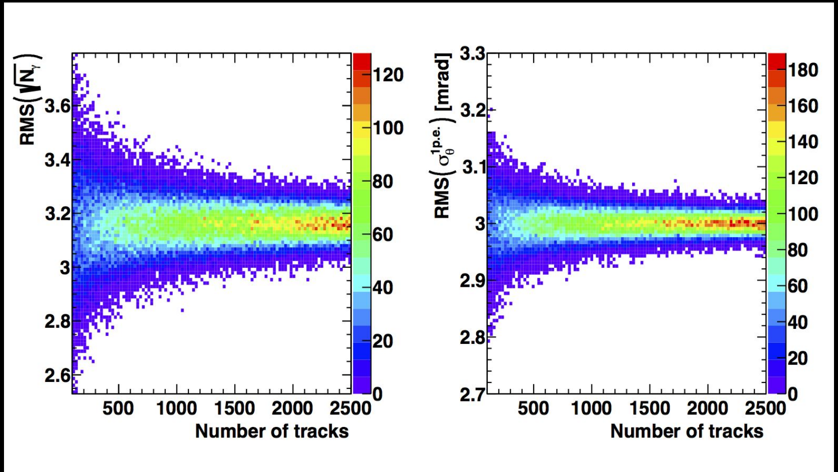
@ p_1 = 14 GeV/c (aerogel) and p_2 = 60 GeV/c (gas) considering the two parts disentangled

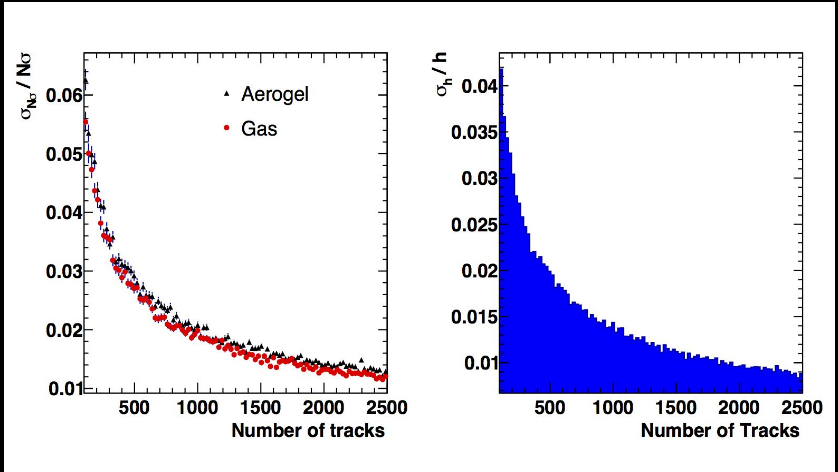
Noise Studies

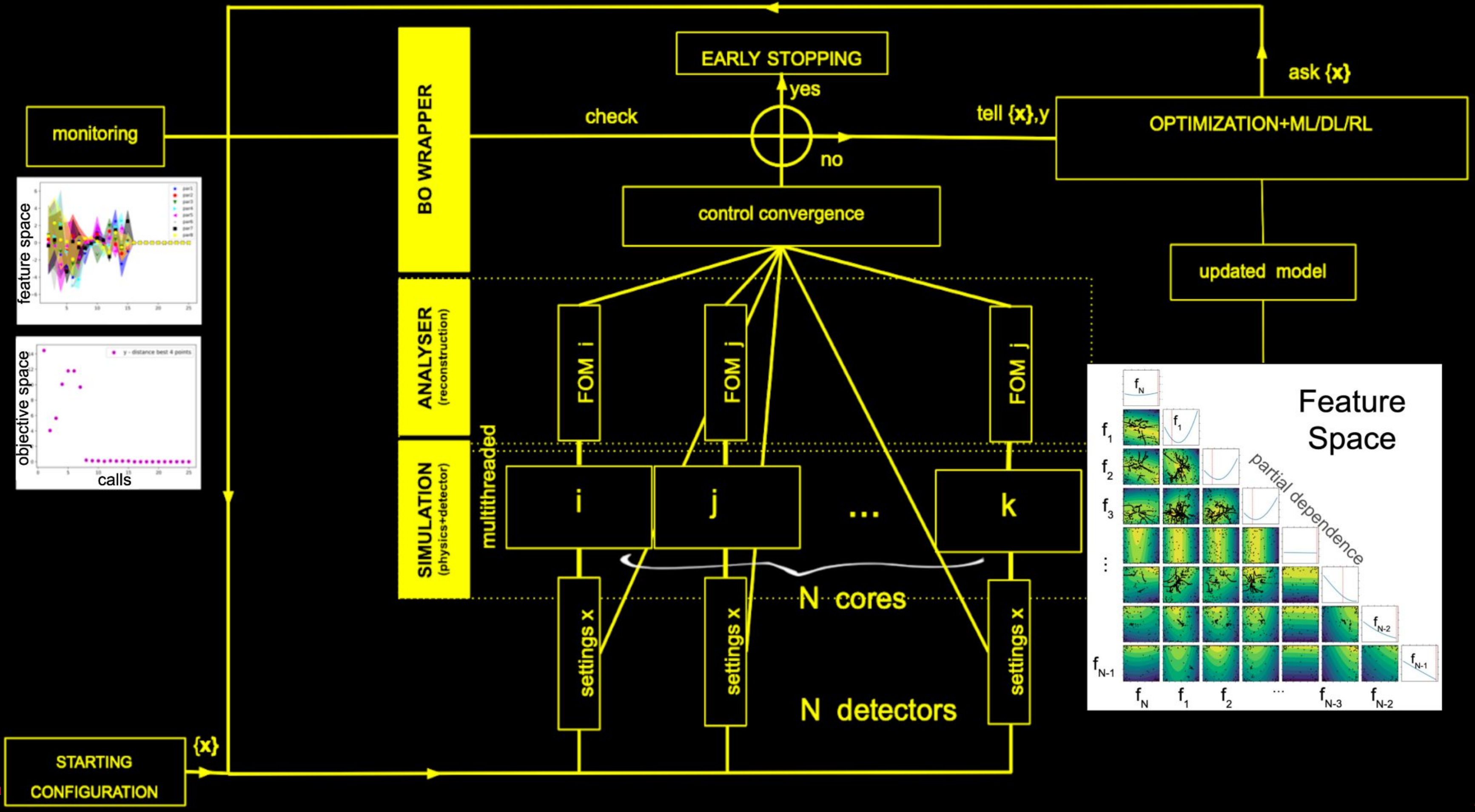
$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

- 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.



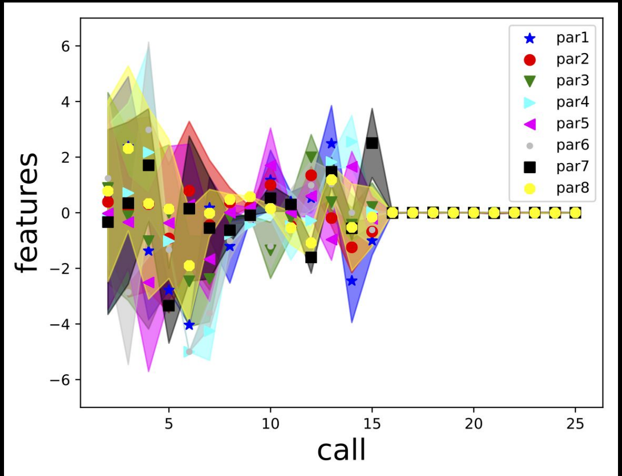


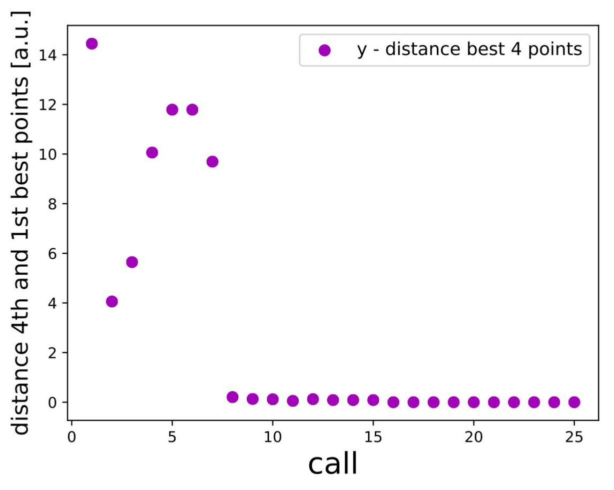


EARLY STOPPING
yes
check
no
tell {x},y
ask {x}
OPTIMIZATION+ML/DL/RL
monitoring
BO WRAPPER
control convergence
updated model
feature space
objective space
calls
ANALYSER
(reconstruction)
FOM i
FOM j
FOM j
SIMULATION
(physics+detector)
multithreaded
i
j
...
k
N cores
settings x
settings x
settings x
N detectors
Feature Space
partial dependence
STARTING CONFIGURATION
{x}

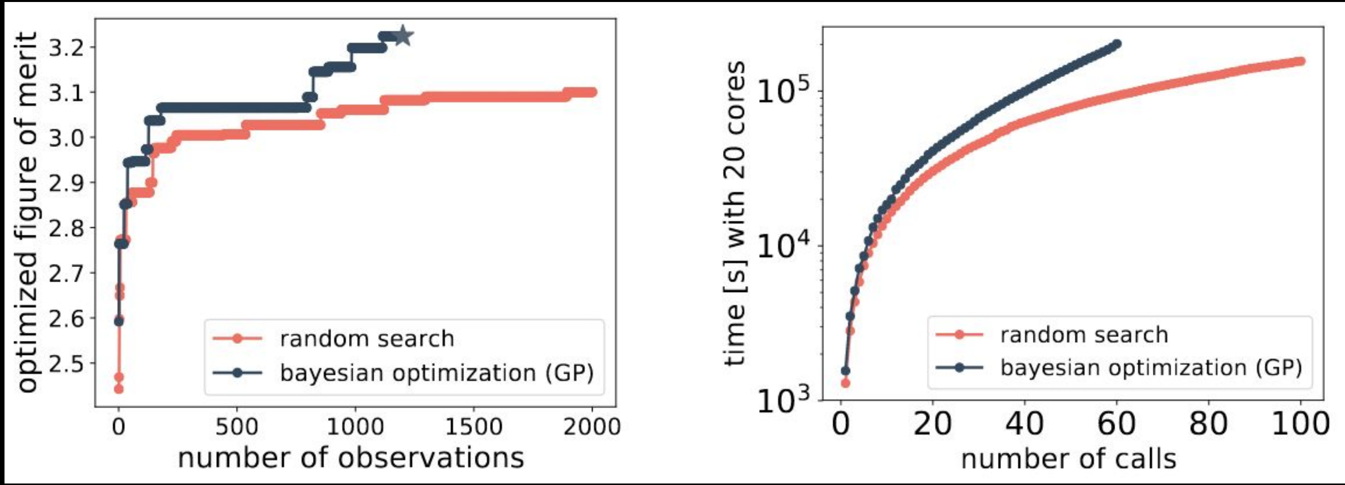
Convergence Criteria

- Can in general be applied in the design space, in the objective space, or looking at the behavior of the acquisition function.
- 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.





Comparison with Random Search



Each call:
400 tracks generated/core
20 cores

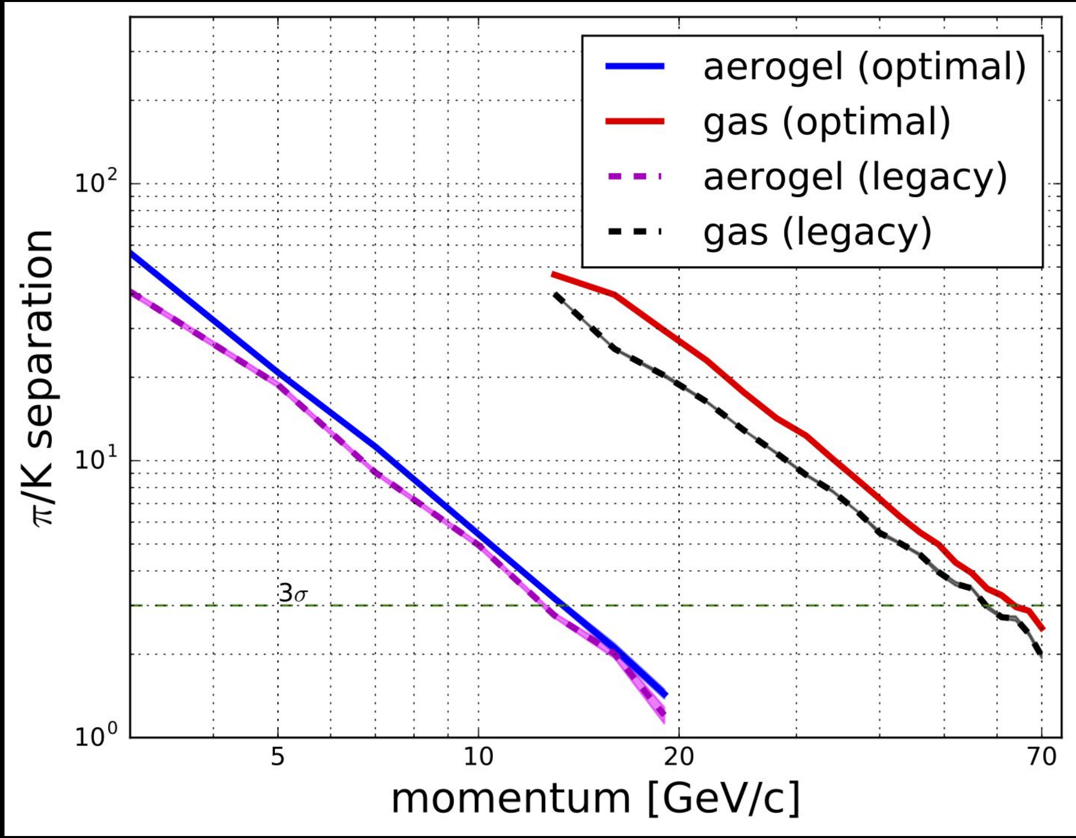
1 design point ~ 10 mins/CPU

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.

dRICH Performance at the optimal design point

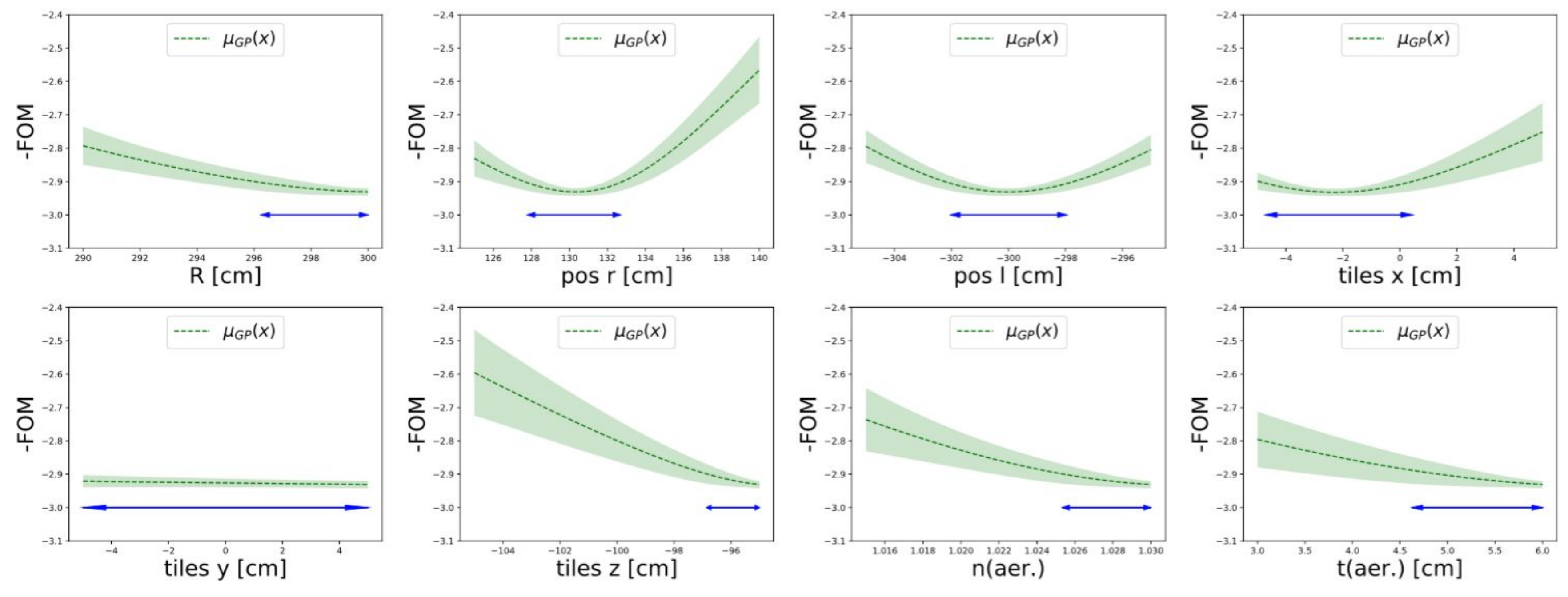


- Statistically significant Improvement in both parts.
- In particular in the gas region where the 5σ threshold shifted from 43 to 50 GeV/c and the 3σ one extended up to
- Notice that before this study we did not know “how well” the legacy design was performing.

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

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).



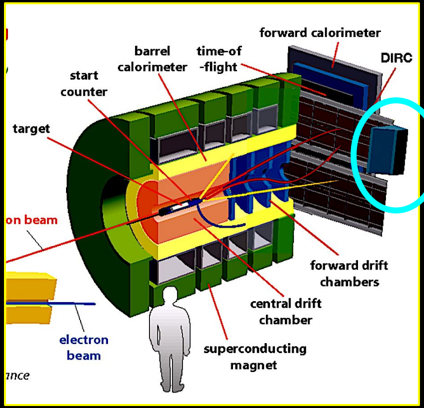
- Larger than the construction tolerances on each parameter.
Notice a small lateral shift of the tiles has negligible impact on the PID capability.

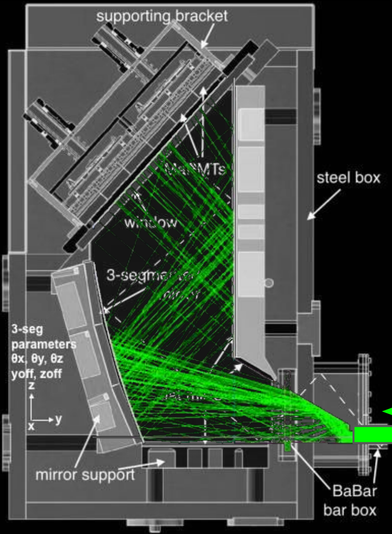
Frameworks and Deployment in the Industry

- 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.

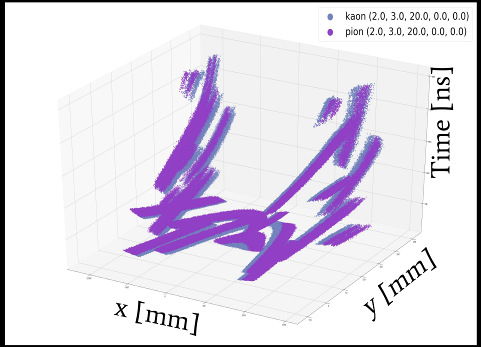
Other Applications: DIRC Alignment

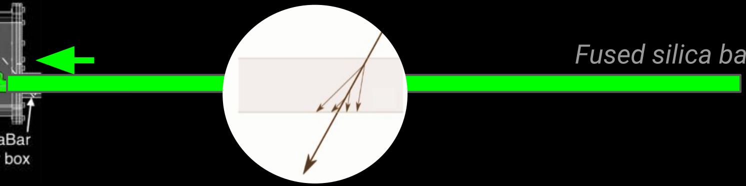




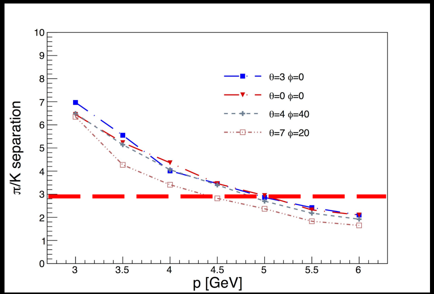
Optical box

3D Readout





Fused silica bars



π/K separation with DIRC

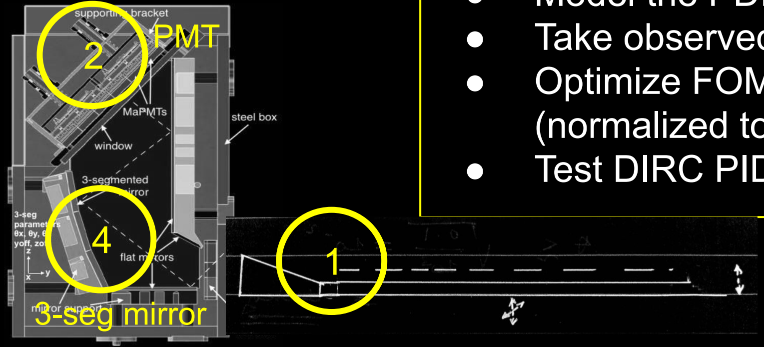
3D (x,y,t) readout allows to separate spatial overlaps.

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.

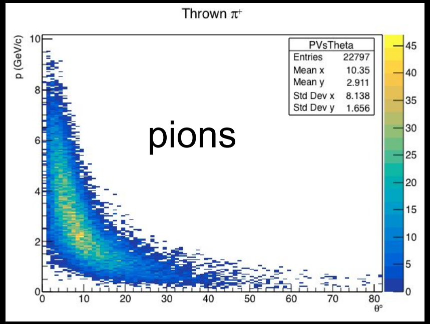
H12700 PMTs have a time resolution of O(200 ps) and read-out electronics giving time information in 1 ns buckets.

Alignment

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

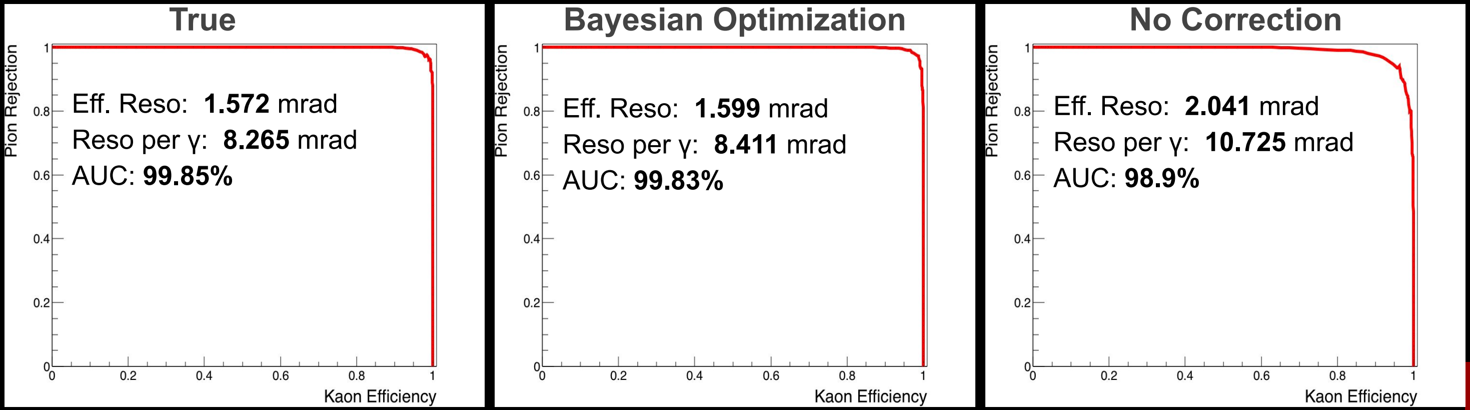


True:
3-seg mirror:
θx,θy,θz=(0.25,0.50,0.15) deg,
y = 0.50 mm;
bar: z = 2.00 mm;
PMT: (r,θ)=(1.50 mm,1.00 deg)

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)

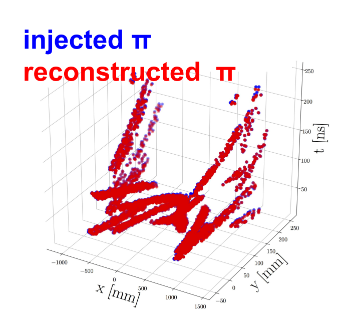
Pion rejection vs Kaon efficiency at large P

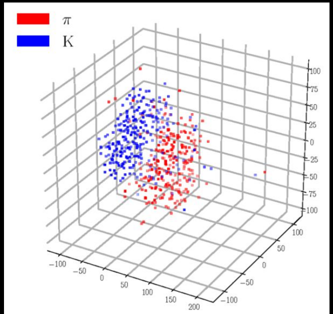


Hyperparameter Tuning

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

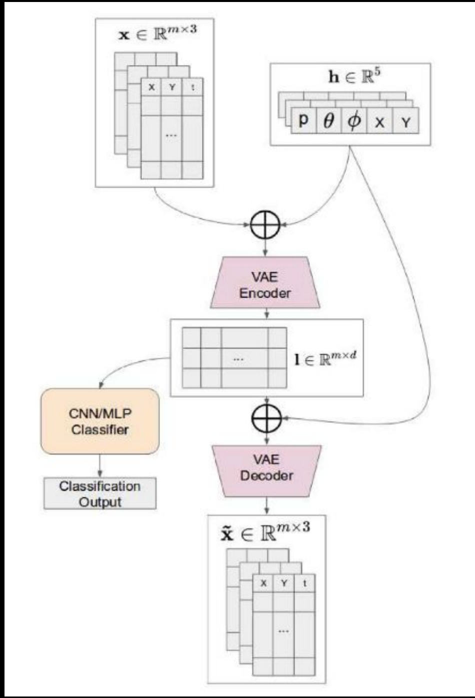
DeepRICH





latent space
t-SNE used for 3D visualization

injected



reconstructed

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]	$6.6\cdot 10^{-4}$

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 $\varepsilon_S * \varepsilon_B$. 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

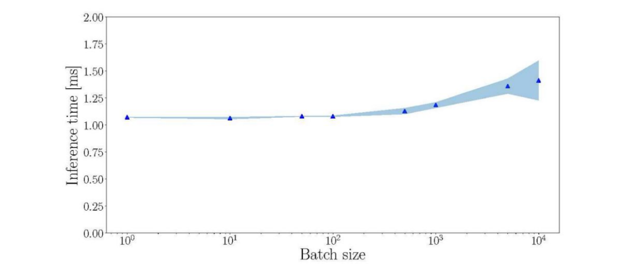


Figure 9. After training, the inference time is almost constant as a function of the batch size, meaning that the effective inference time—*i.e.*, the reconstruction time per particle—can be lower than a μs, the architecture being able to handle 10^4 particles in about 1.4 *ms* in the inference phase.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

Event generator tuning using BO

- MC event generators contain large number of parameters that must be determined by comparing the output of generator with experimental data.
- Generating enough events is extremely CPU intensive, prohibits performing a simple brute-force grid-based tuning of the parameters.
- MC event generator parameters can be accurately obtained using BO and minimal expert-level physics knowledge.
- A tune of the Pythia 8 event generator using e^+e^- events, with 20 tunable parameters, can be run on a modern laptop in just 2 days.
- Combining the BO approach with expert knowledge should enable faster tuning and facilitate the study discrepancies between MC and experimental data.

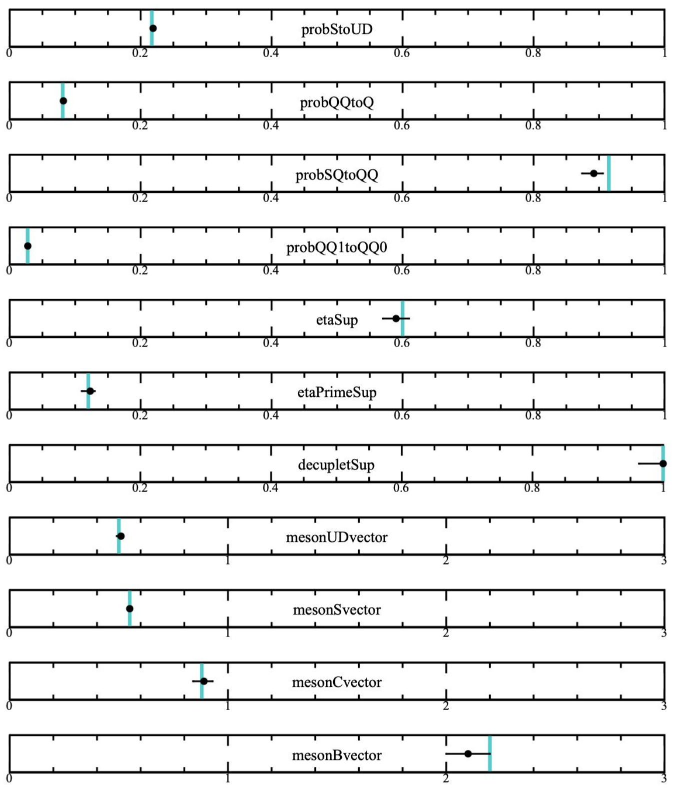
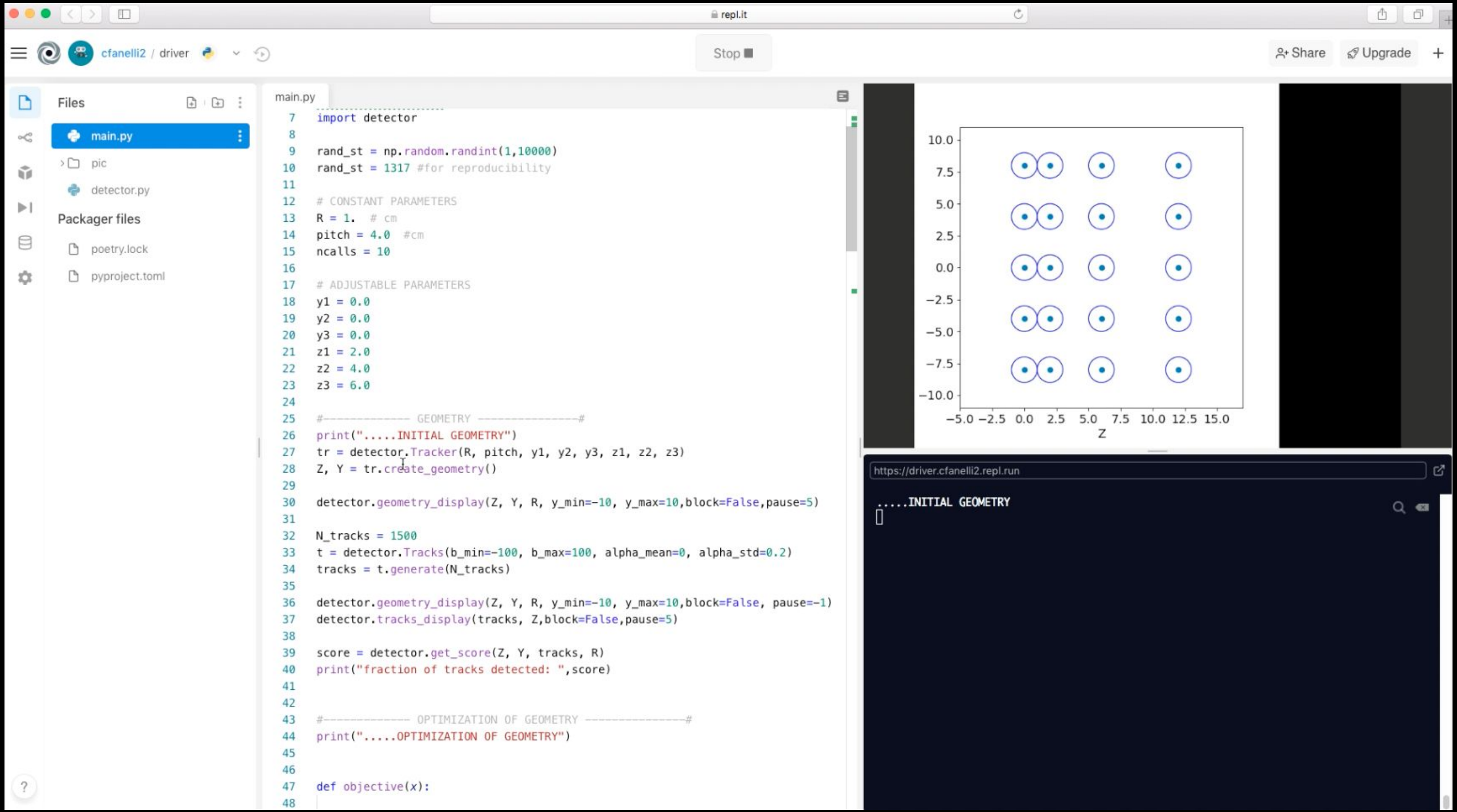


Figure 8. (Black points) Block 3 parameters from our optimal tune compared to their (vertical cyan lines) Monash values. The horizontal-axis ranges are the regions considered by SPEARMINT during tuning.

P. Ilten, M. Williams, Y, Yang JINST 12.04 (2017): P04028.

Toy Model



repl.it
cfanelli2 / driver
Stop
Share
Upgrade
Files
main.py
pic
detector.py
Packager files
poetry.lock
pyproject.toml
main.py
import detector

rand_st = np.random.randint(1,10000)
rand_st = 1317 #for reproducibility

# CONSTANT PARAMETERS
R = 1. # cm
pitch = 4.0 #cm
ncalls = 10

# ADJUSTABLE PARAMETERS
y1 = 0.0
y2 = 0.0
y3 = 0.0
z1 = 2.0
z2 = 4.0
z3 = 6.0

#------------ GEOMETRY ---------------#
print(".....INITIAL GEOMETRY")
tr = detector.Tracker(R, pitch, y1, y2, y3, z1, z2, z3)
Z, Y = tr.create_geometry()

detector.geometry_display(Z, Y, R, y_min=-10, y_max=10,block=False,pause=5)

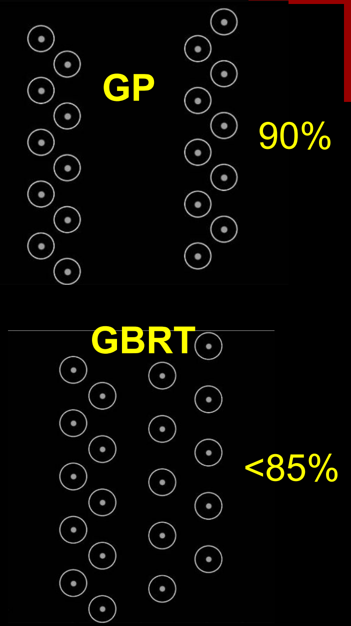
N_tracks = 1500
t = detector.Tracks(b_min=-100, b_max=100, alpha_mean=0, alpha_std=0.2)
tracks = t.generate(N_tracks)

detector.geometry_display(Z, Y, R, y_min=-10, y_max=10,block=False, pause=-1)
detector.tracks_display(tracks, Z,block=False,pause=5)

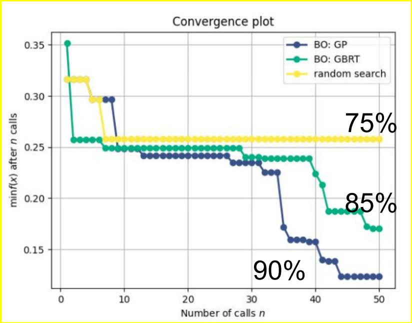
score = detector.get_score(Z, Y, tracks, R)
print("fraction of tracks detected: ",score)

#------------ OPTIMIZATION OF GEOMETRY ---------------#
print(".....OPTIMIZATION OF GEOMETRY")

def objective(x):
https://driver.cfanelli2.repl.run
.....INITIAL GEOMETRY
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−5.0 −2.5 0.0 2.5 5.0 7.5 10.0 12.5 15.0
Z

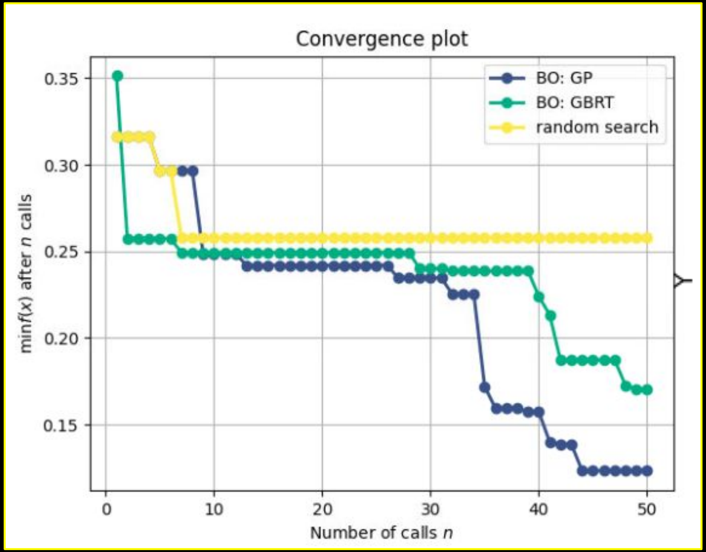


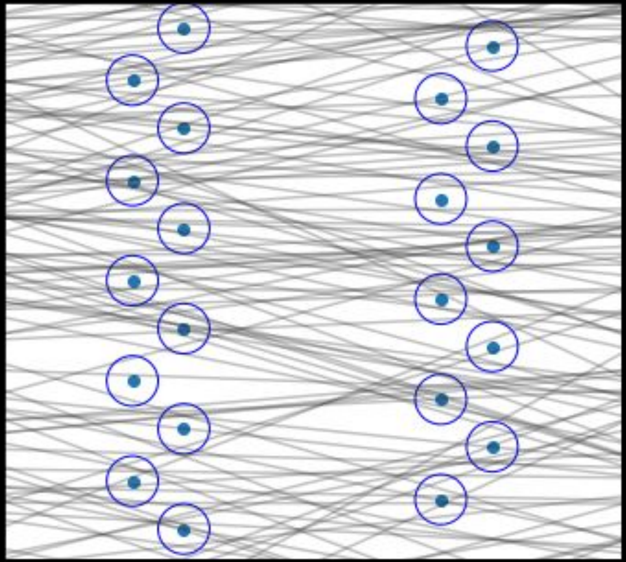
GP
90%
GBRT
<85%



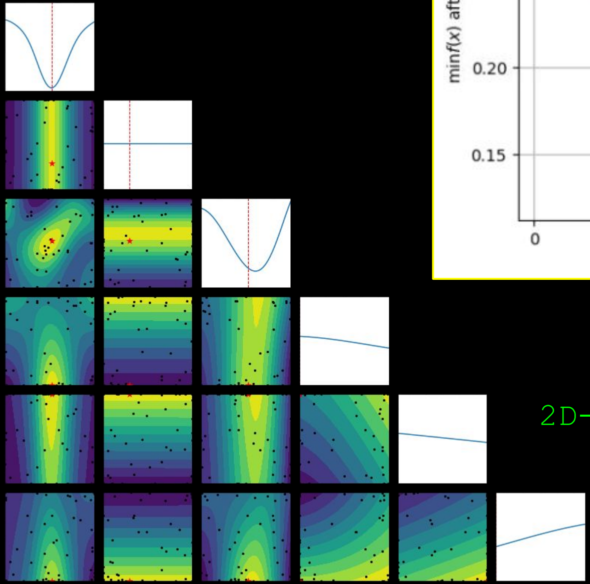
Convergence plot
BO: GP
BO: GBRT
random search
75%
85%
90%
min f(x) after n calls
Number of calls n

Toy Model





Objective: Efficiency is defined as at least two wires are hit

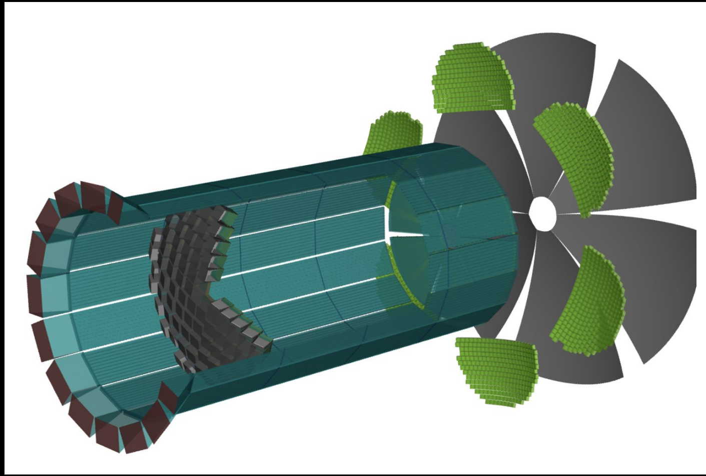


2D-plots of objective function and partial dependencies

Backup

Particle Identification with Cherenkov

η	θ	Nomenclature			Tracking						Electrons and Photons			π/K/p		HCAL		Muons
					Resolution	Relative Momentum	Allowed X/X_0	Minimum-pT	Transverse Pointing Res.	Longitudinal Pointing Res.	Resolution σ_E/E	PID	Min E Photon	p-Range (GeV/c)	Separation	Resolution σ_E/E	Energy	
< -4.6		↓ p/A	Far Backward Detectors	low-Q2 tagger														
-4.6 to -4.0					Not Accessible													
-4.0 to -3.5						Reduced Performance												Muons useful for bkg, improve resolution
-3.5 to -3.0			Central Detector	Backward Detector		σ_p/p ~0.2%×p⊕5%	~5% or less X	70-150 MeV/c (B=1.5 T)			1%/E ⊕ 2.5%/√E ⊕ 1%	π suppression up to 1:1E-4	20 MeV	≤ 10 GeV/c	≥ 3 σ	50%/ √E⊕10%	~500MeV	
-3.0 to -2.5																		
-2.5 to -2.0						σ_p/p~ 0.04%×p⊕2%												
-2.0 to -1.5									dca(xy) ~ 40/pT μm ⊕ 10 μm	dca(z) ~ 100/pT μm ⊕ 20 μm	2%/E ⊕(4-8)%/ √E ⊕ 2%	π suppression up to 1:(1E-3 - 1E-2)	50 MeV					
-1.5 to -1.0																		
-1.0 to -0.5				Barrel		σ_p/p ~0.04%×p⊕1%		200 MeV/c	dca(xy) ~ 30/pT μm ⊕ 5 μm	dca(z) ~ 30/pT μm ⊕ 5 μm	2%/E ⊕(12-14)%/ √E ⊕ (2-3)%	π suppression up to 1:1E-2	100 MeV	≤ 6 GeV/c		100%/ √E+10%		
-0.5 to 0.0																		
0.0 to 0.5																		
0.5 to 1.0																		
1.0 to 1.5				Forward Detectors		σ_p/p ~0.04%×p⊕2%		70 - 150 MeV/c (B = 1.5 T)	dca(xy) ~ 40/pT μm ⊕ 10 μm	dca(z) ~ 100/pT μm ⊕ 20 μm	2%/E ⊕ (4*-12)%/√E ⊕ 2%	3σ e/π up to 15 GeV/c	50 MeV	≤ 50 GeV/c		50%/ √E+10%		
1.5 to 2.0																		
2.0 to 2.5																		
2.5 to 3.0						σ_p/p ~0.2%×p⊕5%												
3.0 to 3.5																		
3.5 to 4.0		↑ e		Instrumentation to separate charged particles from photons		Reduced Performance												
4.0 to 4.5					Not Accessible													
> 4.6			Far Forward Detectors	Proton Spectrometer Zero Degree Neutral Detection														

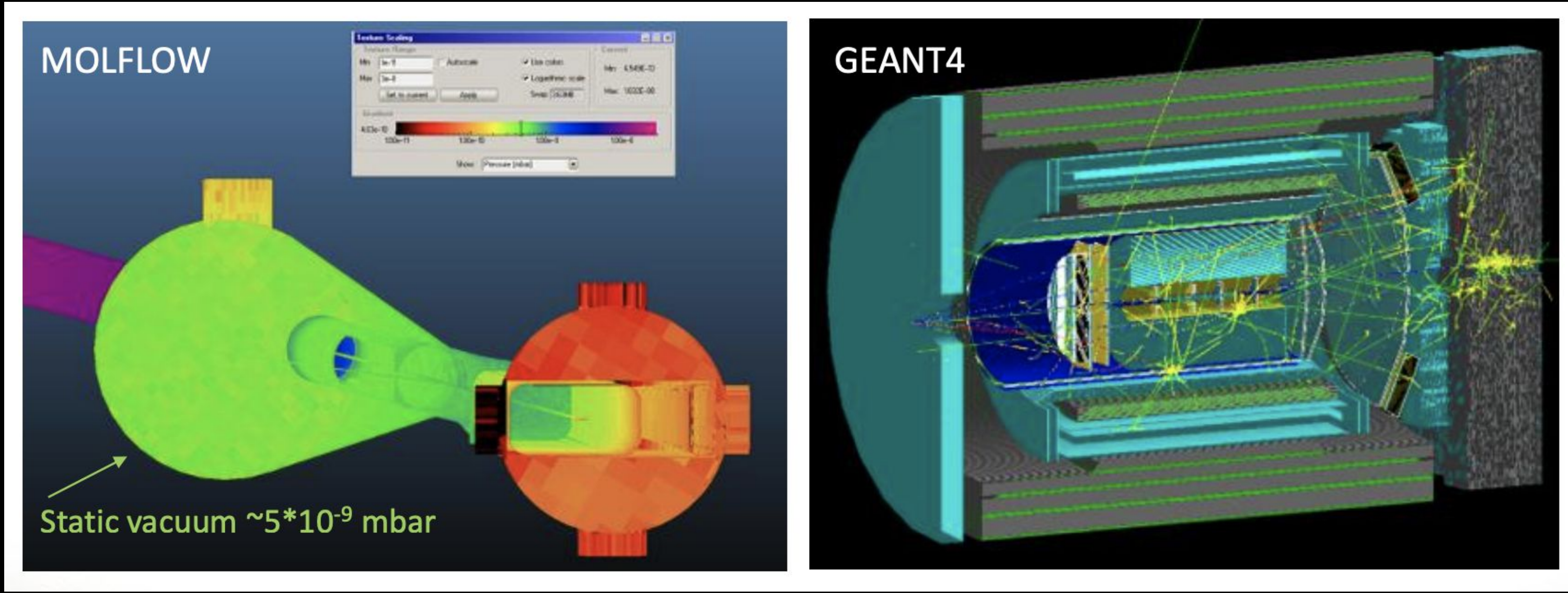


- Cherenkov detectors form the backbone of PID at EIC
  - Currently, all EIC detector designs use a dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction, a DIRC (detection of internally reflected Cherenkov light) in the barrel, and a modular RICH in the electron direction.
  - <u>Simulating these detectors is typically compute expensive</u>, involving many photons that need to be tracked through complex surfaces.
  - All three rely on pattern recognition of ring images in reconstruction, and <u>the DIRC is the one having the more complex ring patterns</u>!

Beam-gas induced background

Courtesy of Y. Furletova

Beam-gas interactions can cause hadronic showers, which produce high multiplicity events in the central detector apparatus

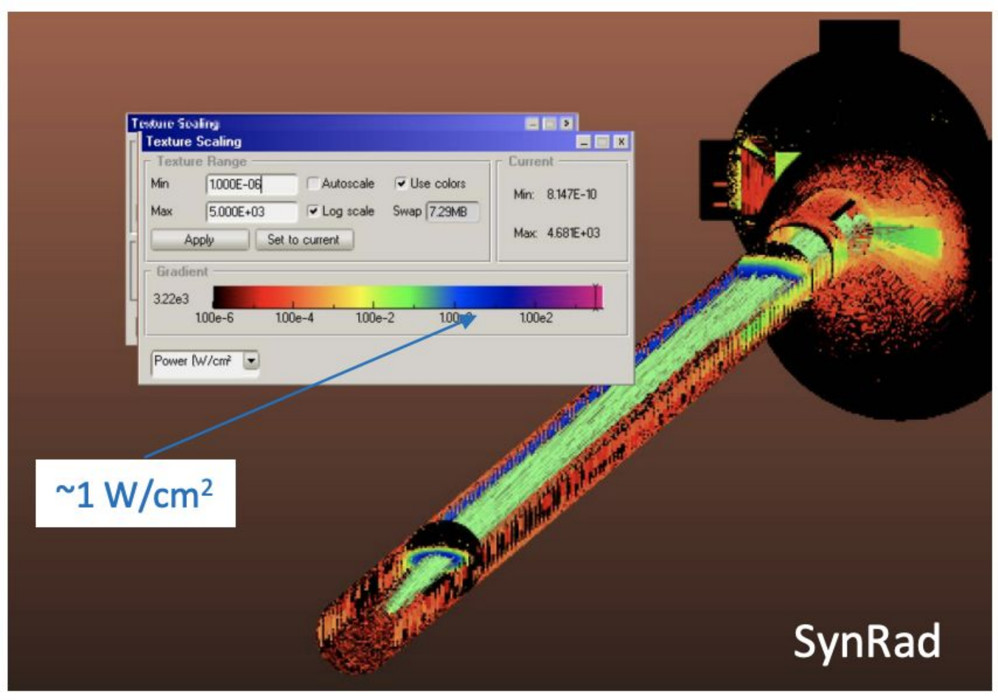


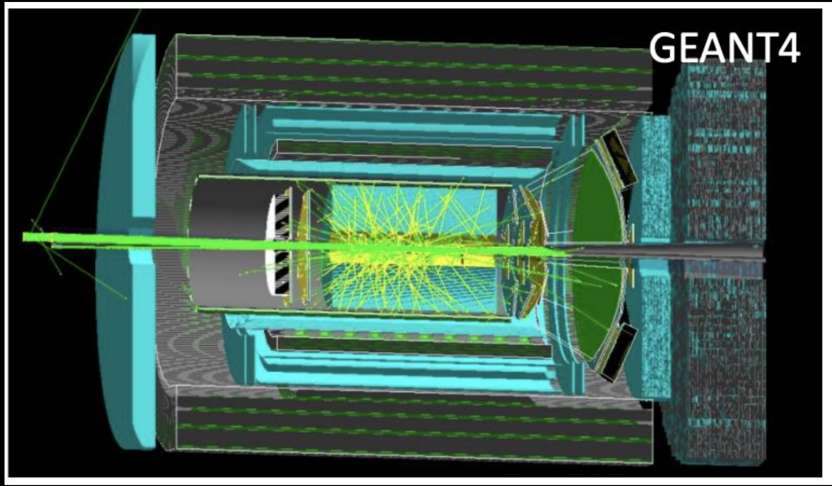
The GEANT simulation shows that for 10^{-9} mbar vacuum the contribution of such events to the data stream is relatively small compared to the physics collisions

Synchrotron radiation

Courtesy of Y. Furletova

Even in a configuration with the crossing angle, incoming electron trajectory bending in the upstream dipole and quadrupole magnetic fields produces substantial synchrotron radiation load

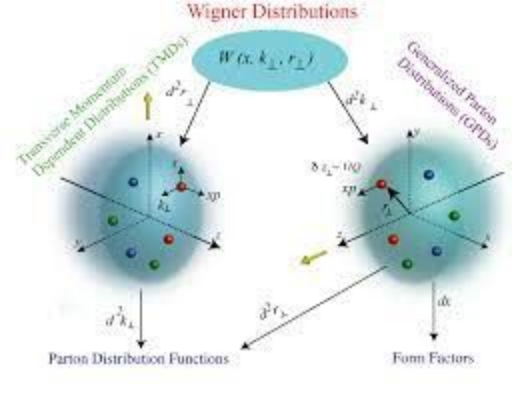




18 GeV electron beam 0.26 A

The design of absorbers and masks must be modeled thoroughly

Nucleon Tomography



5D tomography:
Wigner distribution— the "mother distribution"

Belitsky, Ji, Yuan (2003);
Lorce, Pasquini (2011)

$$W(x,\vec{k}_\perp,\vec{b}_\perp) = \int \frac{d^2\Delta_\perp}{(2\pi)^2} e^{i\vec{b}_\perp\cdot\vec{\Delta}_\perp} \int \frac{dz^- d^2 z_\perp}{16\pi^3} e^{ixP^+z^- - i\vec{k}_\perp\cdot\vec{z}_\perp} \langle P-\frac{\Delta}{2}|\bar{q}(-z/2)\gamma^+ q(z/2)|P+\frac{\Delta}{2}\rangle$$

$\int d\vec{b}_\perp$ $\qquad$ $\int d\vec{k}_\perp$

TMD $f(x,\vec{k}_\perp)$ $\qquad$ GPD $f(x,\vec{b}_\perp)$

$\int d\vec{k}_\perp$ $\qquad$ $\int d\vec{b}_\perp$ $\qquad$ $\int dx$

$f(x)$
PDF

$F(\vec{b}_\perp)$ Form factor

$\int dx$ $\qquad$ $\int d\vec{b}_\perp$

Q
charge