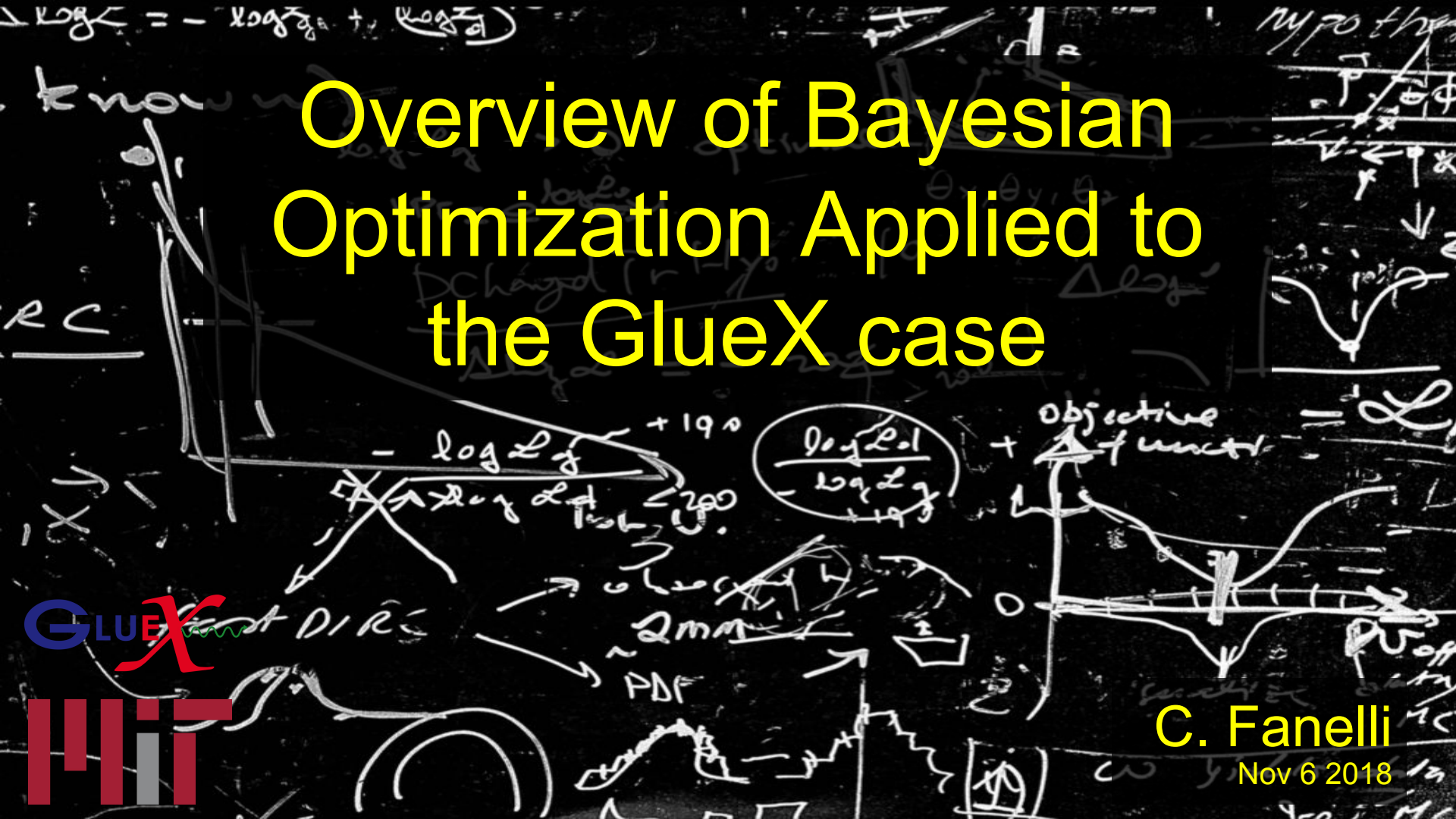
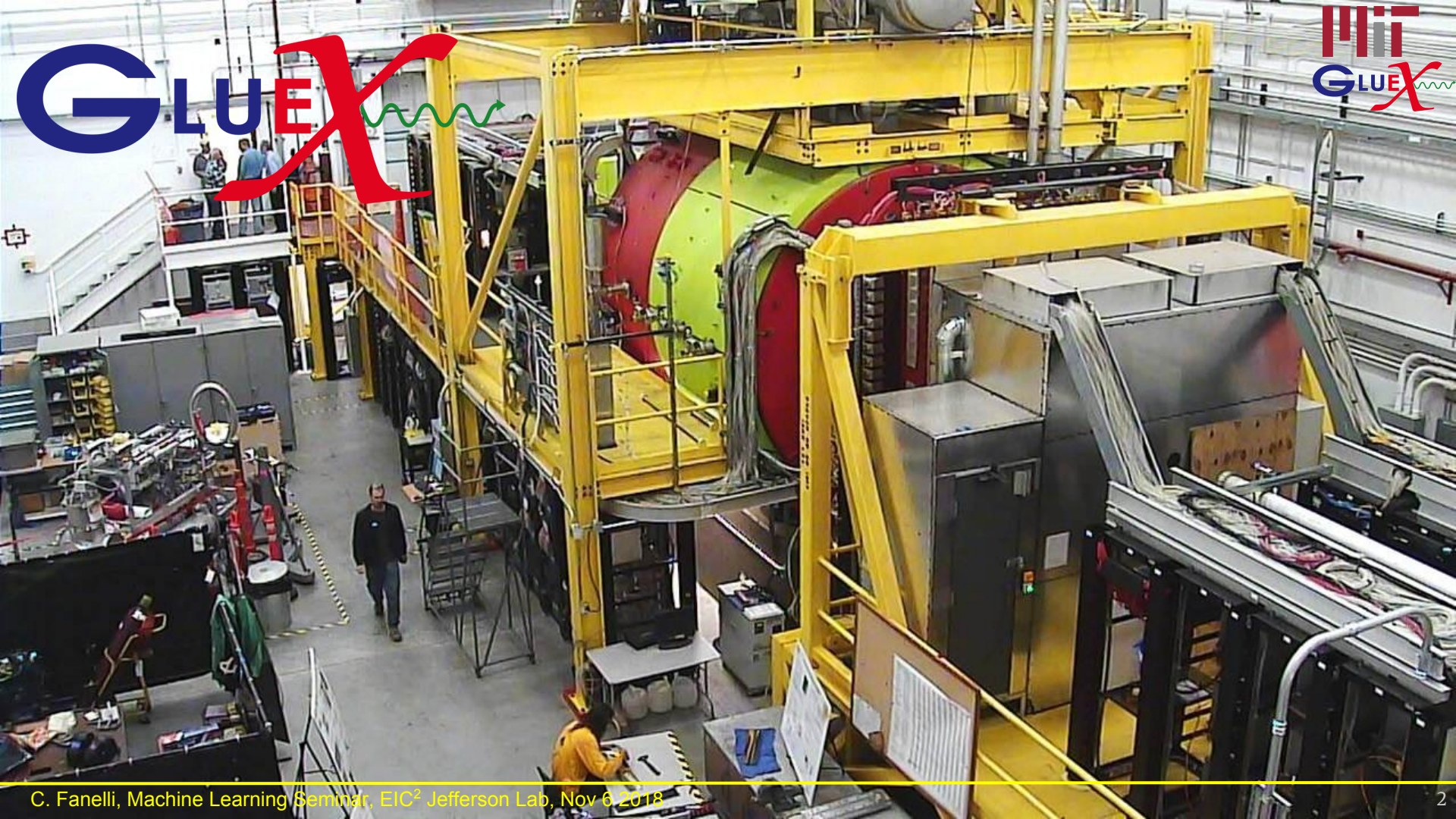
Overview of Bayesian Optimization Applied to the GlueX case

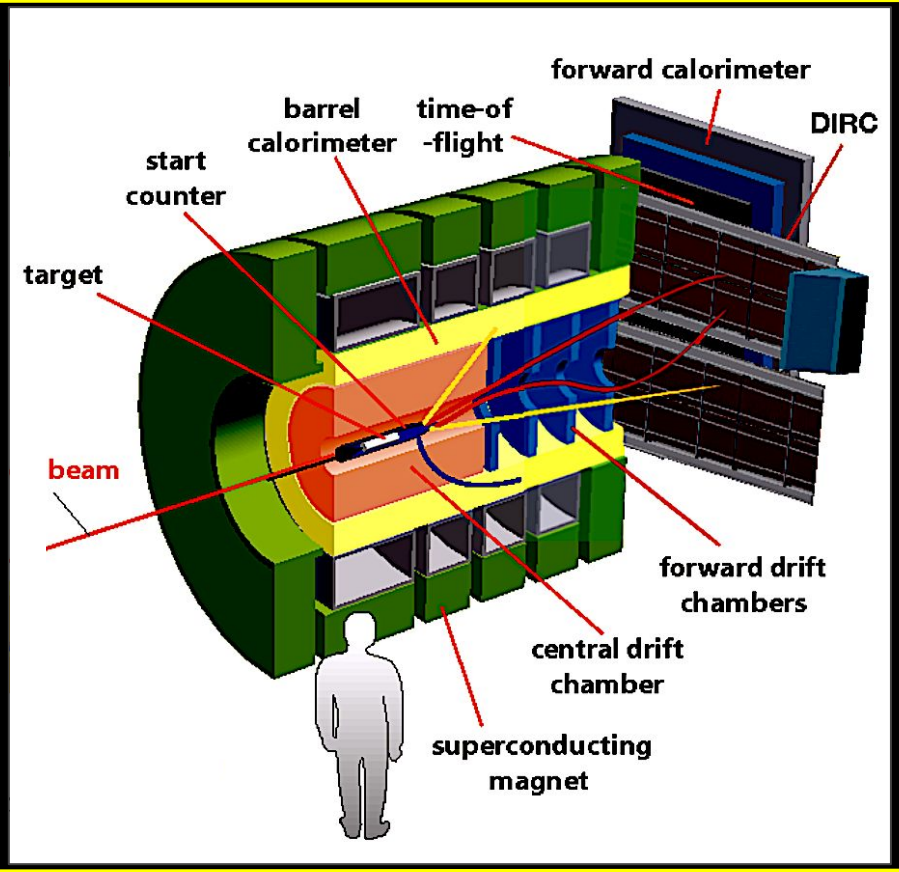


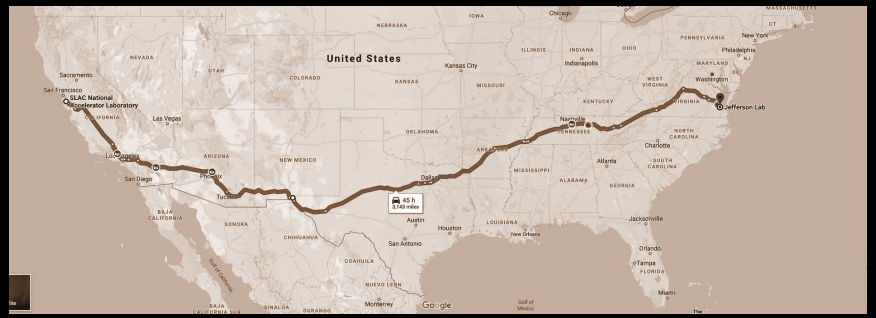
C. Fanelli

Nov 6 2018

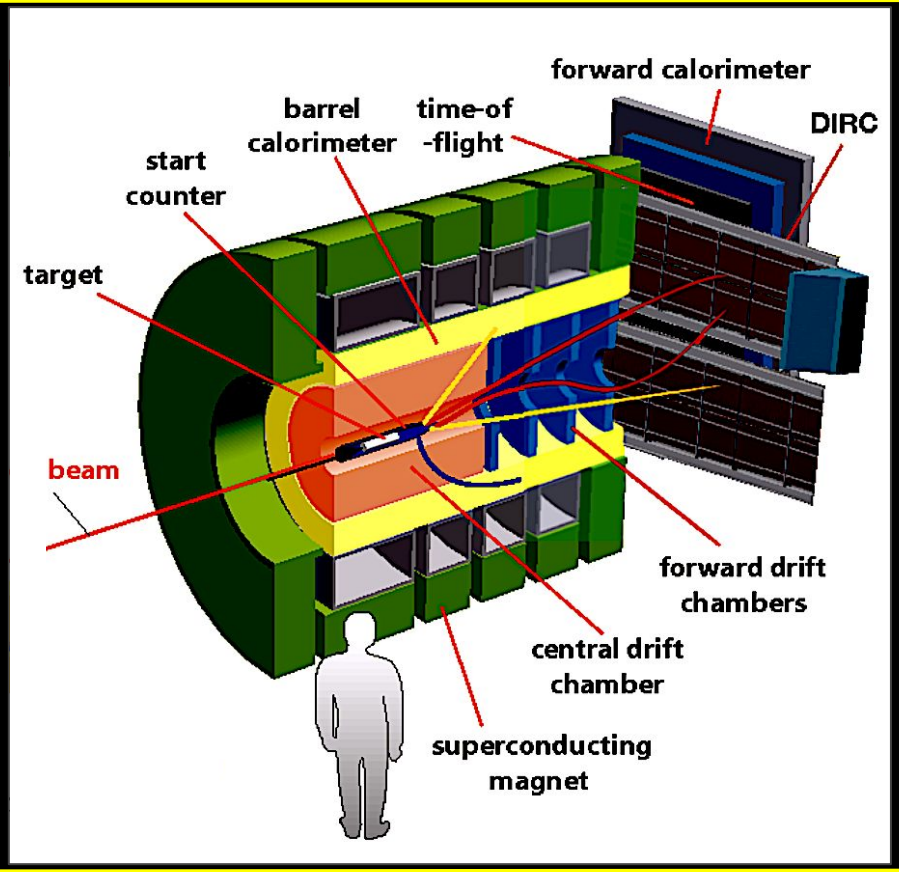

GLUEX
MIT
GLUEX

The GlueX detector in Hall D and the DIRC



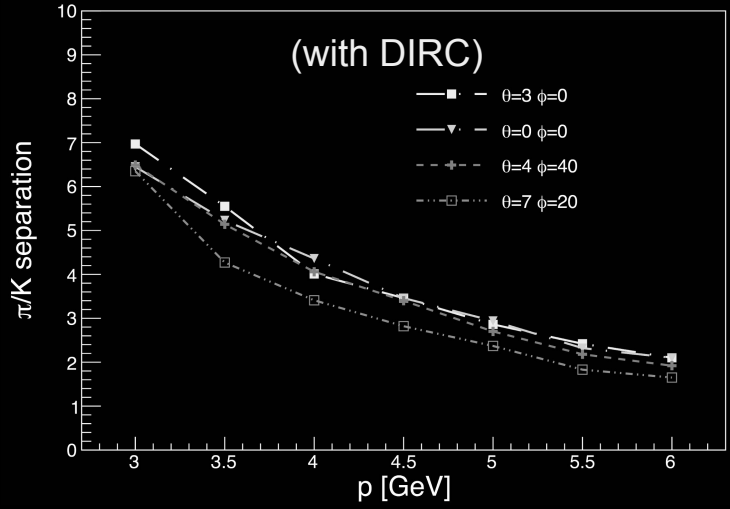


The GlueX detector in Hall D and the DIRC

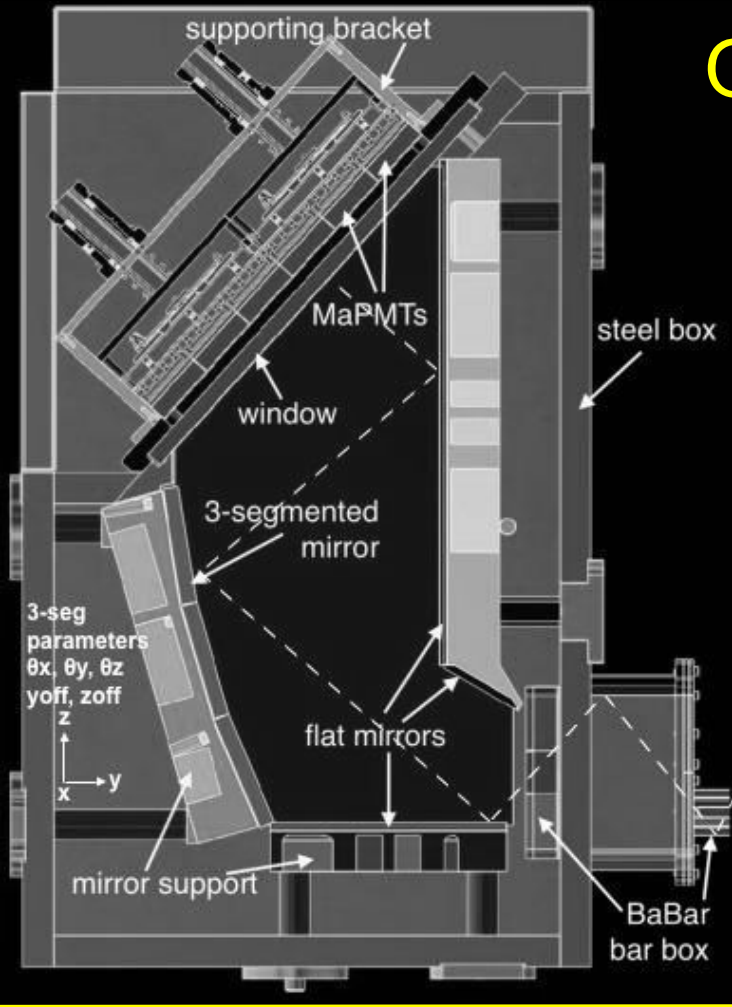


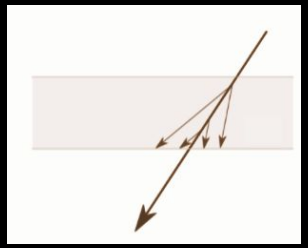
DIRC will improve GlueX PID capabilities

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

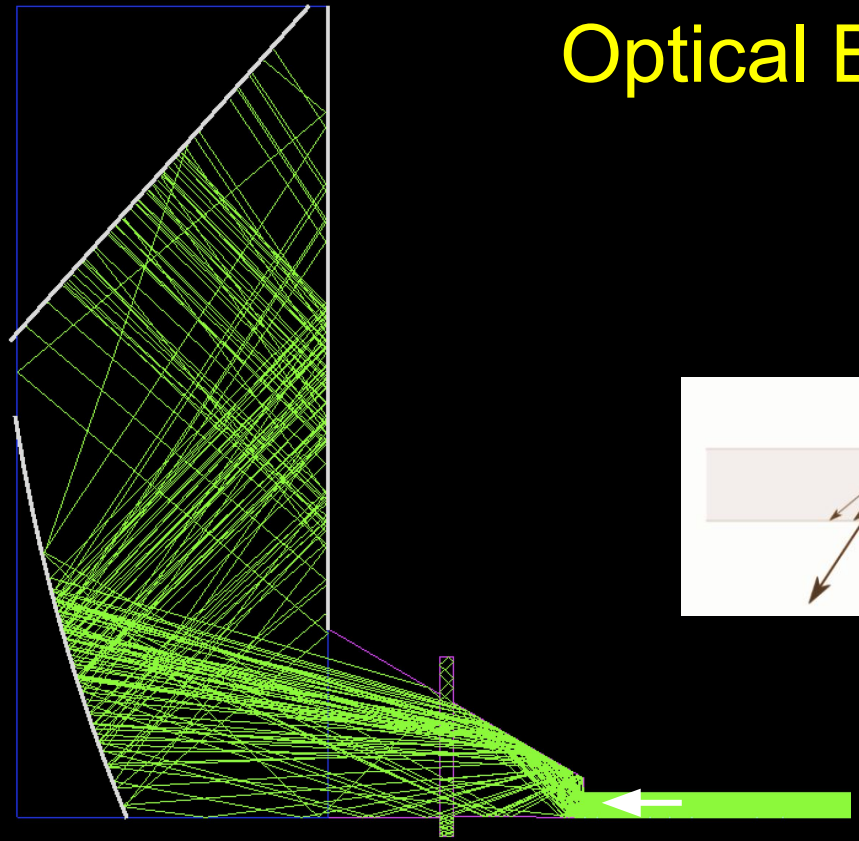


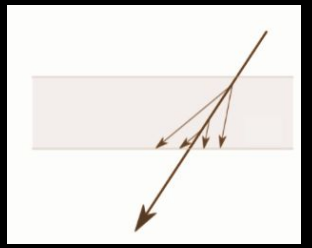
Optical Box


supporting bracket
MaPMTs
steel box
window
3-segmented mirror
3-seg parameters
θx, θy, θz
yoff, zoff
z
y
x
flat mirrors
mirror support
BaBar bar box

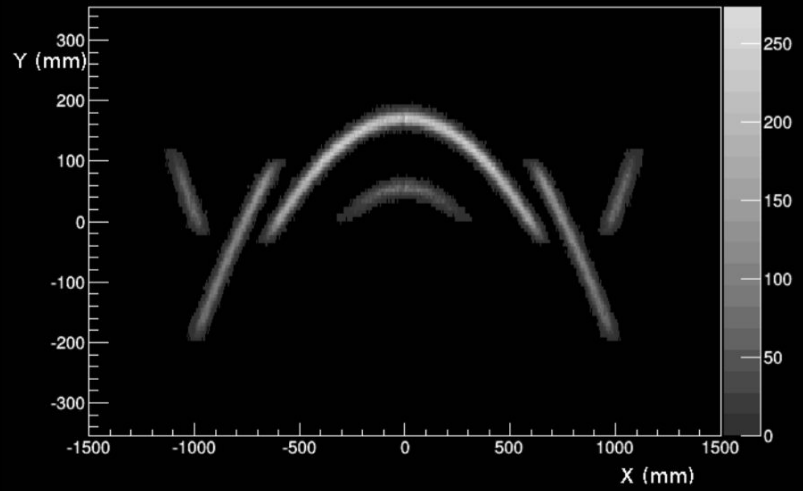


Optical Box

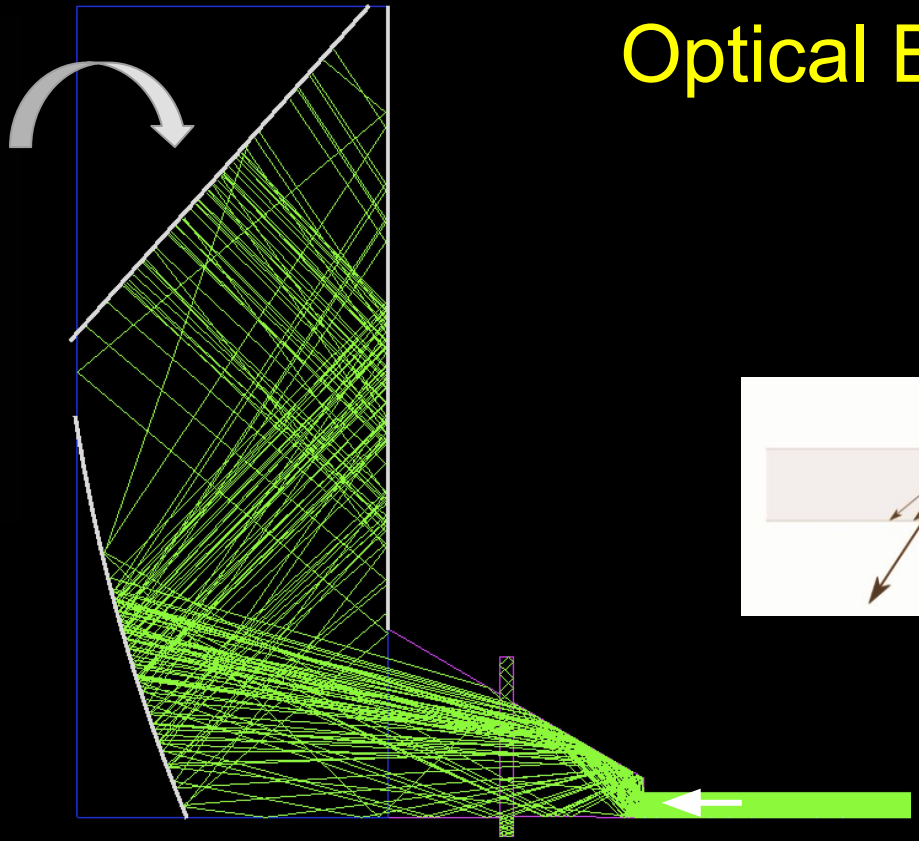


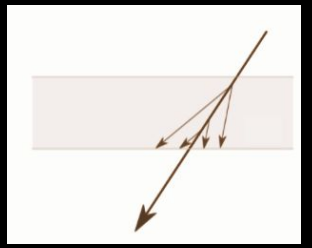


Optical Box



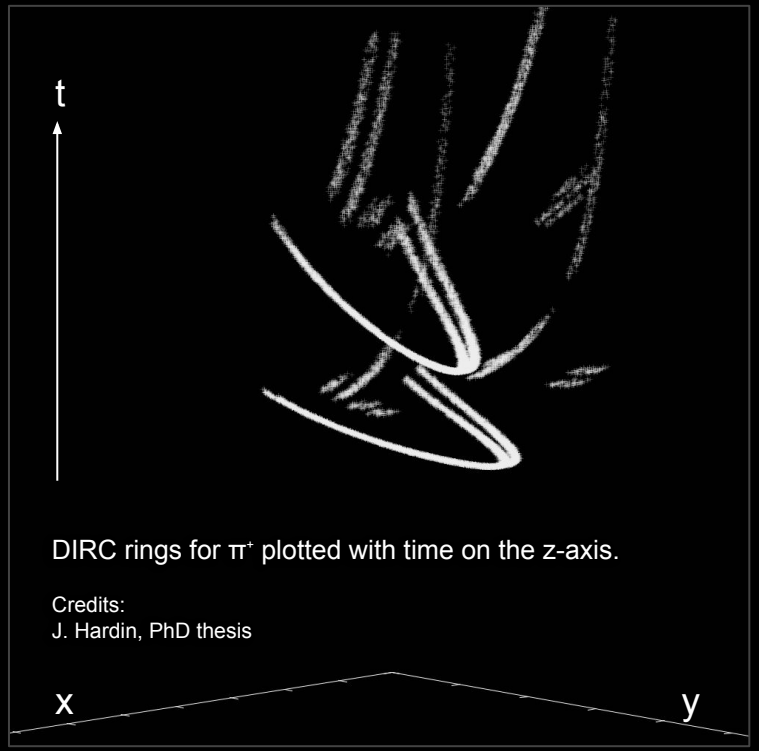
Cherenkov Photon “Ring”
in PMT plane





On average about ~ only O(50) photons detected per particle!

Hit Patterns



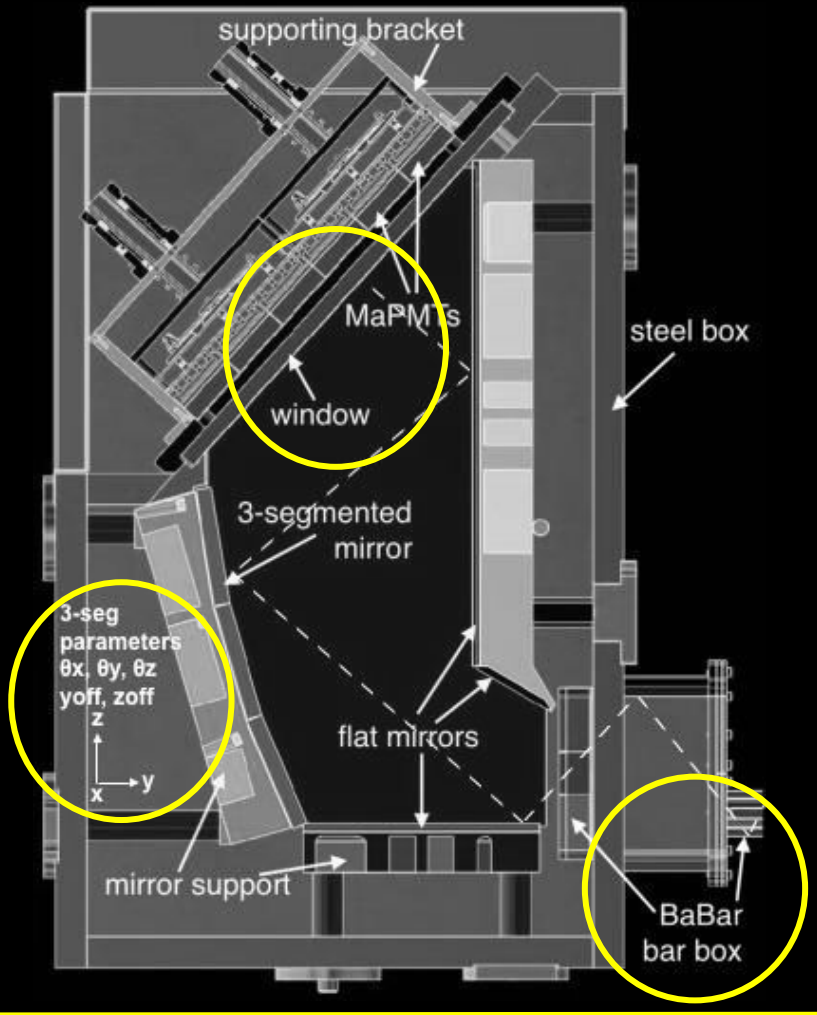
DIRC rings for π^+ plotted with time on the z-axis.

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)

Misalignments



- The optical box will be filled by distilled water (refractive index close to bars).
- Optical box made by several components, system for calibration.
- During data-taking this becomes a 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

 # offsets $\gtrsim$ O(10)

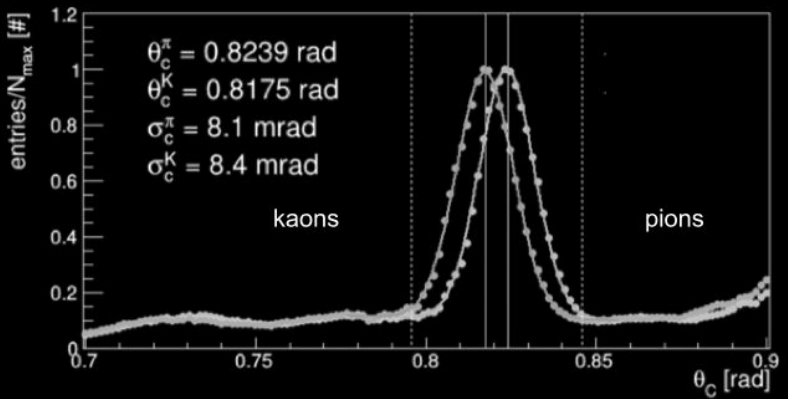
Reconstruction Algorithms and PID

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)

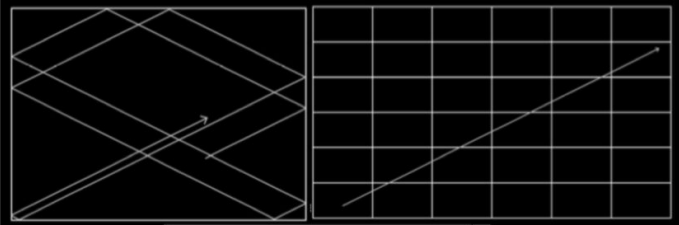


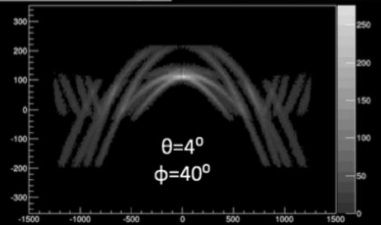
KDE-based

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

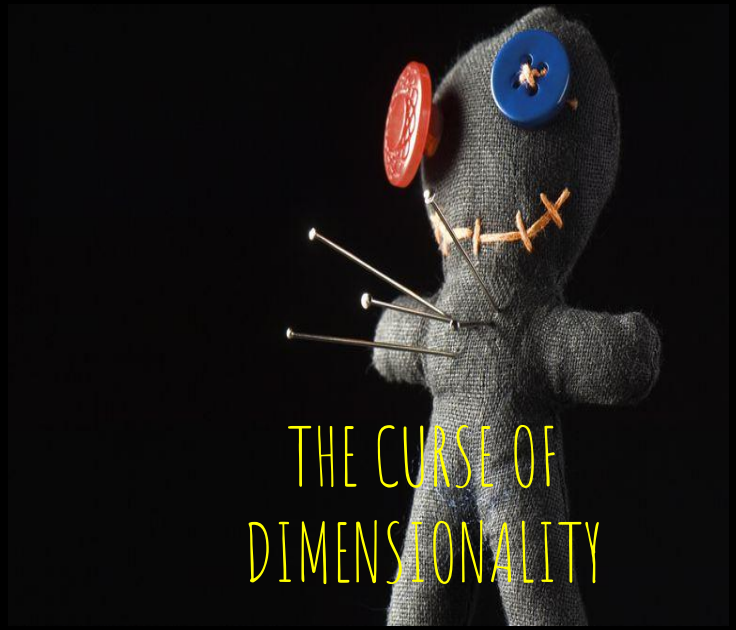
Optimization problems

- e.g. tuning hyperparameters of a ML problem is often more of an art than science.
  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.



Optimization problems

- e.g. tuning hyperparameters of a ML problem is often more of an art than science.
  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.

Likelihood
How probable is the evidence given that our hypothesis is true?

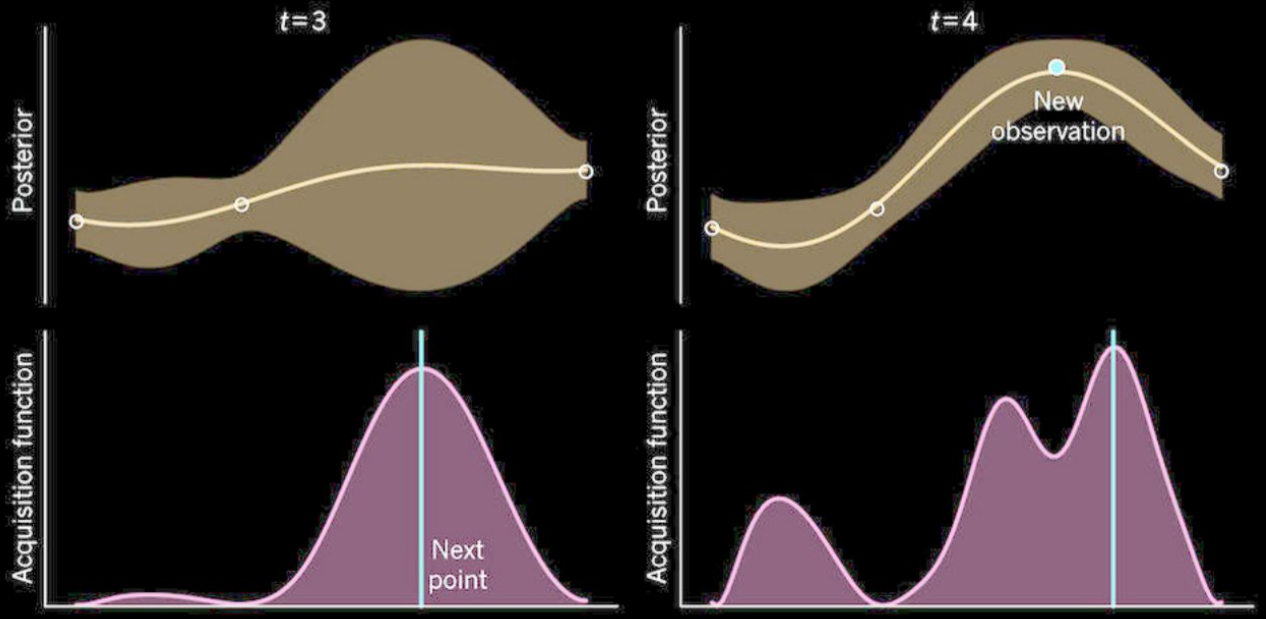
Prior
How probable was our hypothesis *before* observing the evidence?

$$P(H \mid e) = \frac{P(e \mid H)\, P(H)}{P(e)}$$

Posterior
How probable is our hypothesis given the observed evidence?
(Not directly computable)

Marginal
How probable is the new evidence under all possible hypotheses?
$P(e) = \sum P(e \mid H_i)\, P(H_i)$

Bayesian Optimization

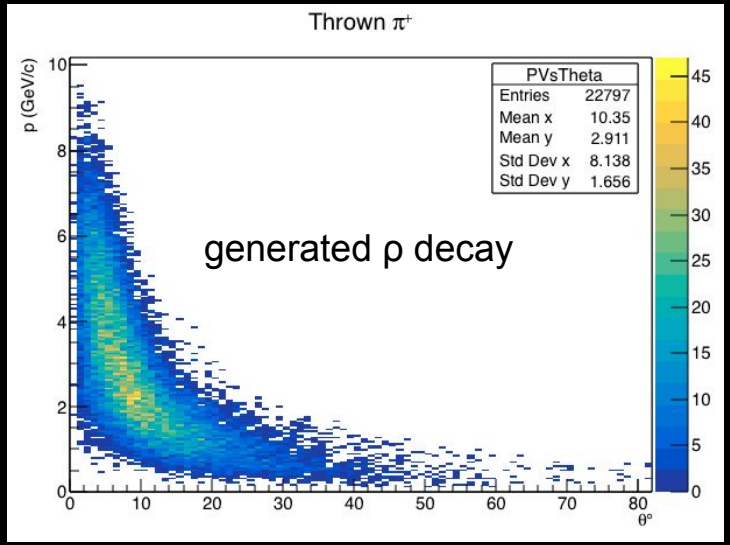


- BO worked amazingly well for tuning MC (Ilten, Williams, Yang [1610.08328]).
- BO could align the DIRC detector leveraging current reconstruction algorithms.

BO is part of wider class called Sequential model based algorithms

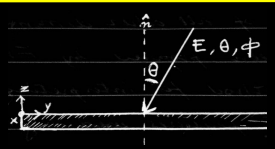
Pure sample of particles for alignment

- 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.)



3D combining different particles

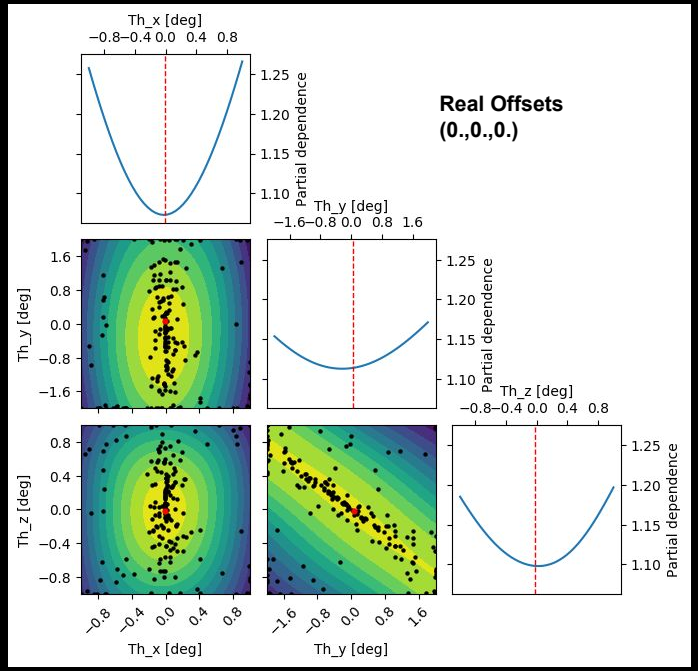
Toy-model: sampling 100 pions/call in range
E [GeV], θ [deg], φ [deg]: [2.5,3.0], [2,4], [1,45]

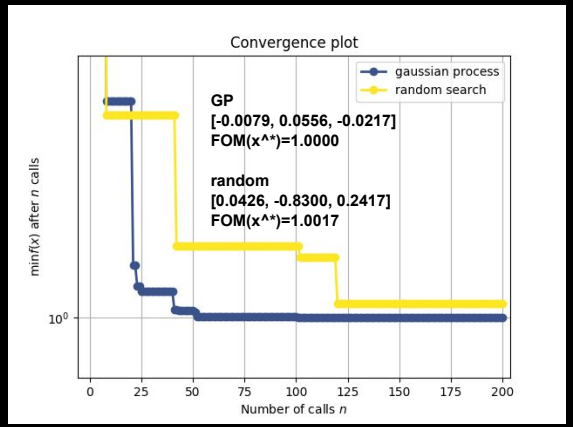


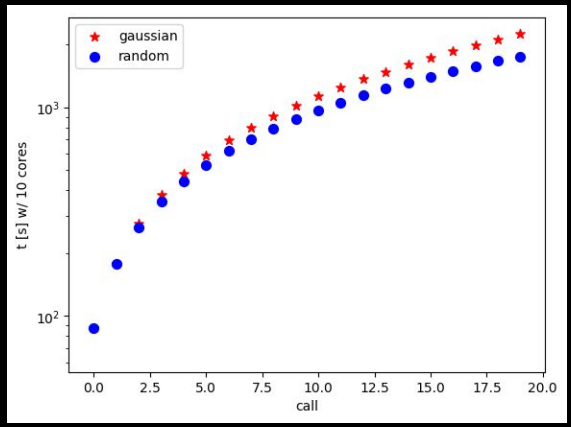
Recipe: For each call of the optimizer, M offset points are explored using N different particles (for each call). The total number of calls is T

T=20 M=10 N=100
Particles used = 2000
Points explored = 200

FoM = LogL normalized to a default alignment

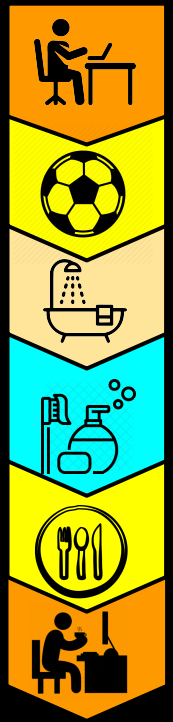




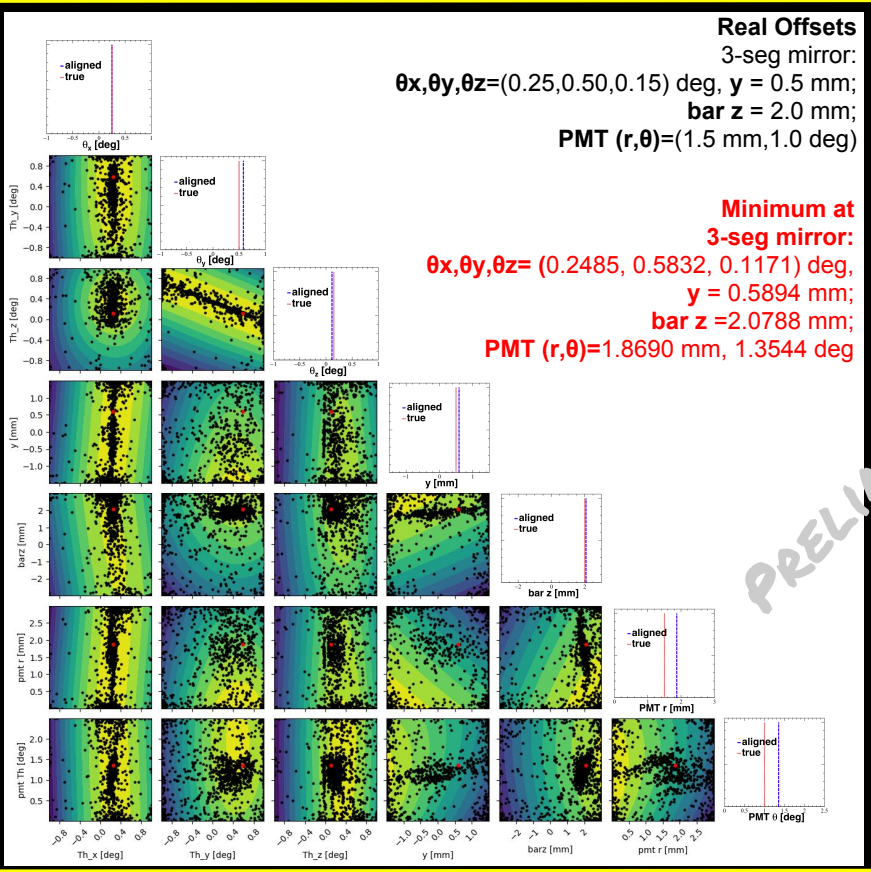


Multi-dimensional case

Launch and
wait for results



7D with main offsets - preliminary



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

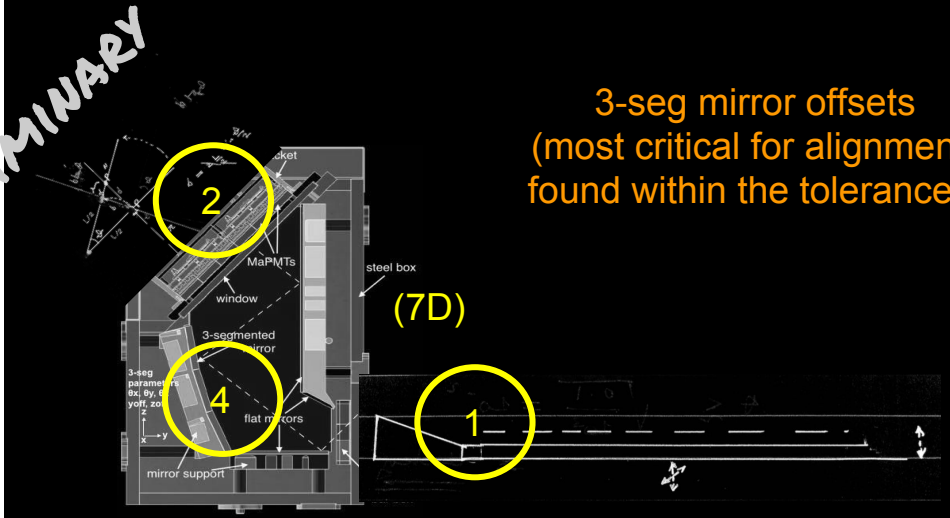
Recipe: For each call of the optimizer, M offset points are explored using N different particles (for each call). The total number of calls is T

T=120 M=10 N=125

Particles used = 15000

Points explored = 1200

FoM = LogL normalized to a default alignment

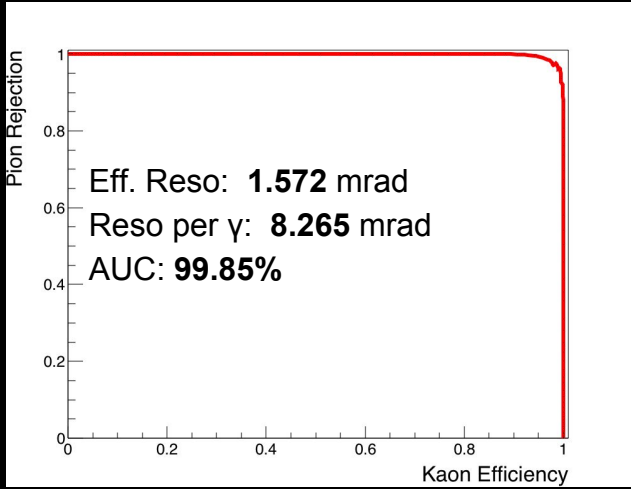


3-seg mirror offsets
(most critical for alignment)
found within the tolerances.

Resolutions Vs Offsets

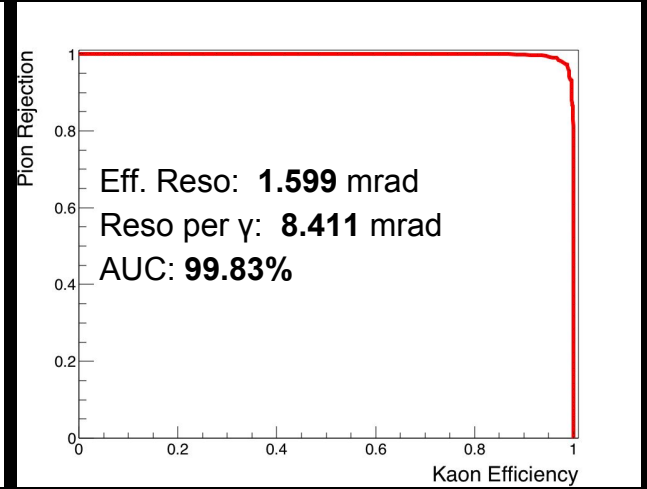
correct

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)



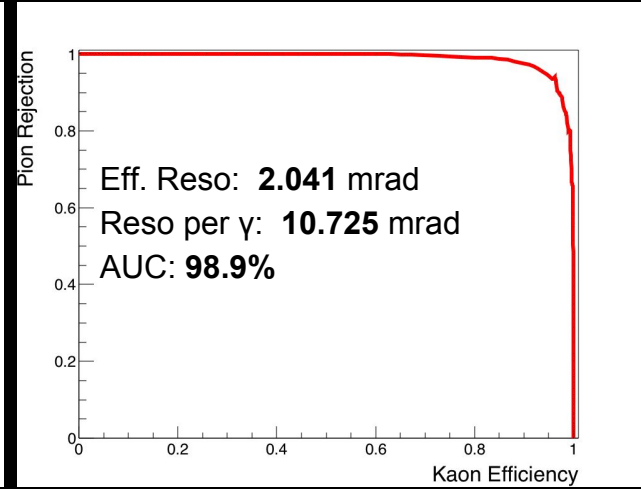
calibrated

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)



nominal

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)

Final Remarks

- This work just started, currently designing a final strategy.
  - BO can be “divided” into blocks of parameters.
  - test different BOs, FOMs (e.g. χ^2).
  - explorations can be distributed in parallel.

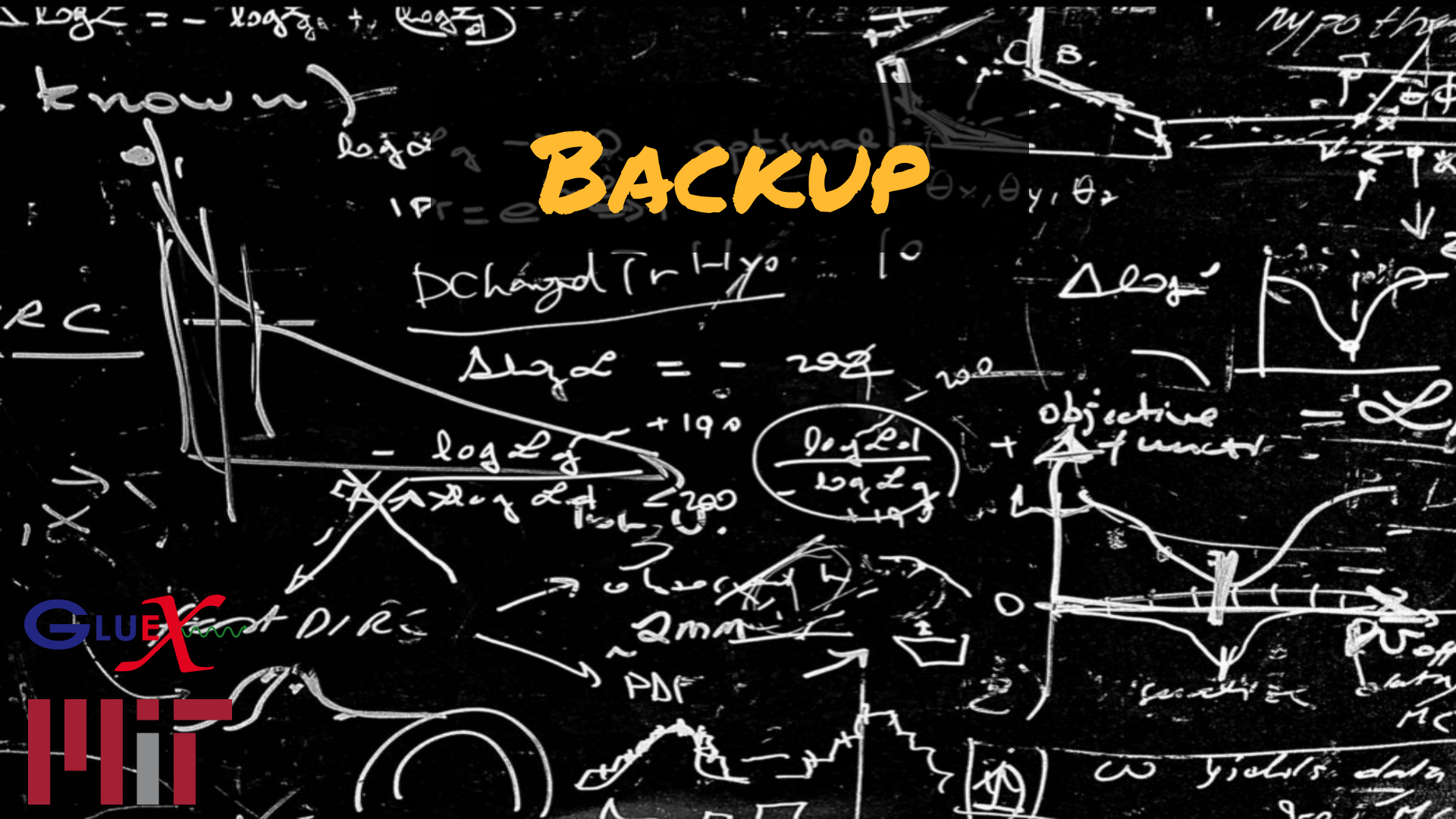
- We end up introducing some “hyperparameters” (N,M,...) - beyond those of the BO, which in principle we could tune.

- Several ways to assign uncertainty to the calibrated parameters (didn’t discuss) (e.g. varying each parameter independently and holding all other parameters fixed to their tuned values, running multiple calibrations in different bins of kinematics, etc…).

Summary

- Bayesian Optimization is a viable approach to optimize noisy black-box functions.
- In this preliminary work, a first closure test has been shown for the DIRC alignment.
- This work will encompass several aspects and reflects understanding of detector and development of reconstruction code, data preprocessing, testing of BOs, etc.
- Aiming to extend alignment to a larger space (e.g. PMTs), and run with real data after commissioning of detector.

Backup



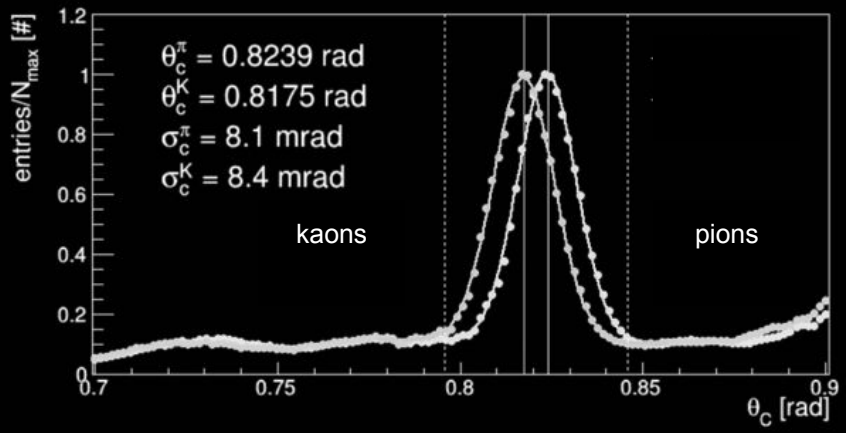
Geometrical Reconstruction

Kaons and pions with p = 4 GeV/c, theta = 11°, phi = 90°

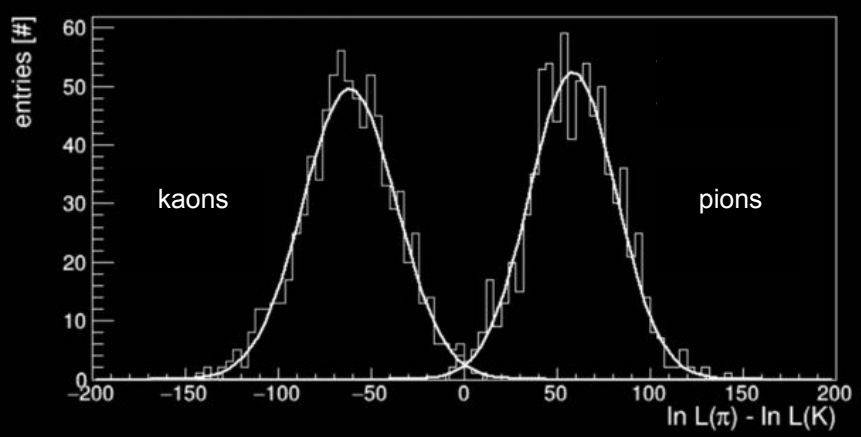
Log likelihood is based on the Cherenkov angle and the number of detected photons

Timing information is used to cut out some solutions

Reconstructed Cherenkov angle



Separation between kaons and pions = 4.95 s. d.



$$\sigma_c^2 = \sigma_{tr}^2 + \frac{\sigma^2}{N}$$

Design goal: >= 3 s.d. between pions and kaons for momenta up to 4 GeV/c

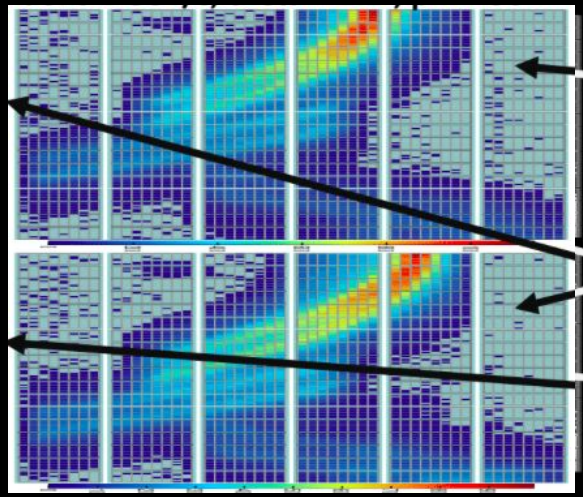
Time-based imaging

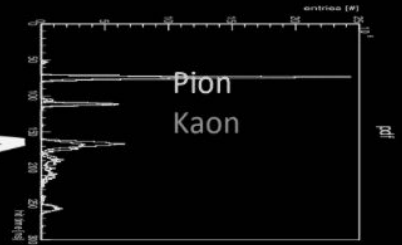
Calculate log likelihoods for each particle hypothesis directly from the time spectrum in each hit pixel

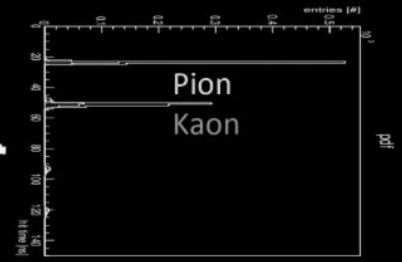
In advance need time spectra for each particle type and the given track configuration (momentum, direction)

Such time spectra can be **simulated or calculated analytically**

kaons 2 GeV/c, θ=1.2º, φ=90º





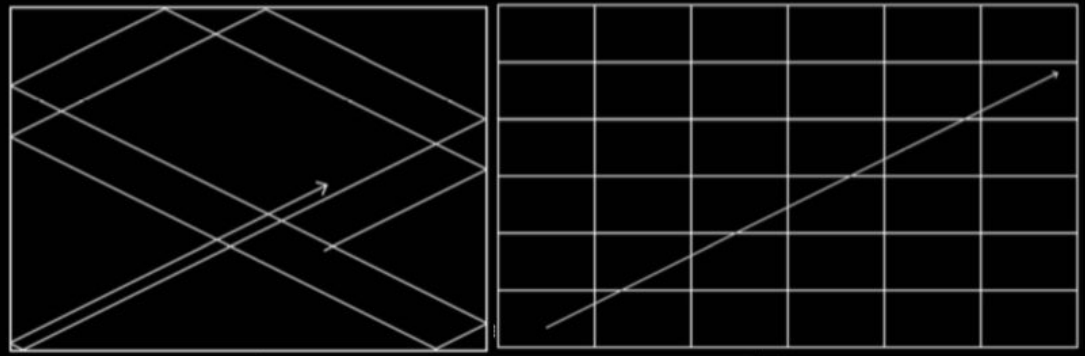


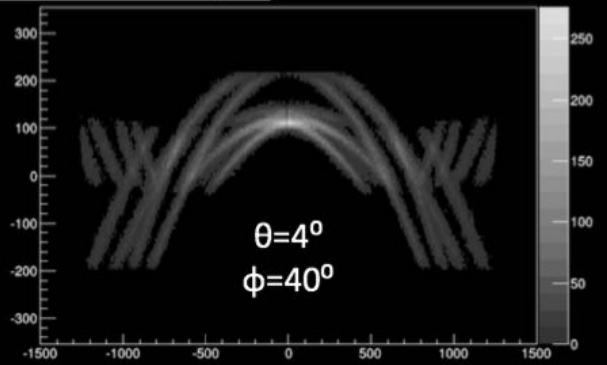
pions 2 GeV/c, θ=1.2º, φ=90º

FastDIRC - KDE based Reconstruction

- Use Kernel Density Estimation to compare particle hypothesis patterns
- Use fast tracing to implement KDE
  - O(1) speed vs O(100) bounces
- Compare Log Likelihoods

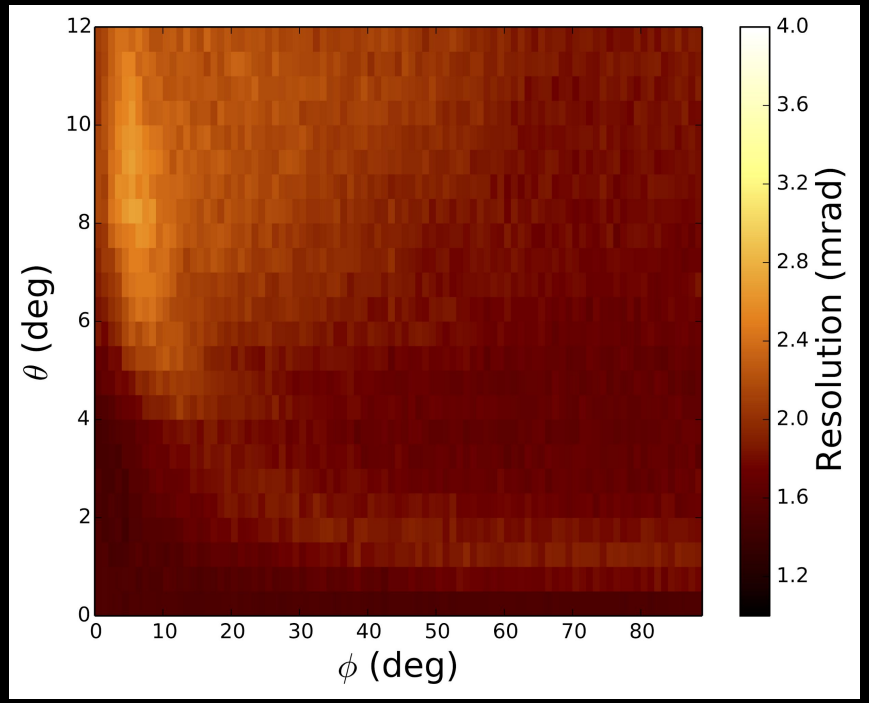
$$P(x) \approx \sum_{i}^{n} K(x - s_i)$$



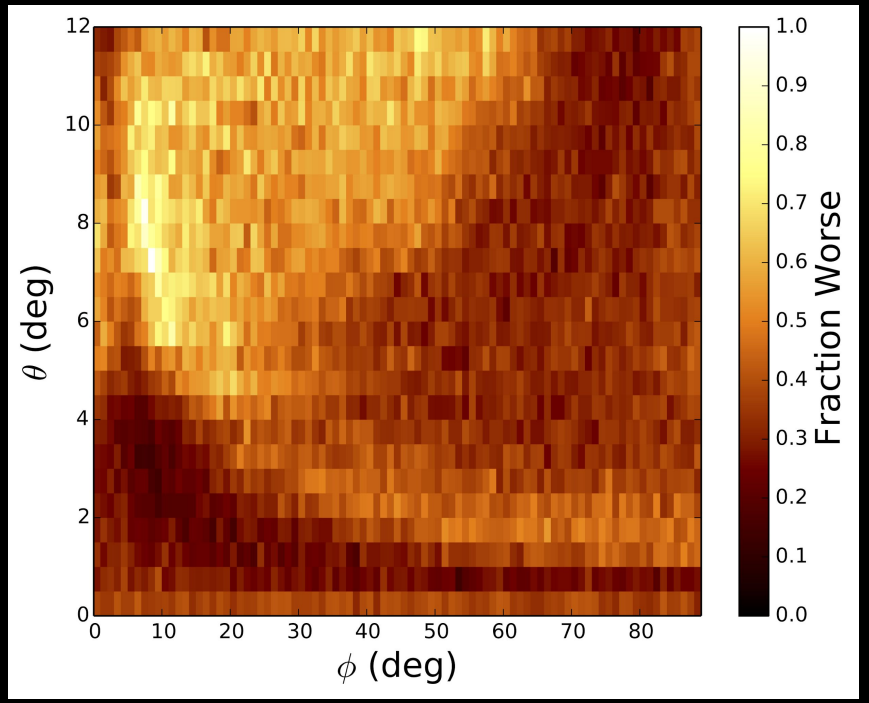


See the FastDIRC paper (J Hardin, M Williams 1608.01180)
Github: https://github.com/jmhardin/FastDIRC

FastDIRC - KDE based Reconstruction

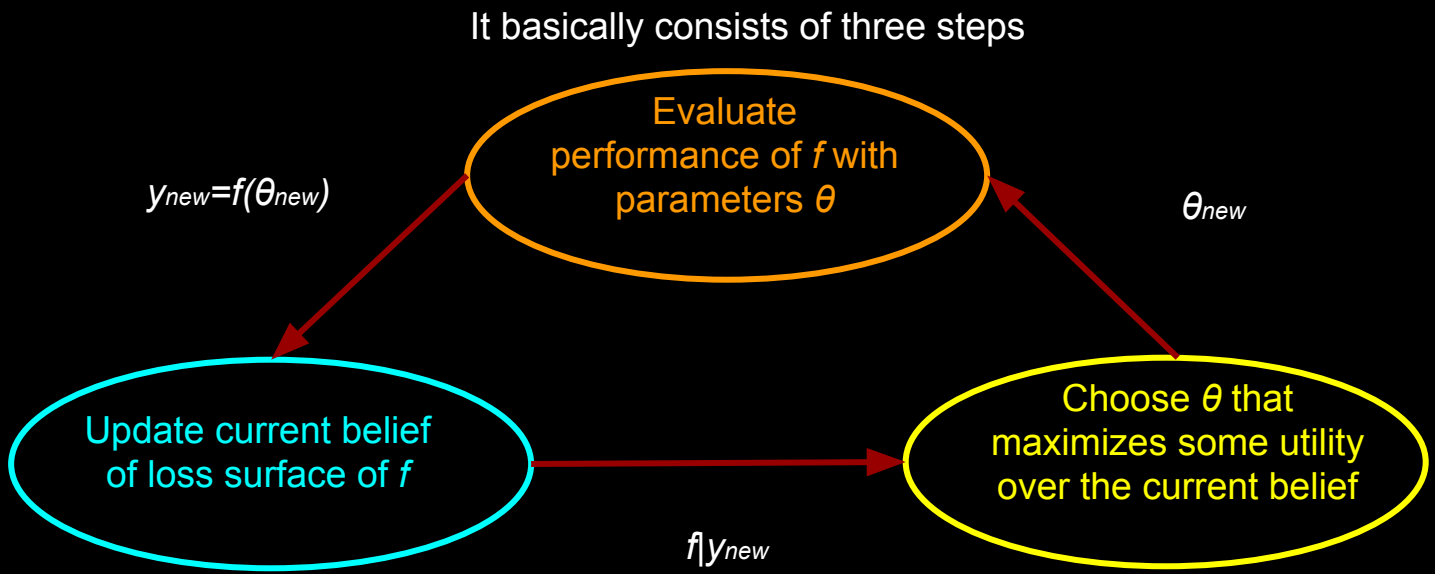


FastDIRC

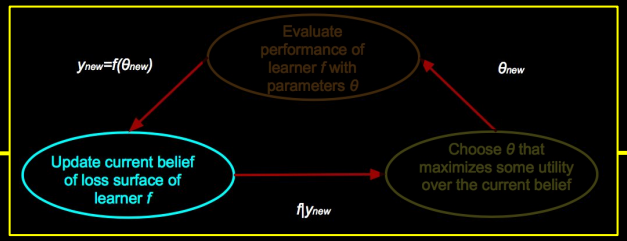


Fraction Worse = (LUTres - FastDIRCres)/LUTres

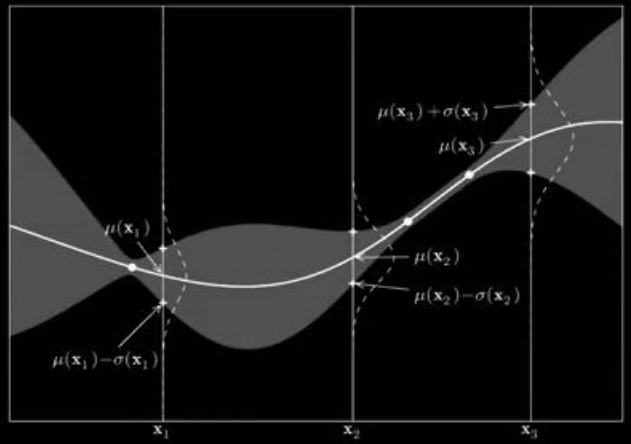
Bayesian Optimization



Update

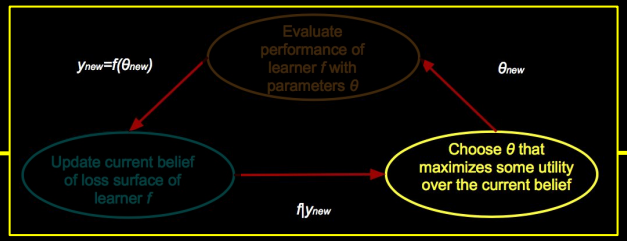


- GPs are the generalization of a Gaussian distribution to a distribution over functions, instead of random variables.
- GP is completely specified by its mean function and covariance function.

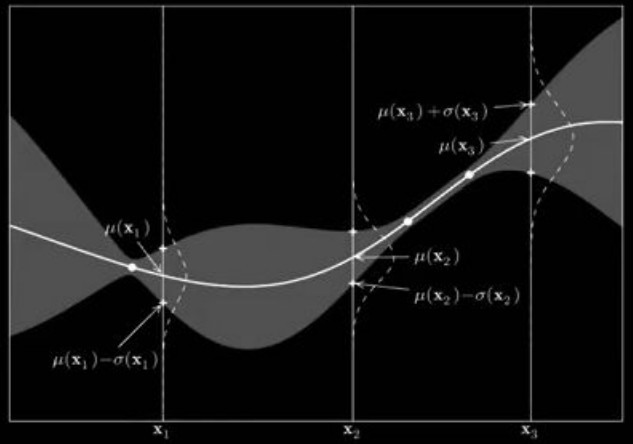


- How should I read this?
  - **Solid line**: function we are trying to min/max
  - S**haded region:** probability model (we know the actual points already evaluated but we are more uncertain in regions where we haven't).
  - In every point a normal distribution of the potential performance function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

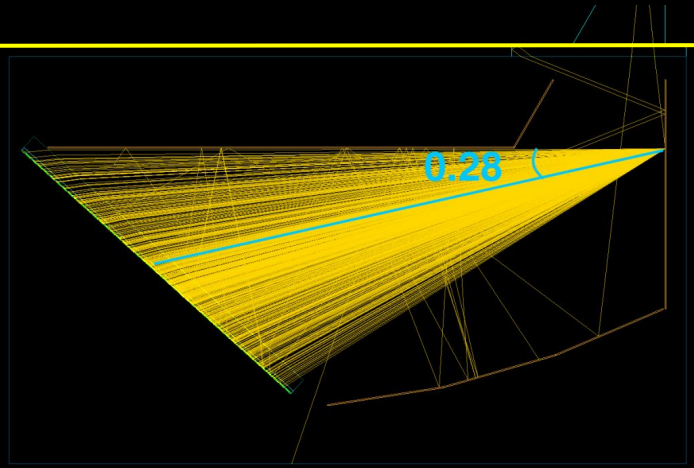
EXPLORATION
Sample a point where uncertainty is high

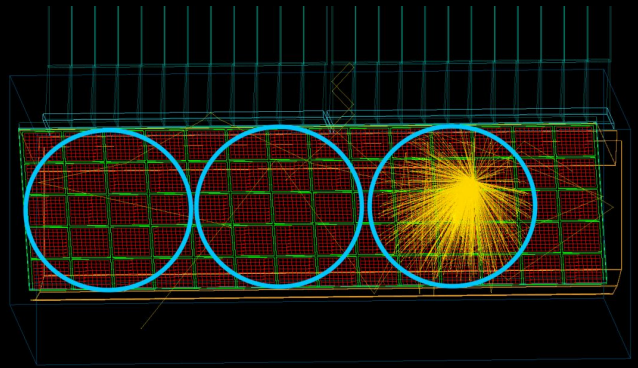
$$EI(\theta) = \begin{cases} \left(\mu(\theta) - f(\hat{\theta})\right)\Phi(Z) + \sigma(\theta)\phi(Z), & \sigma(\theta) > 0 \\ 0, & \sigma(\theta) = 0 \end{cases}$$

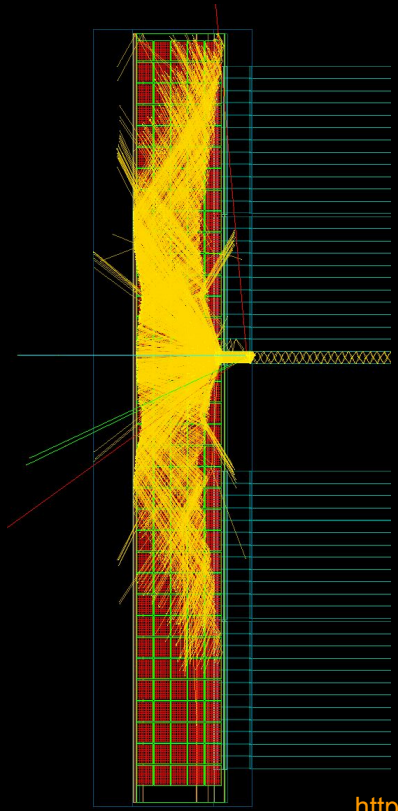
best value we found so far

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

DIRC calibration

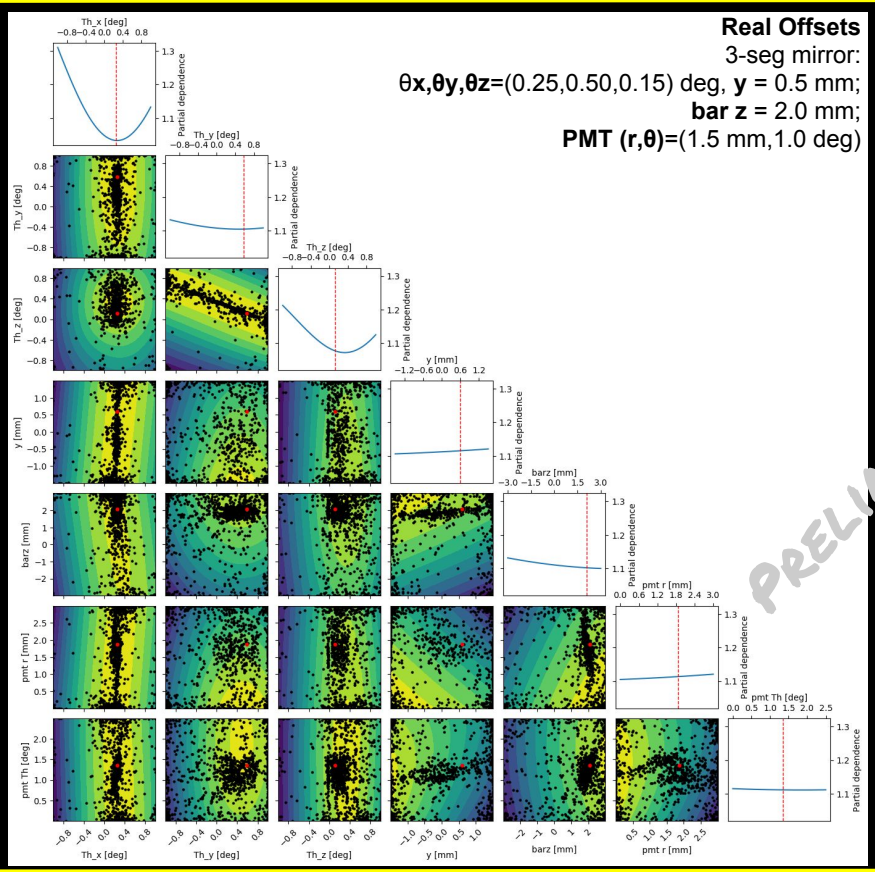


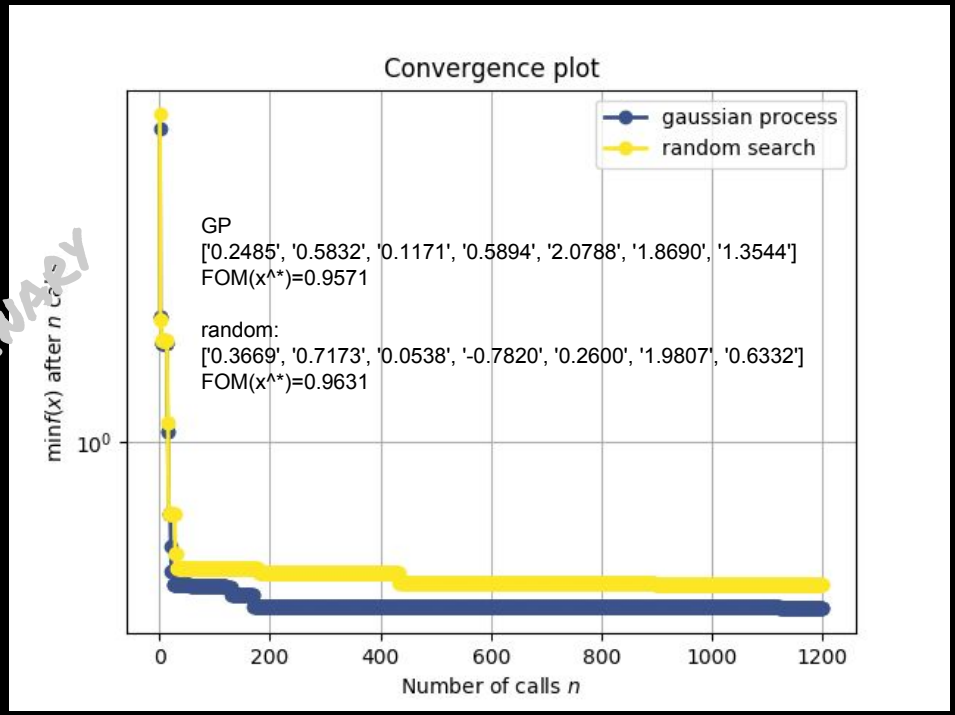




https://gluexdirc.mit.edu/laser

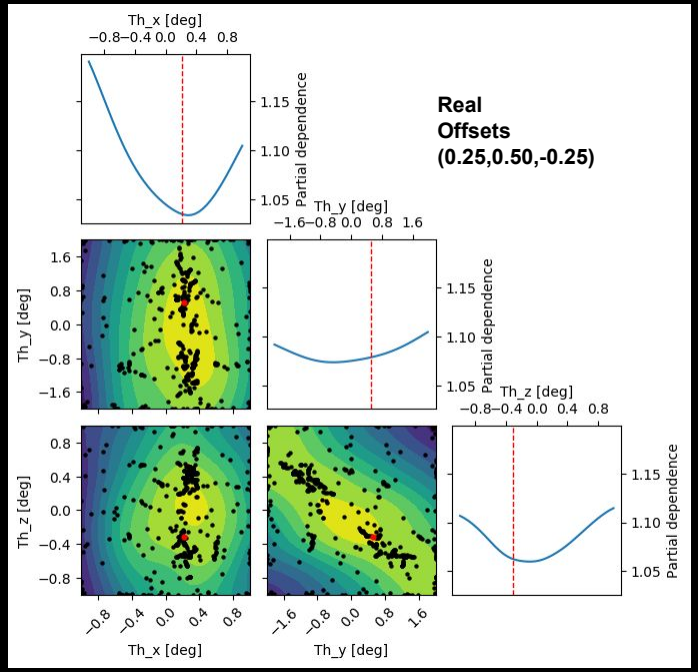
7D preliminary





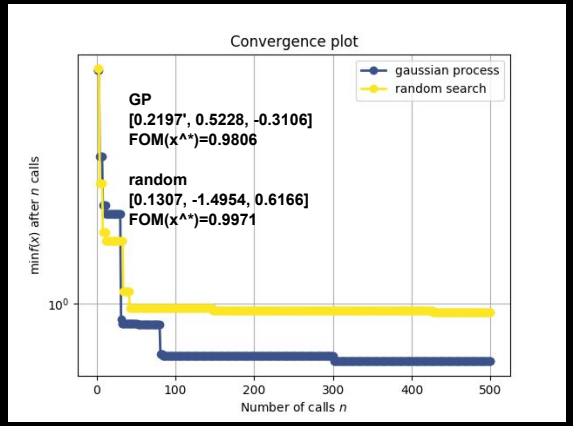
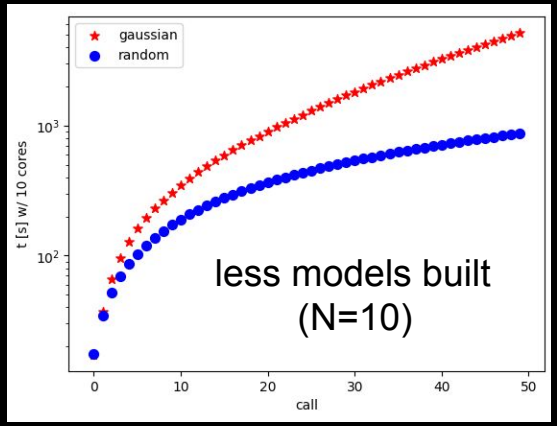
3D combining different particles

Toy-model: sampling 10 particles/call in range
E [GeV], θ [deg], φ [deg]: [2.5,3.0], [2,4], [1,45]



https://github.com/scikit-optimize/scikit-optimize

Recipe: For each call of the optimizer, M offset points are explored using N different particles (for each call). The total number of calls is T

T=50 M=10 N=10
Particles used = 500
Points explored = 500

FoM = LogL normalized to a default alignment