Adaptive Experimentation to assist detector design at EIC

from ECCE to ePIC and beyond

Cristiano Fanelli [1,2], Zisis Papandreou [3], Karthik Suresh [3]

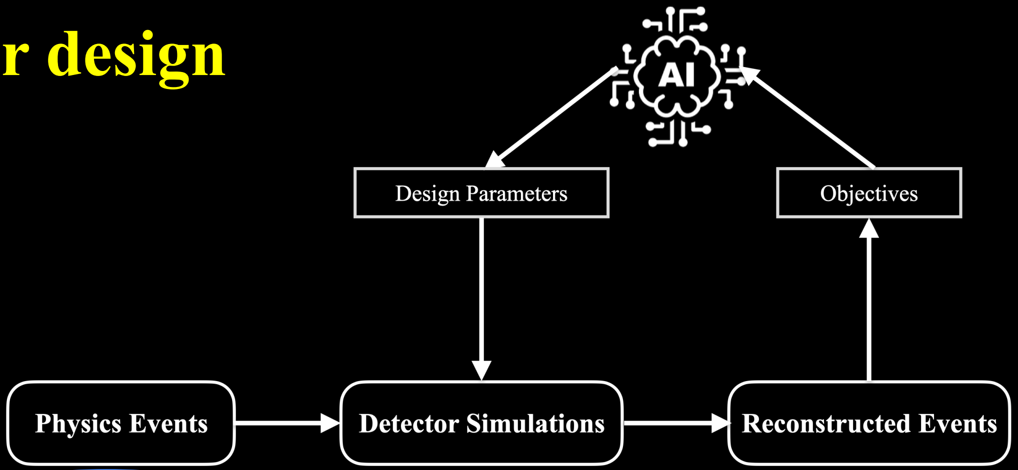
[1] College of William and Mary, VA, USA

[2] Thomas Jefferson National Accelerator Center (TJNAC), VA, USA

[3] University of Regina, SK, Canada

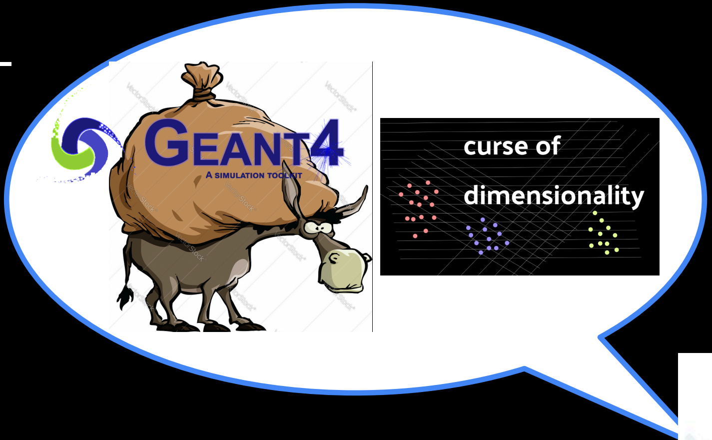
William & Mary Chartered 1693 | Jefferson Lab | University of Regina | Faculty of Science | AI4EIC

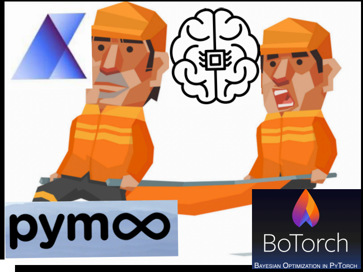
Optimization of EIC Detector design



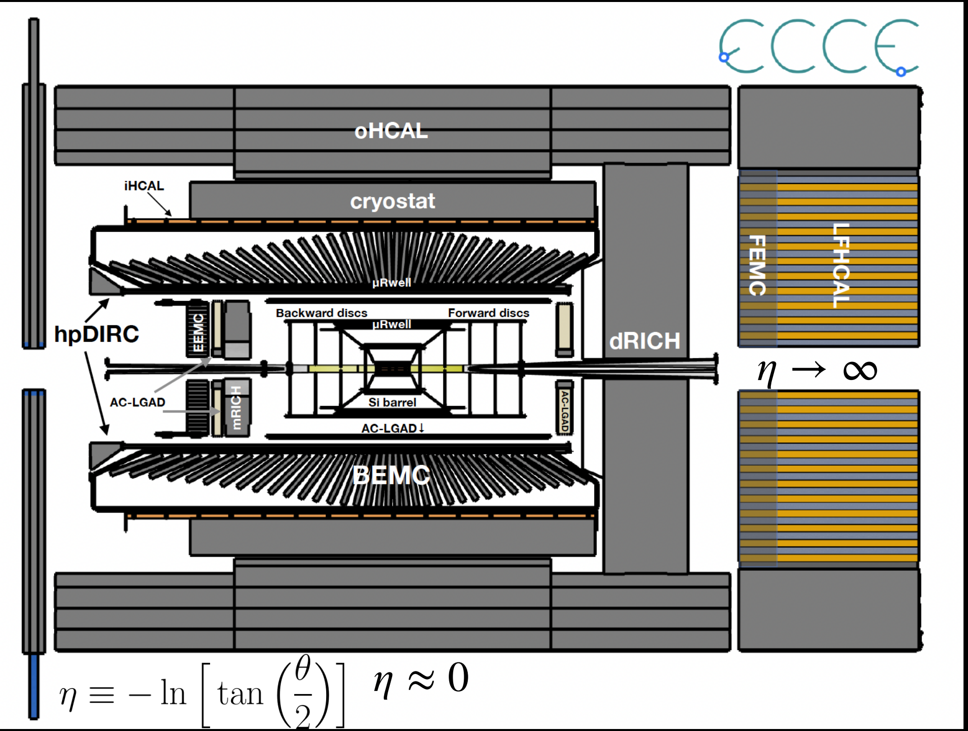
GEANT4 simulations are typically computation intensive.

In order to explore a multi-dimensional parameter space in a multi-dimensional objective space, AI can assist in this search in a more efficient way.

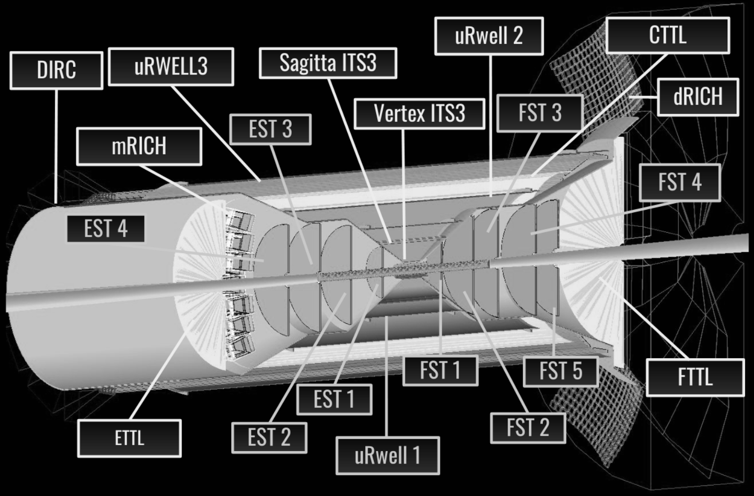




Example: The ECCE Detector - the Tracking System



Tracker System



AI been used to steer the design

arxiv:2205.09185, arxiv:2203.04530

The tracking system reconstructs charged particle tracks. It combines different technologies.

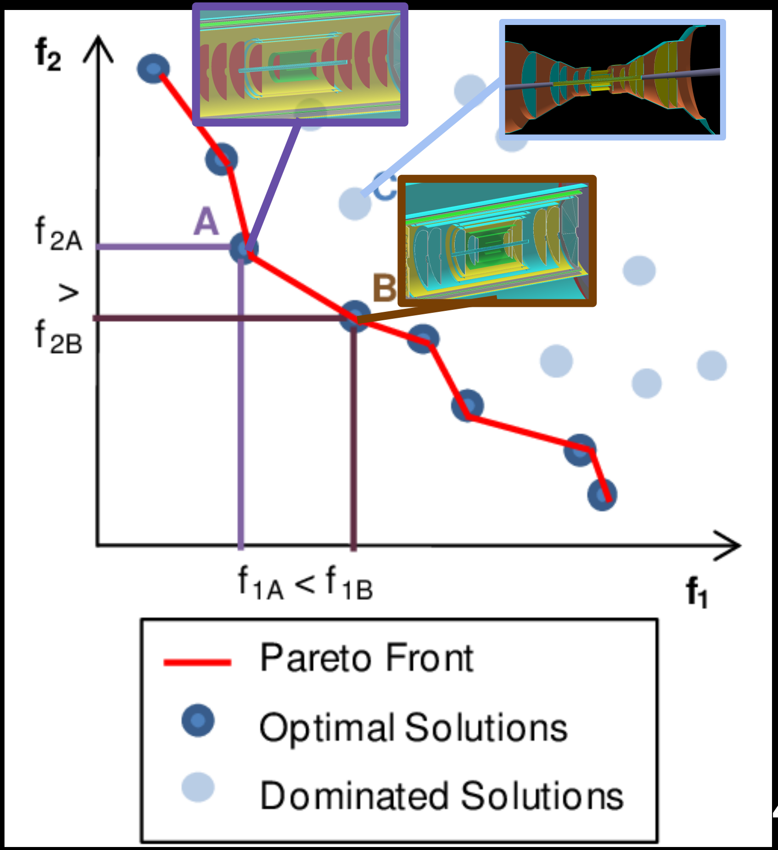
ECCE design was chosen as reference. Lots of updates/progress from Tracking WG (ePIC) since then.

Multi Objective Optimization

$$
\begin{aligned}
&\min \ \mathbf{f_m}(\mathbf{x}) && m = 1, \cdots, M \\
s.t.\ \ &\mathbf{g_j}(\mathbf{x}) \le 0, && j = 1, \cdots, J \\
&\mathbf{h_k}(\mathbf{x}) = 0, && k = 1, \cdots, K \\
&x_i^L \le x_i \le x_i^U, && i = 1, \cdots, N
\end{aligned}
$$

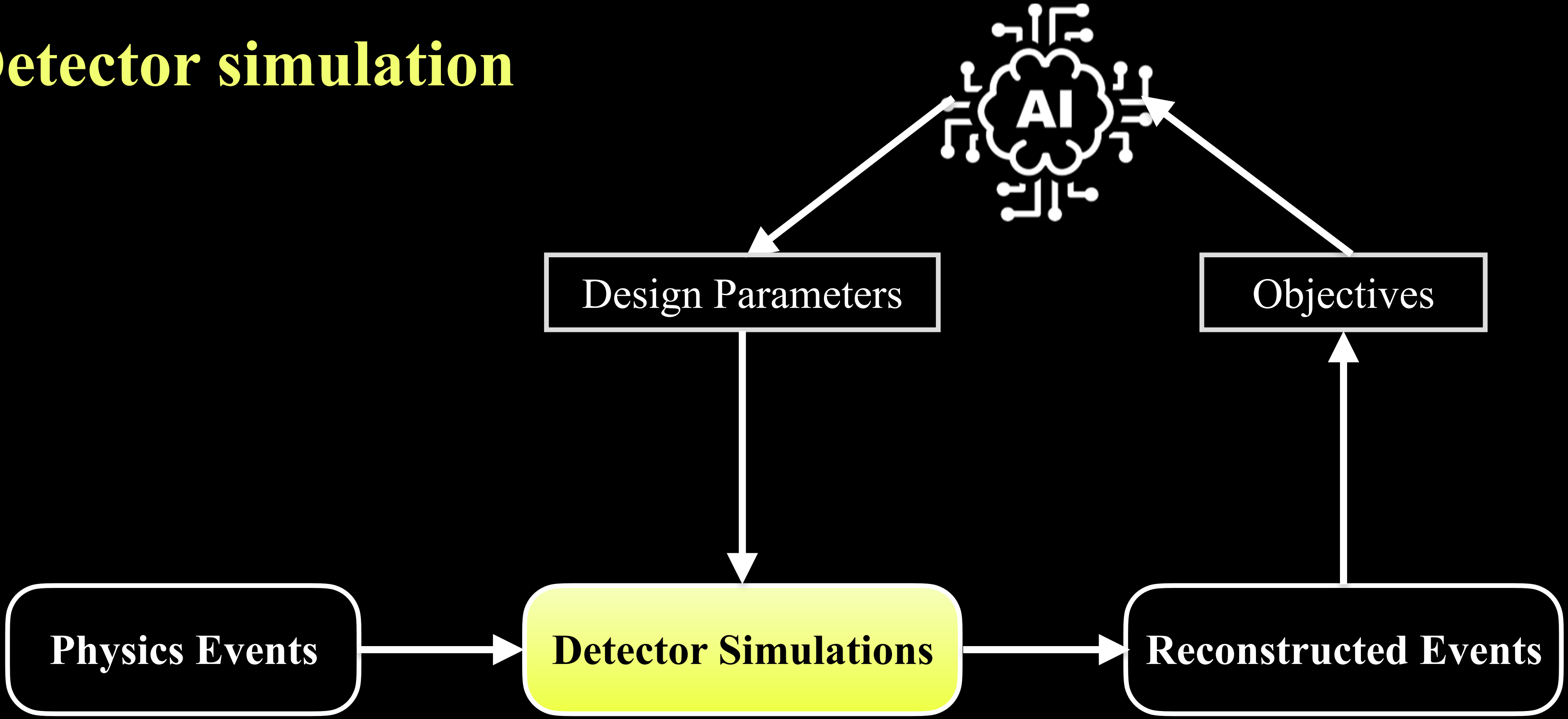
Ref: Max Balandat's talk

- Multiple "objectives", e.g., weighted avg momentum resolution, $\boldsymbol{\theta}$ resolution, KF efficiency, projected $\boldsymbol{\theta}$ resolution at PID location. Objectives could be conflicting. (This can be extended to other objectives, e.g., physics)
- Pareto-optimal solutions. Locus of points in Objective Space which are non-dominating to one another.
- Leverage on SOTA open source tools to solve complex Multi Objective Optimizations problems. Adapt these tools and integrate within GEANT4 based simulations for efficient detector design



BoTorch
Bayesian Optimization in PyTorch
PyGMO
nextorch
NEXTorch
latest
Ax
Multi-objective Optimization in Python
pymoo
GPflow
jMetalPy

Detector simulation

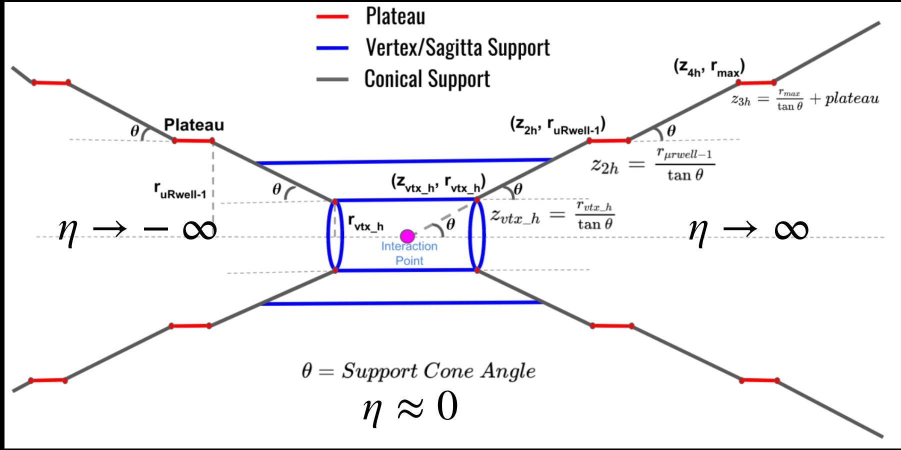


Desired kinematic range

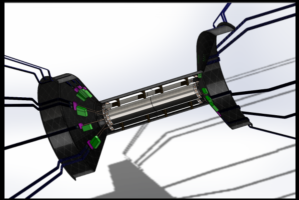
Parametrization

Parametrization is an essential part of an automated optimization:

- explores different designs
- avoids overlaps of volumes
- encodes constraints

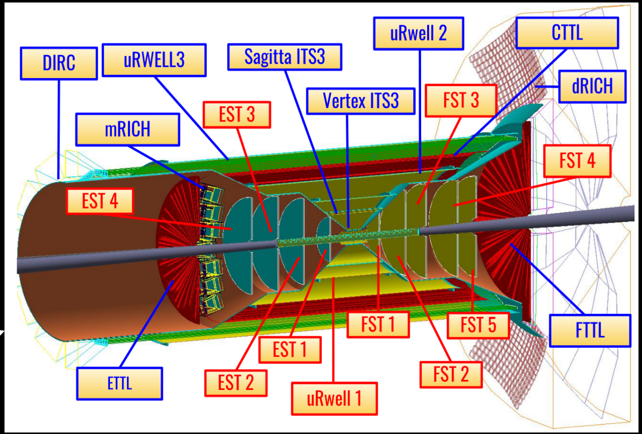


$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$

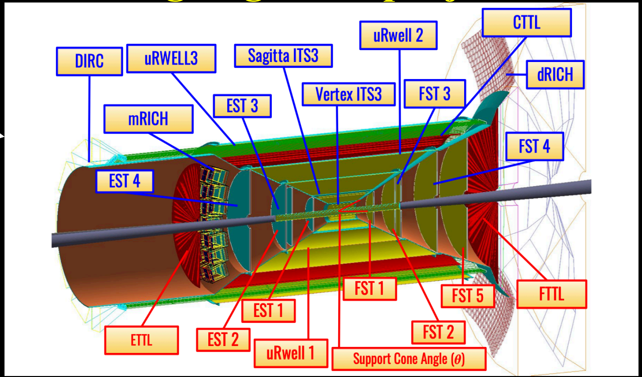


Implementation of support structures with realistic material budgets

Reference design

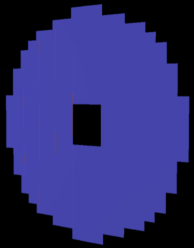


Ongoing R&D projective

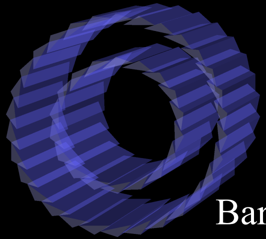


Variable pars; Fixed pars

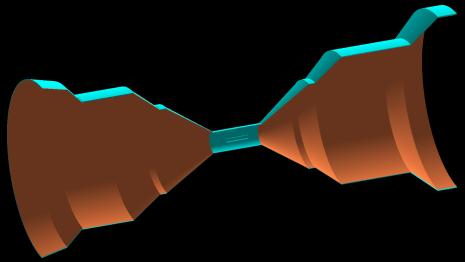
Constraints



FST/EST Disks



Barrel Si Layer



- **Design Parameters** (O(N) ≈ 10)
  - Based on an extensive parameterization.

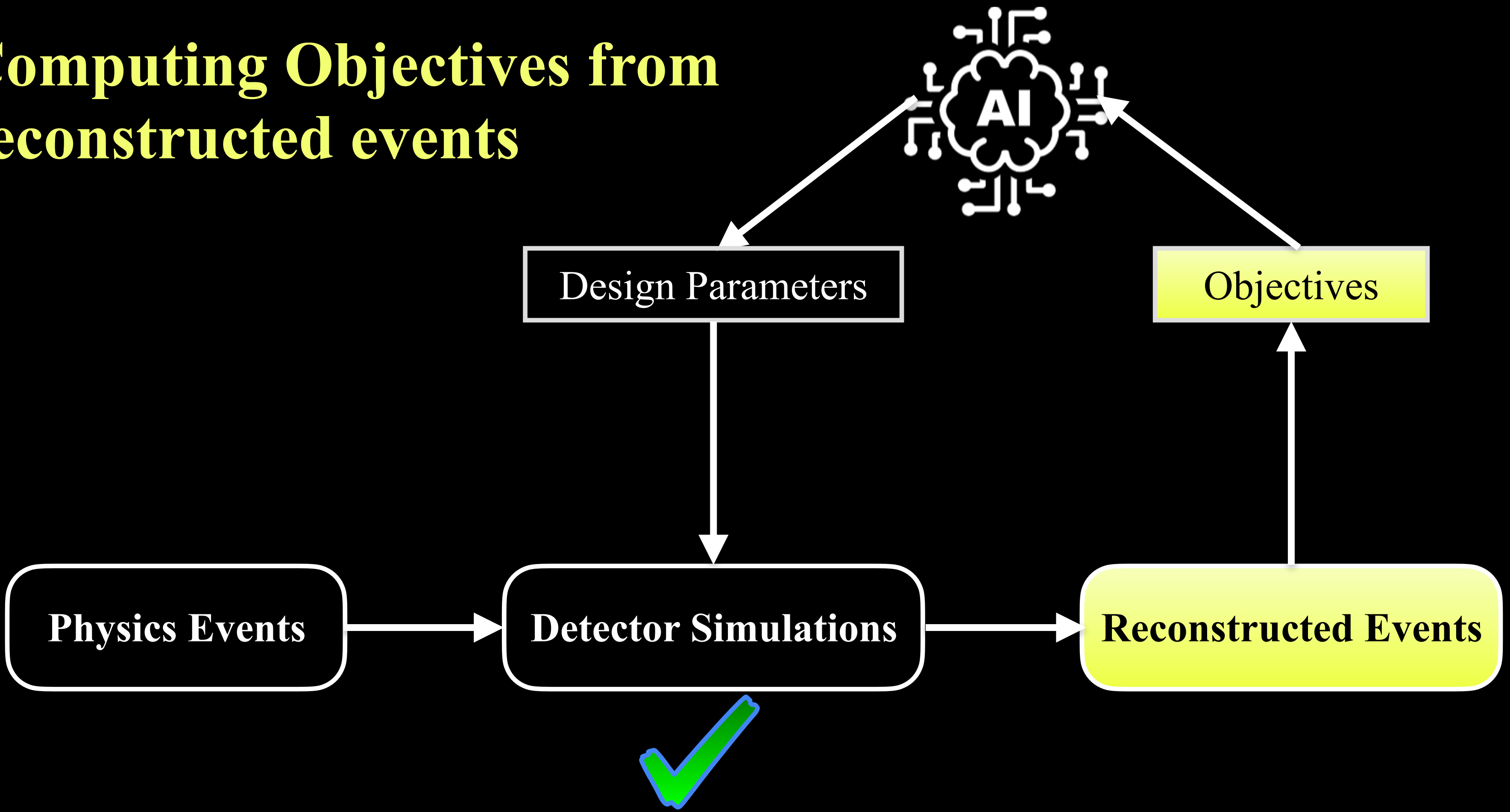
- **Constraints** being used (n_const ≥ 3)
  - HARD The minimum distance between 2 disks should be >= 10 cm (giving room for services)
  - SOFT The Rmax-Rmin for the disks have to be multiple of 3.00 cms and 1.8 cms (Tiling of pixels)

- **Overlaps checks**
  - GEANT4 unstable when overlaps are detected in volumes.
  - Overlaps are checked for every design explored and penalized.

sub-detector	constraint	description
EST/FST disks	$min\left\{\sum_i^{disks}\left\|\frac{R^i_{out}-R^i_{in}}{d}-\left\lfloor\frac{R^i_{out}-R^i_{in}}{d}\right\rfloor\right\|\right\}$	**soft constraint**: sum of residuals in sensor coverage for disks; sensor dimensions: d = 17.8 (30.0) mm
EST/FST disks	$z_{n+1} - z_n >= 10.0$ cm	**strong constraint**: minimum distance between 2 consecutive disks
sagitta layers	$min\left\{\left\|\frac{2\pi r_{sagitta}}{w}-\left\lfloor\frac{2\pi r_{sagitta}}{w}\right\rfloor\right\|\right\}$	**soft constraint**: residual in sensor coverage for every layer; sensor strip width: w = 17.8 mm
μRWELL	$r_{n+1} - r_n >= 5.0$ cm	**strong constraint**: minimum distance between μRwell barrel layers

Extensive details at arXiv:2205.09185

Computing Objectives from reconstructed events



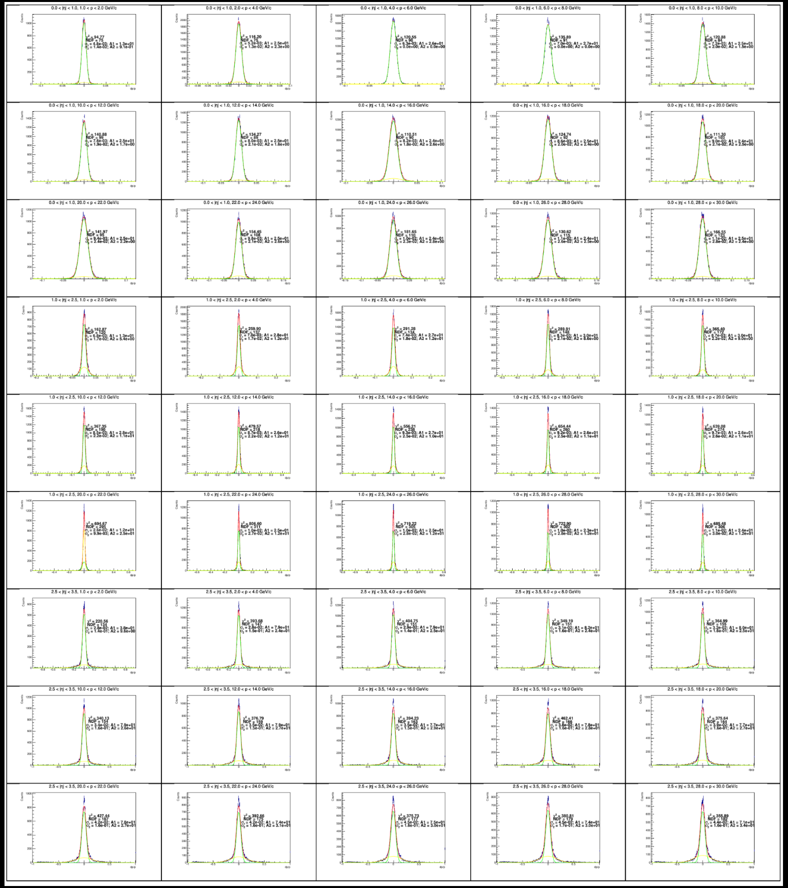
Implementing Objectives

- **Objective functions** Average of Weighted Averages (n_obj $\geq$ 2)
  - **Momentum resolution dp/p**
  - **Theta resolution dθ/θ**
  - **Projected dθ/θ at PID location.**
  - **Kalman Filtering inefficiency** (improving the tracking reconstruction ability of the algorithm)

- **Validation** of the solutions
  - Validate by comparing optimal vs baseline dφ resolution, vertex resolution and reconstruction efficiency

Weighted sum with errors along η

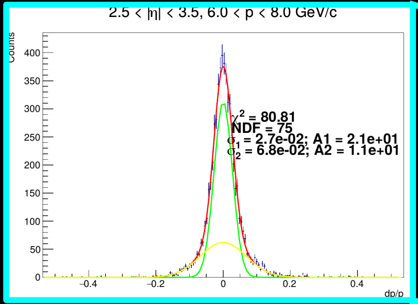
Weighted sum with errors along p



Implementing Objectives

Robust fitting procedure in fine-grained phase-space

Propagate uncertainties in fits throughout



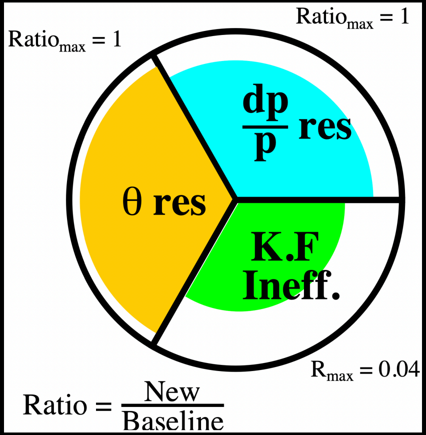
Weighted sum with errors along η

Weighted sum with errors along p

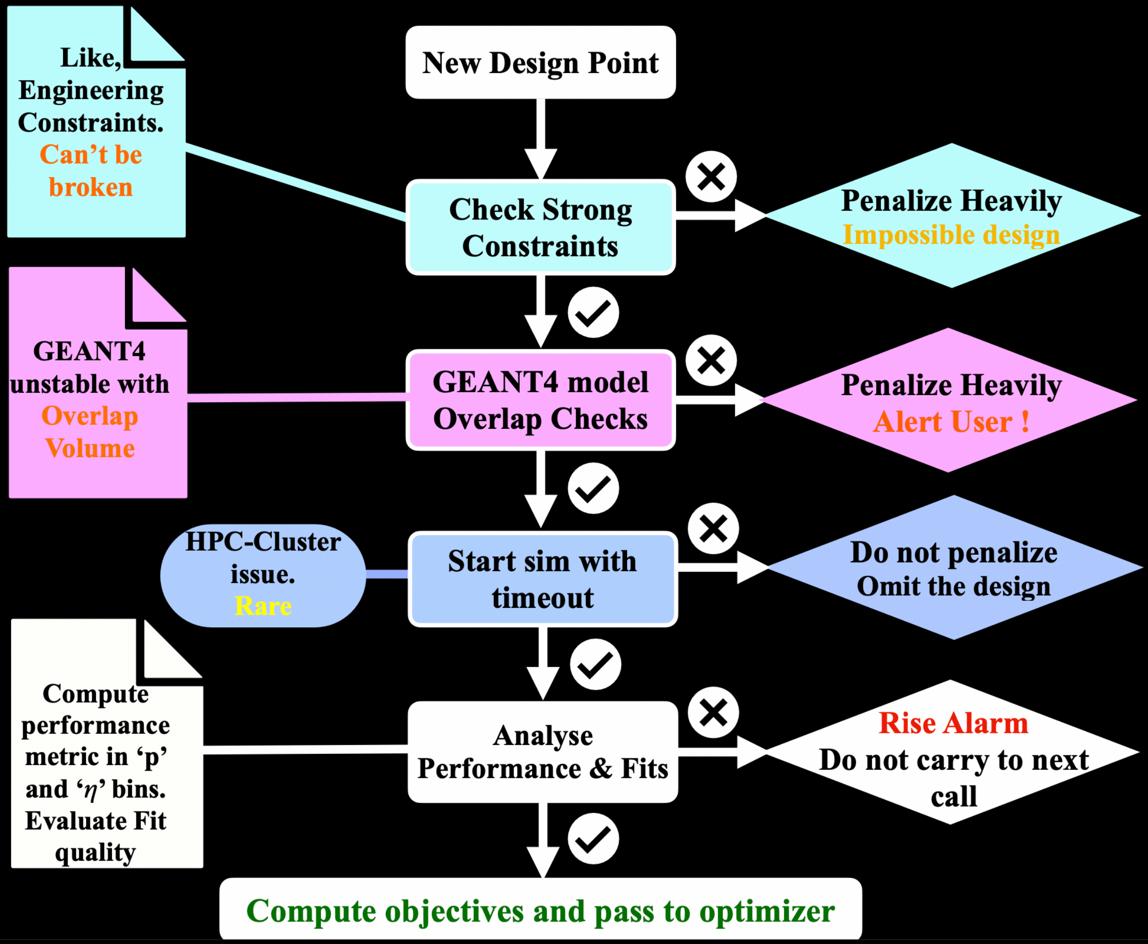
$$\bar{x}_\eta = \frac{\Sigma_p x_p w_p}{\Sigma_p w_p}$$

Avg in a η bin

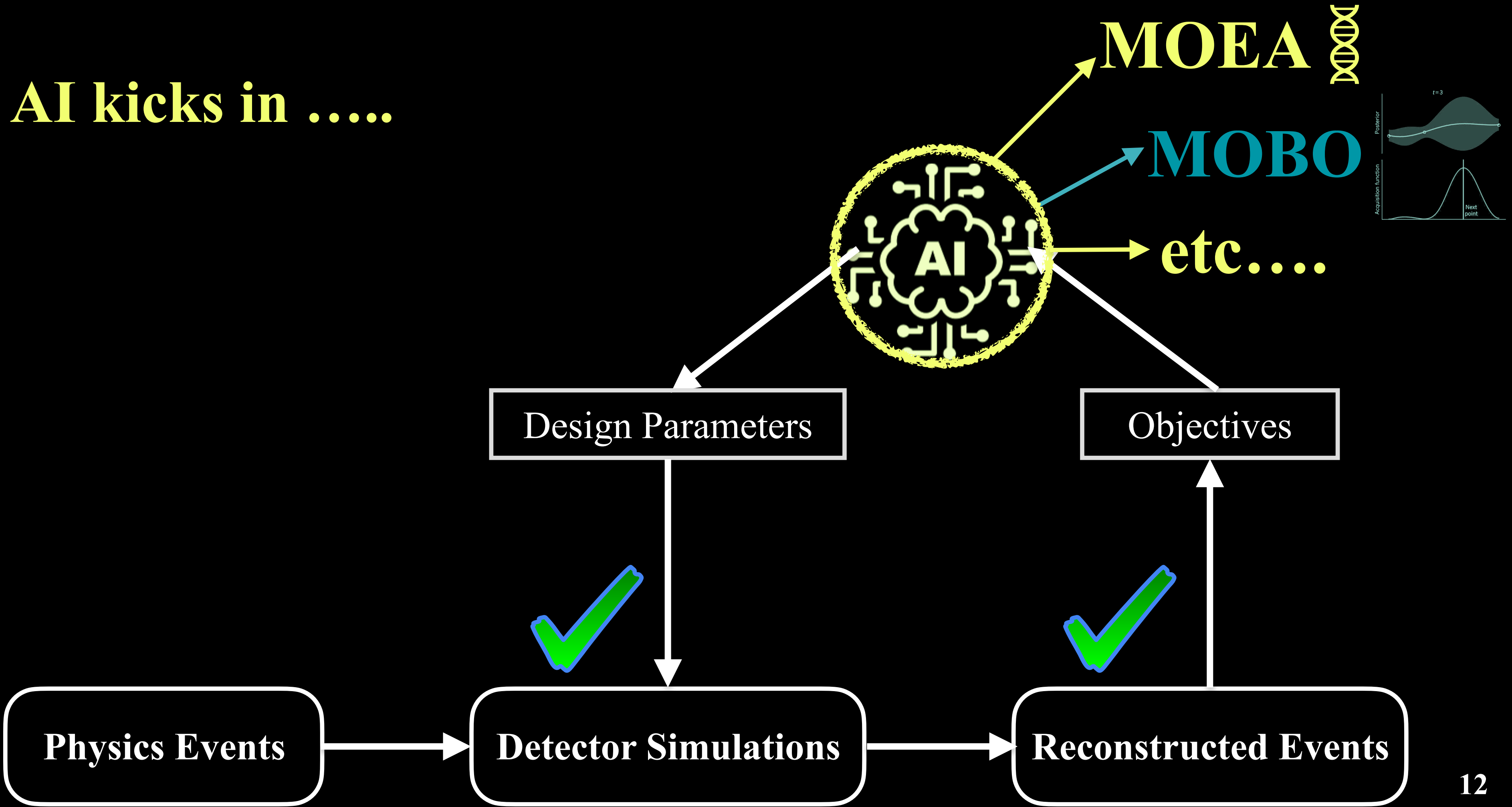
$$\bar{x} = \frac{\Sigma_\eta^{N_\eta} \bar{x}_\eta}{N_\eta}$$



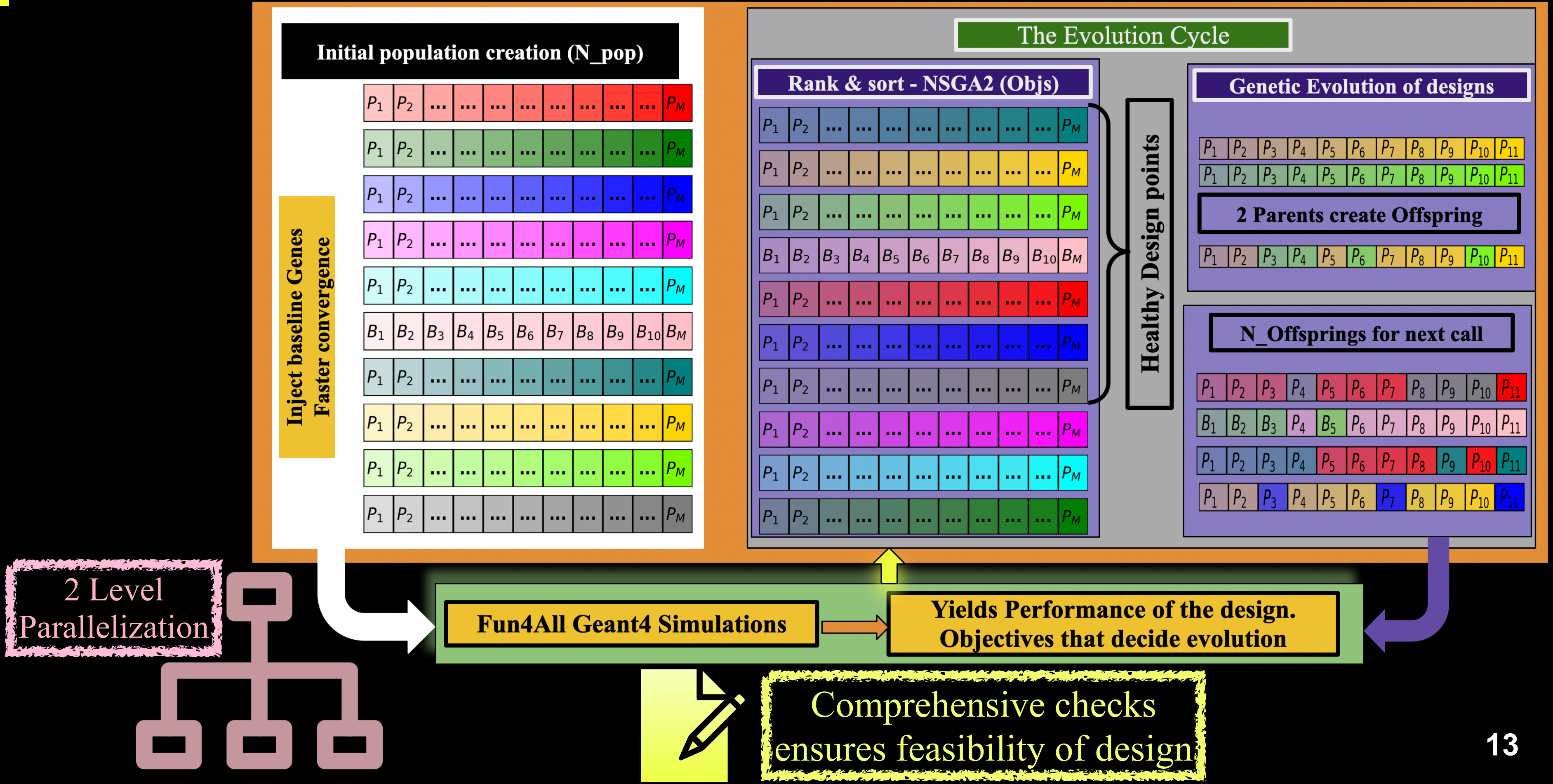
Checks performed



AI kicks in .....


MOEA
MOBO
etc….
AI
Posterior
Acquisition function
Next point
Design Parameters
Objectives
Physics Events
Detector Simulations
Reconstructed Events

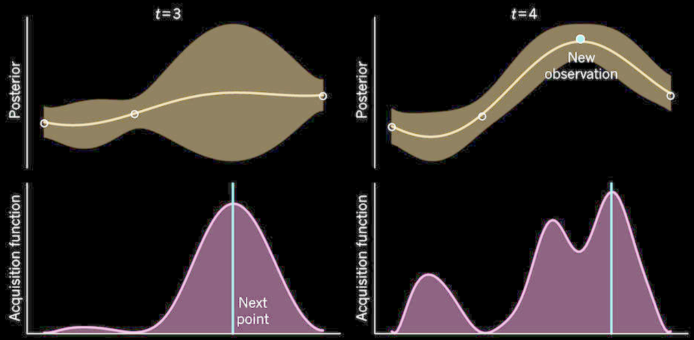
MOEA Pipeline

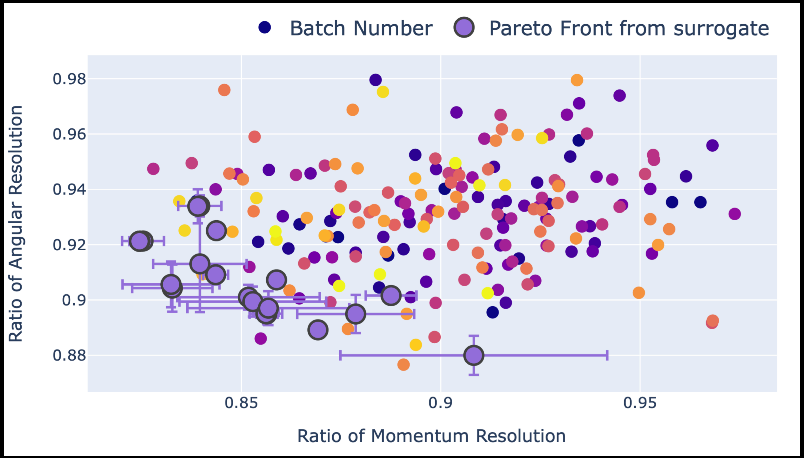


MOBO Pipeline

- **Ax - BoTorch** (Ref: Max Balandat's talk)
  - Apt when evaluations of objectives are costly. Typical for our case.
  - Builds surrogate models that maps objective space to design parameter space.
  - Uses novel qNEHVI acq. function with reduced computational complexity arxiv:2105.08195.

- **Implementation**
  - 1 Level Parallelization (≈ 120cores)
  - BATCH_SIZE - 3 (q)
  - N_BATCH - 50
  - qNEHVI + SAASBO

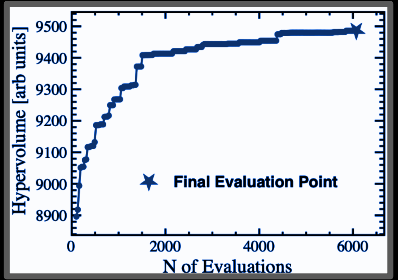




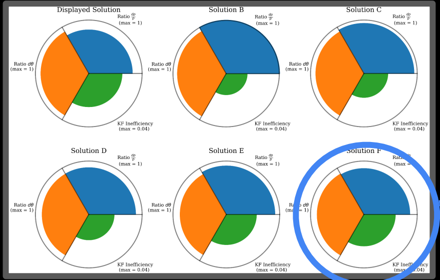
Interactive Visualization of the result

Analyzing results

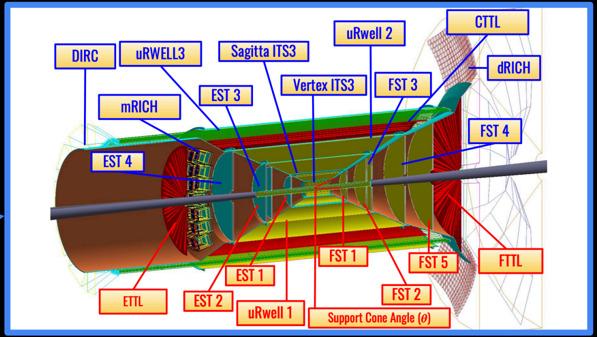
1 Can take a snapshot any time during evaluation



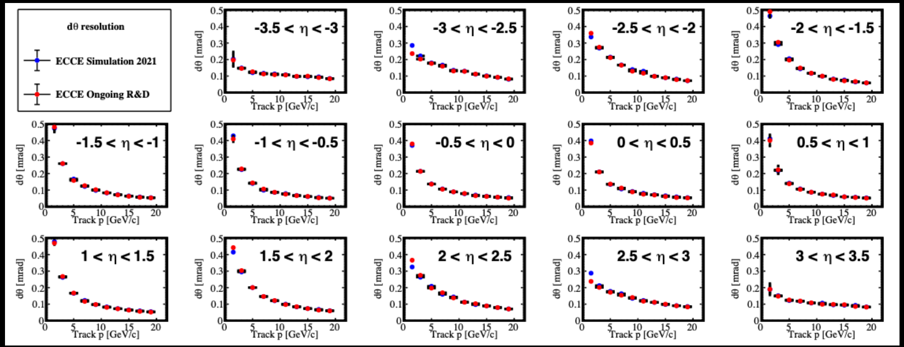
2 Updated Pareto Front at time t

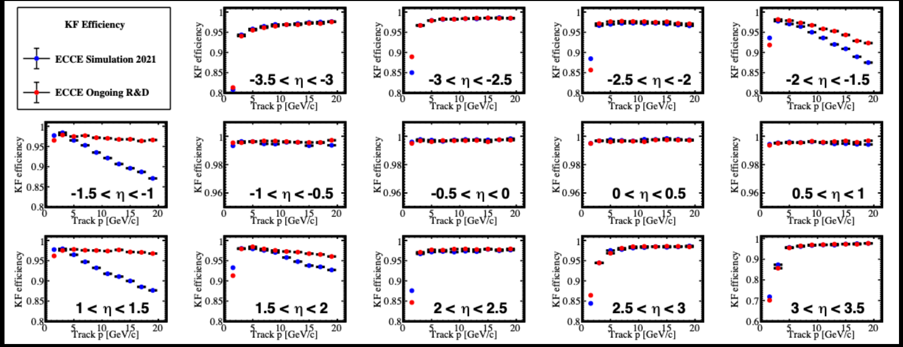


3 Each point is a design

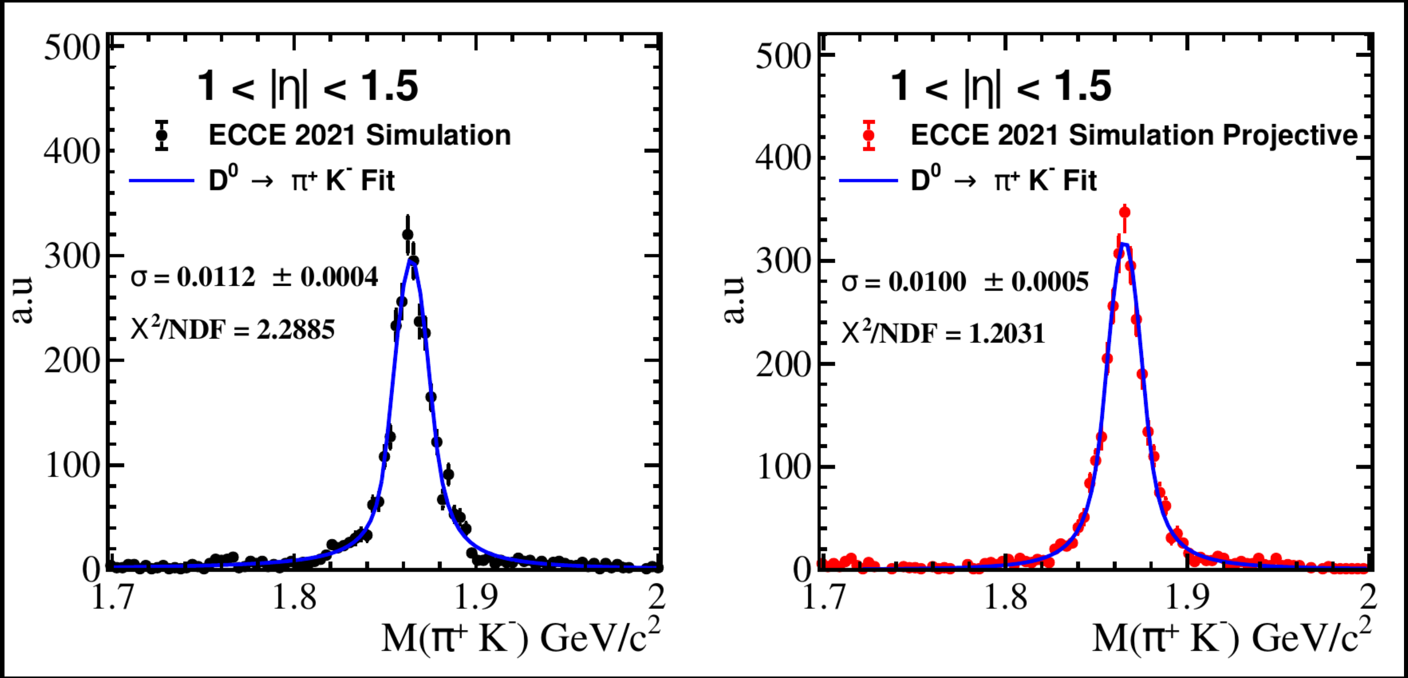


4 Analysis of Objectives (momentum resolution, angular resolution, KF Efficiency)



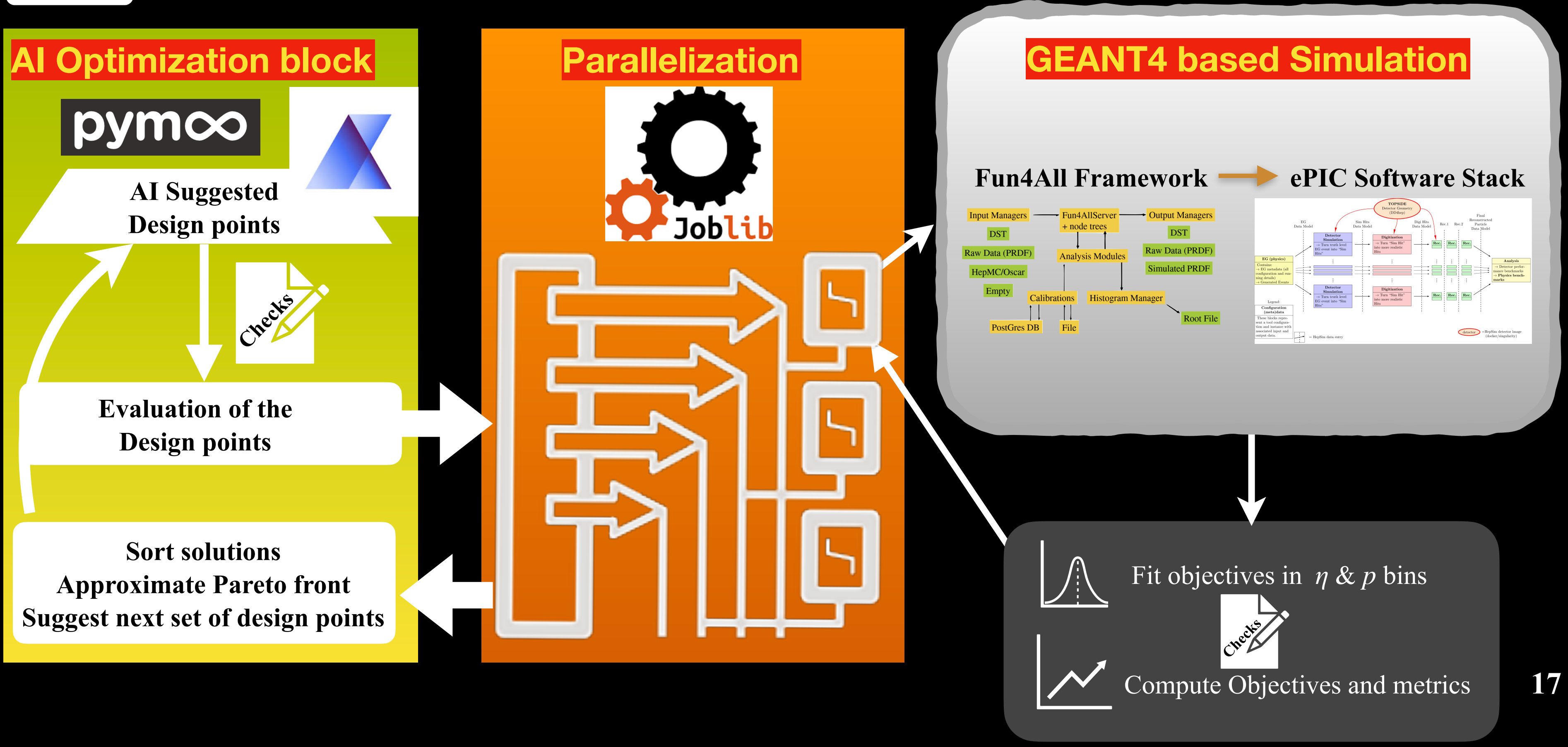


Post-hoc validation on physics observables



The π^+K^- invariant mass obtained from the SIDIS events with updated baseline and optimized projective geometry. A region of eta that is sensitive due to considerable materials for support structure was also taken in to account for this optimization.

AI assisted EIC Detector optimization pipeline



AI assisted EIC Detector optimization pipeline

SINGULARITYCE

AI Optimiza

pymo

AI Sug
Design

Evaluatio
Design

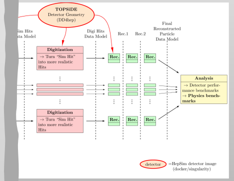
Sort so
Approximate
Suggest next set

ePIC Software Stack

- Geometry implementation via data source (DD4Hep uses ROOT TGeo) makes transparent the coupling of AI to the software stack design parameters; minimal changes needed to run different optimization pipelines.
- Modularity of geometry description reduces complexity of parametrization and therefore computational complexity.
- Effective CI/CD implementation. Need when relevant updates are made, eg. to simulation, or newer aspects to be included in optimization.
- Support for inherent parallelization and heterogenous computation.
- Ease of coupling AI/ML libraries in to the software framework.
- More at [EIC Software Infrastructure Review]

ulation

Software Stack



p bins

es and metrics

Summary

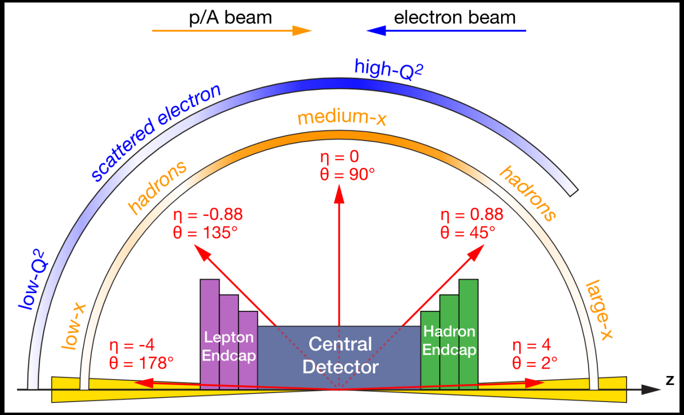
- AI can assist the design and R&D of complex experimental systems by providing more efficient design (considering multiple objectives) utilizing effectively the computing resources needed to achieve that.
- EIC can be one of the first experiments to be designed with the support of AI.
- Optimization could be done in phases. Eg. include one detector system at a time arxiv:2205.09185
- May not have to reinvent the wheel, leverage on existing SOTA tools,
  - Co-develop tools to better adapt and serve our community (EIC Software: Statement of principles)
- Ongoing work:
  - In the process of migrating to the new ePIC software stack*
  - Looking into tracker + PID optimization. e.g, detectors like the dRICH (JINST 15 P05009)
  - More realistic effects in the simulation and reconstruction techniques (effort from ePIC detector WG)

* Beta tests successful.

Thank You

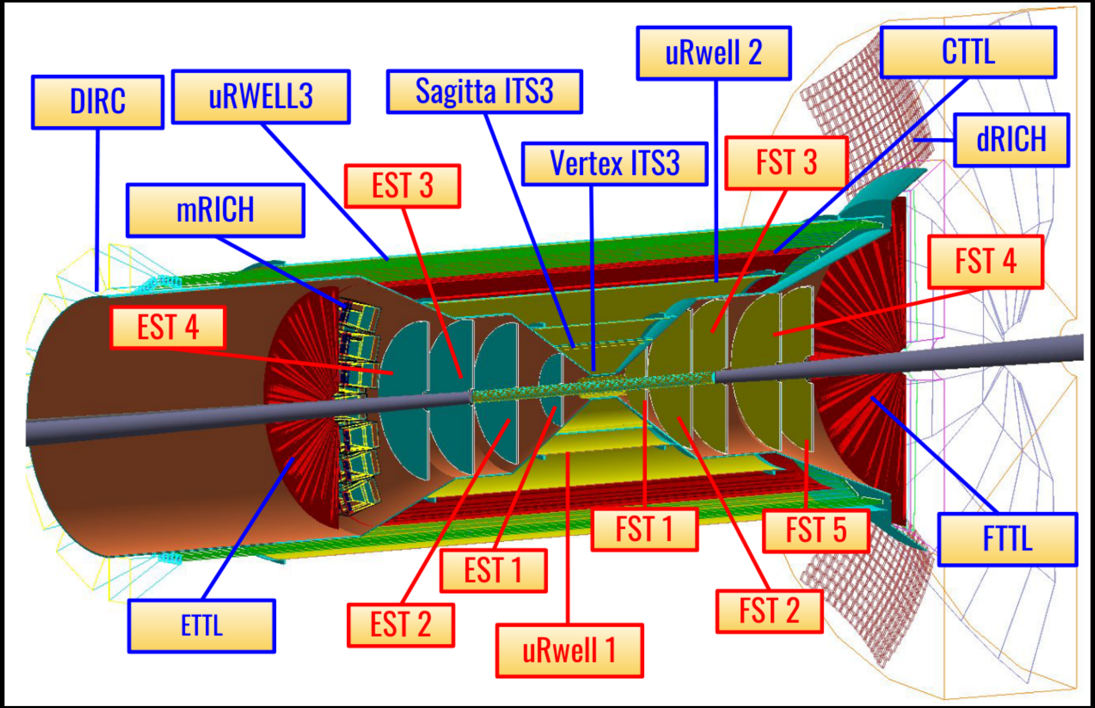
Spares

The ECCE Tracking System

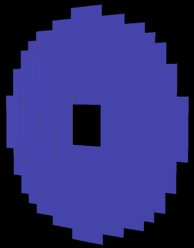


Pseudorapidity

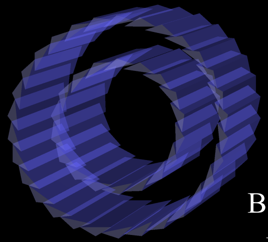
$$\eta \equiv -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$$



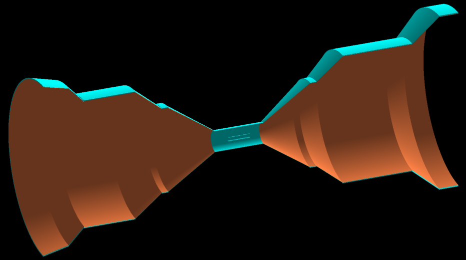
Detector parameters and Constraints



FST/EST Disks



Barrel Si Layer

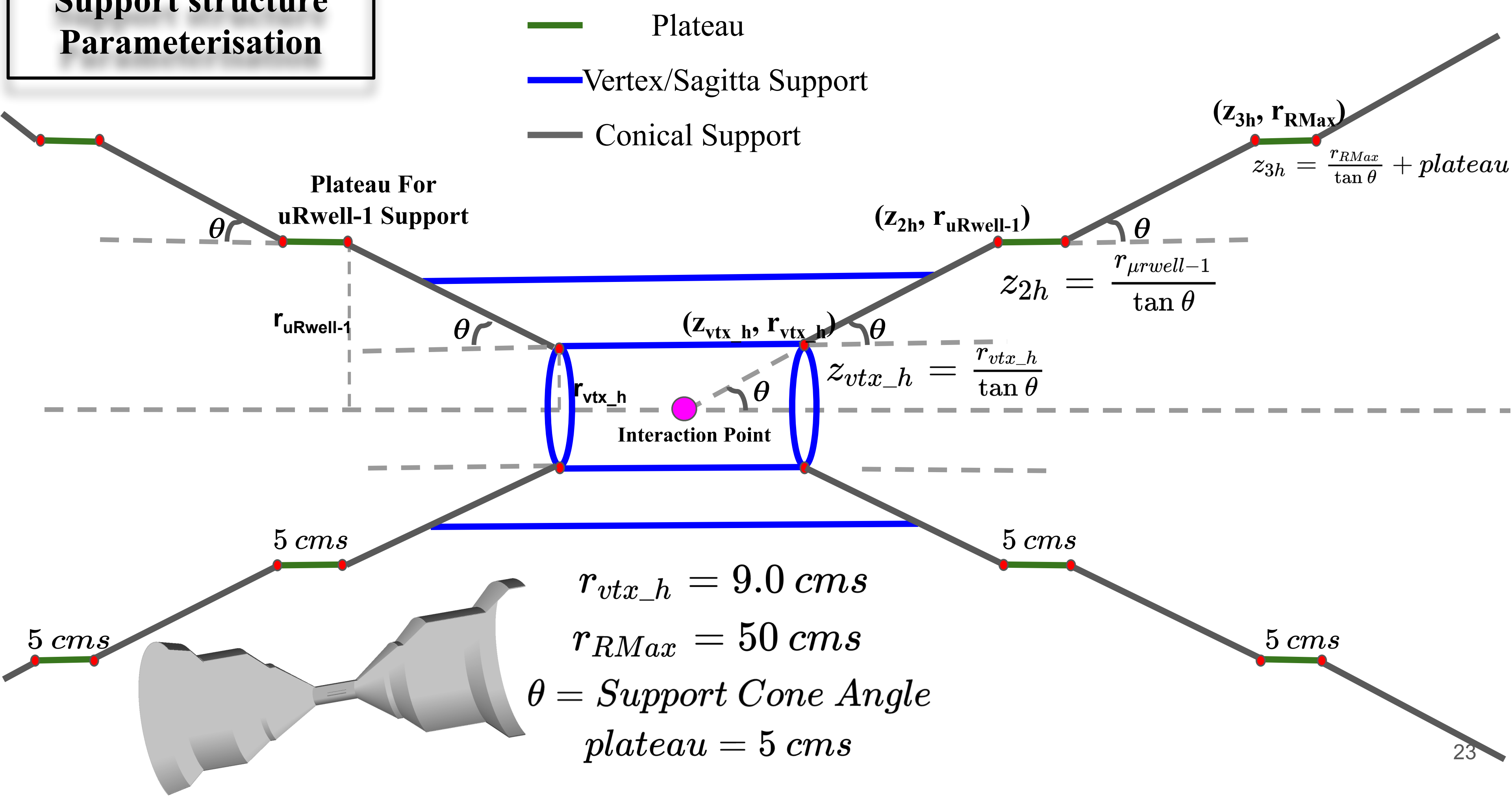


ECCE design (non-projective)	
Design Parameter	**Range**
μRWELL 1 (Inner) (r) Radius	[17.0, 51.0 cm]
μRWELL 2 (Inner) (r) Radius	[18.0, 51.0 cm]
EST 4 z position	[-110.0, -50.0 cm]
EST 3 z position	[-110.0, -40.0 cm]
EST 2 z position	[-80.0, -30.0 cm]
EST 1 z position	[-50.0, -20.0 cm]
FST 1 z position	[20.0, 50.0 cm]
FST 2 z position	[30.0, 80.0 cm]
FST 3 z position	[40.0, 110.0 cm]
FST 4 z position	[50.0, 125.0 cm]
FST 5 z position	[60.0, 125.0 cm]
ECCE ongoing R&D (projective)	
Design Parameter	**Range**
Angle (Support Cone)	[25.0°, 30.0°]
μRWELL 1 (Inner) Radius	[25.0, 45.0 cm]
ETTL z position	[-171.0, -161.0 cm]
EST 2 z position	[45, 100 cm]
EST 1 z position	[35, 50 cm]
FST 1 z position	[35, 50 cm]
FST 2 z position	[45, 100 cm]
FST 5 z position	[100, 150 cm]
FTTL z postion	[156, 183 cm]

sub-detector	constraint	description
EST/FST disks	$min\left\{\sum_{i}^{disks}\left\lvert\frac{R_{out}^{i}-R_{in}^{i}}{d}-\left\lfloor\frac{R_{out}^{i}-R_{in}^{i}}{d}\right\rfloor\right\rvert\right\}$	**soft constraint**: sum of residuals in sensor coverage for disks; sensor dimensions: d = 17.8 (30.0) mm
EST/FST disks	$z_{n+1} - z_n >= 10.0$ cm	**strong constraint**: minimum distance between 2 consecutive disks
sagitta layers	$min\left\{\left\lvert\frac{2\pi r_{sagitta}}{w}-\left\lfloor\frac{2\pi r_{sagitta}}{w}\right\rfloor\right\rvert\right\}$	**soft constraint**: residual in sensor coverage for every layer; sensor strip width: w = 17.8 mm
μRWELL	$r_{n+1} - r_n >= 5.0$ cm	**strong constraint**: minimum distance between μRwell barrel layers

Extensive details at arXiv:2205.09185

Support structure Parameterisation



$$r_{vtx\_h} = 9.0\ cms$$

$$r_{RMax} = 50\ cms$$

$$\theta = Support\ Cone\ Angle$$

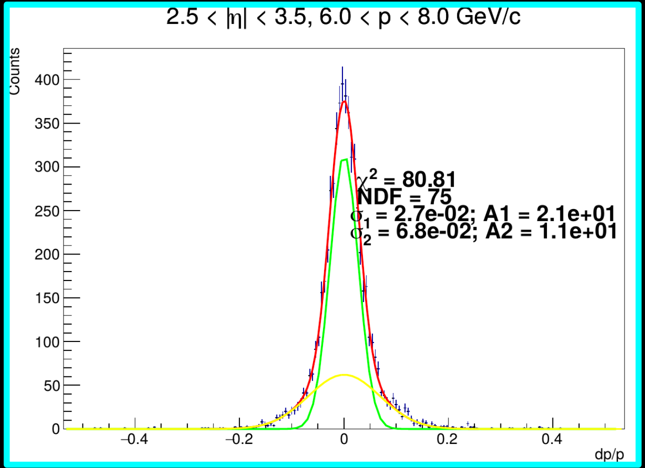
$$plateau = 5\ cms$$

Fitting Procedure

- For resolutions
  - Plot distributions of resolution in bins of eta and p
  - Fit with a double gaussian function
  - Set $A_{1 \text{ or } 2}$ (Amplitude) to 0 if the fit value of A is less than 1% of the $A_{2 \text{ or } 1}$
  - Set $\sigma_{1 \text{ or } 2}$ to 0 if it is greater than the x axis extent of the histogram
  - Calculate the weighted sigma of the fit function and its associated errors.
- For Global KF Inefficiency
  - Calculate the total number of tracks with trackID<0 for the entire simulation
  - Global_KF_Inefficiency = No_of_tracks(trackID<0)/ Total_Events

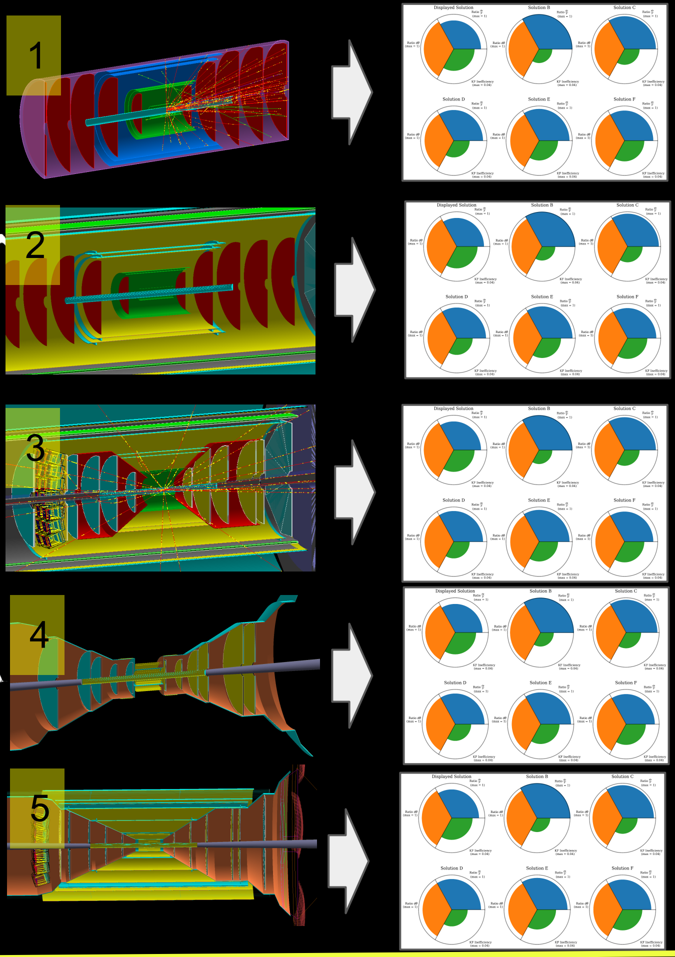
$$D_1 e^{(x-\mu)^2/\sigma_1^2)} + D_2 e^{(x-\mu)^2/\sigma_2{}^2)}$$

$$\sigma = \frac{\sigma_1 A_1 + \sigma_2 A_2}{A_1 + A_2}$$



Phases of Optimisation

Phases of Optimisation



Tracker Optimisation timeline.

- 1: Barrel + technological choices.
- 2: Barrel+Disks. Without any support structures.
- 3. Barrel+Disks. With **fixed** support structures.
- 4. Barrel+Disks and support structure.
- 5. Full tracking system optimisation.

DWG's:
- Technology Selection
- Baseline Design
- Alternate Configuration(s)

PWG's:
- Physics Signal Selection
- Physics Performance Evaluation

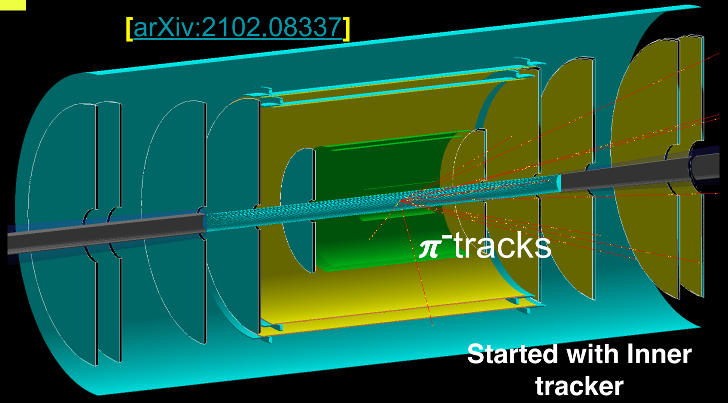
CWG's:
- Simulation Campaign Support
- AI Optimization

updated configurations with any additional requirements

Optimisation phases

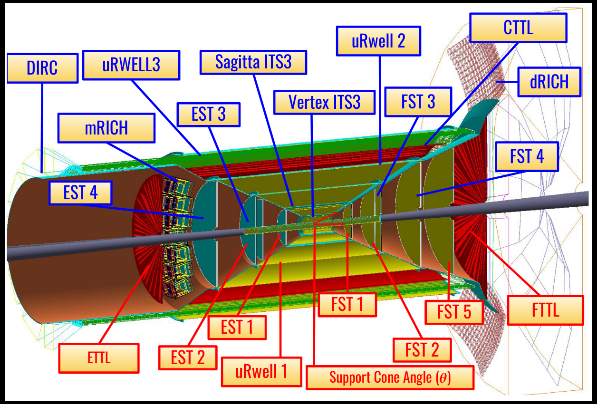
Tracker Optimization

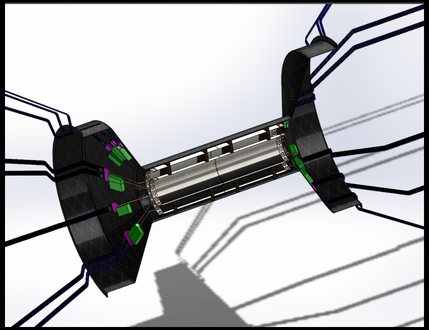
[arXiv:2102.08337]



- The performance of tracker characterized by detector's response (eg. resolution, reconstruction efficiency for the tracks). Often more than one metric.
- Geometric/Design parameters have significant impact in the performance of the tracker.
- Optimization is a continuous and iterative process. Each time add more subsystems (and services) when available. 11 parameters in this example.

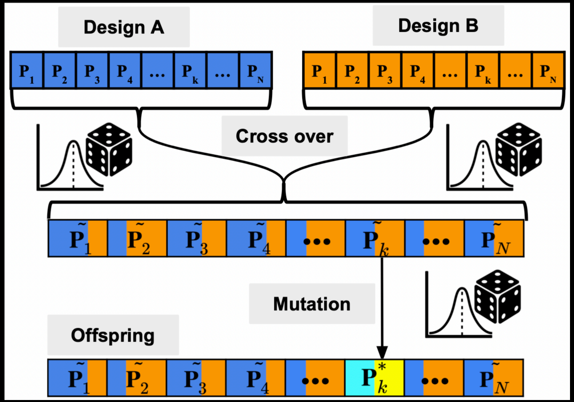
B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_M

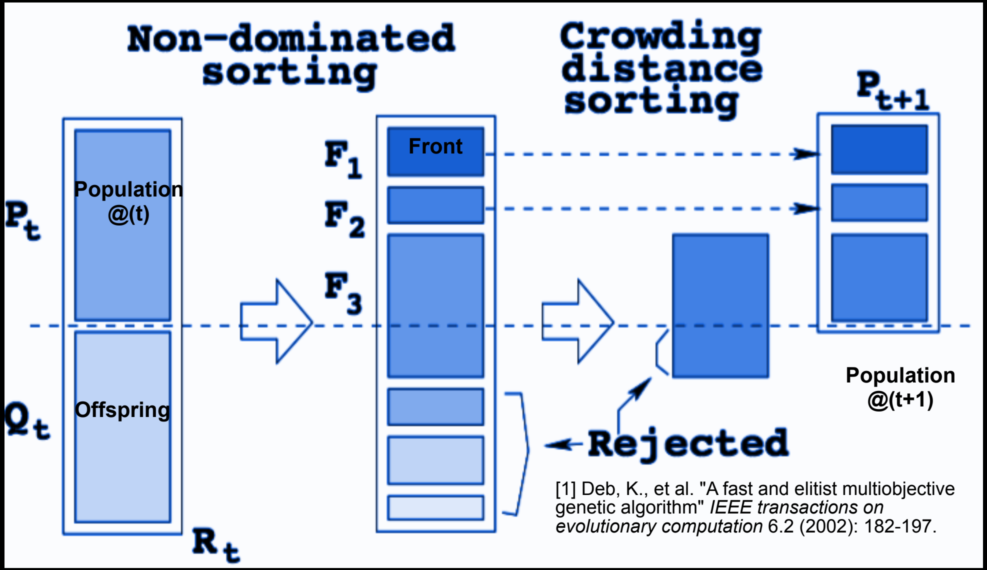




Implementation of Support Structures with realistic material Budgets.

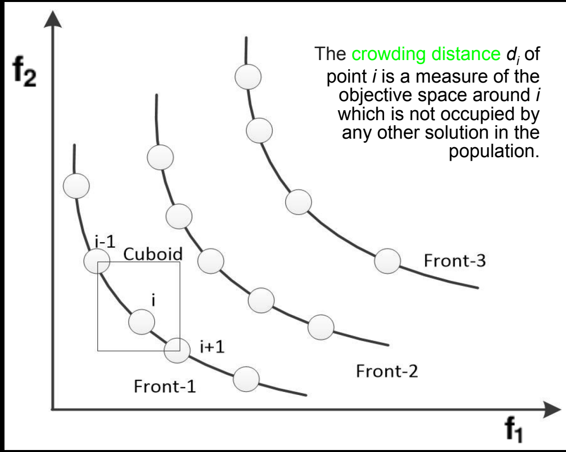
Elitist Non-Dominated Sorting Genetic (NSGA)





This is one of the most popular approach (>35k citations on google scholar), characterized by:

- Use of an elitist principle
- Explicit diversity preserving mechanism
- Emphasis in non-dominated solutions



The population R_t is classified in non-dominated fronts.
Not all fronts can be accommodated in the N slots of available in the new population P_{t+1}. We use **crowding distance** to keep those points in the last front that contribute to the highest diversity.

Software Stack & MOEA Ana

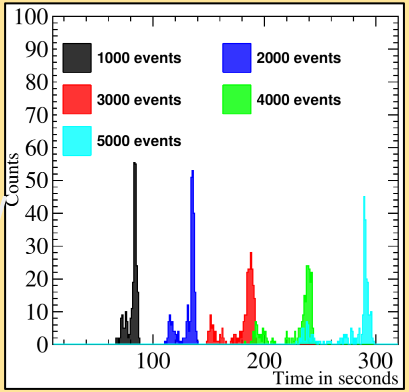
The Wrapper

Initialize Design Population
(Can "modify" gene in population)

AI-assisted design

Evaluate Design Points
Parallelize the Evaluations

Multi-objective Optimization

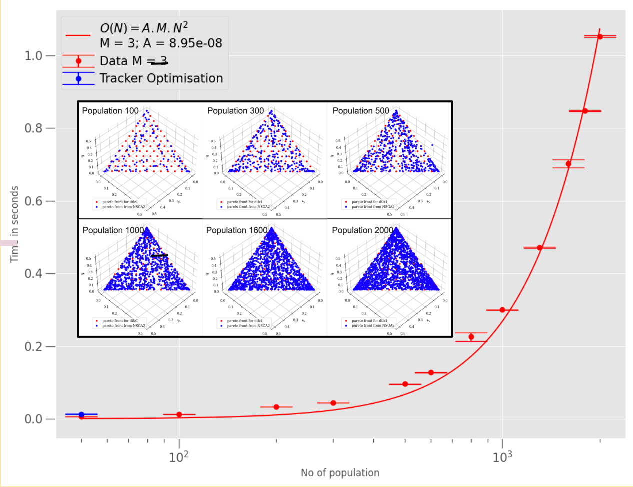


Simulating 80000 in total for each evaluation, 1 evaluation is <=80 mins

N_vars ≥ 11
N_gen = 200
N_population = 100
Offspring ≥ 30
N Cores ≥ 30
N tracks = 80k

Dask natively scales Python

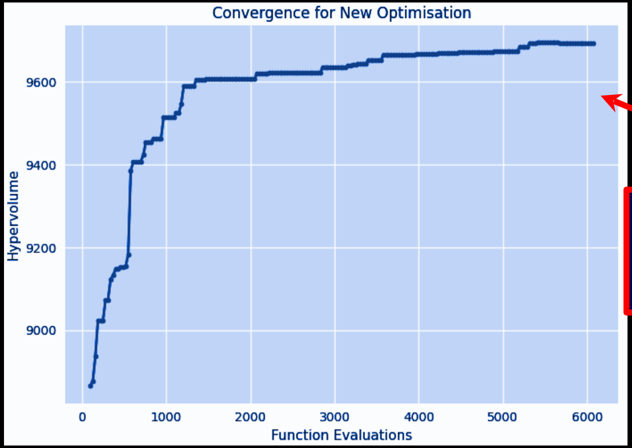
- Characterization of time taken by GA + sorting

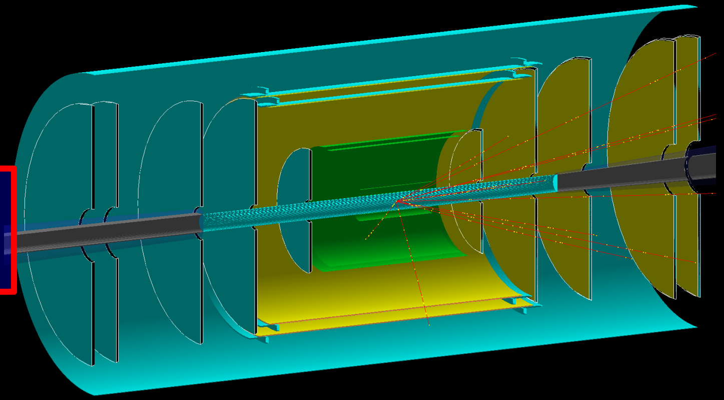


- Used a test problem DTLZ1
- Verified scaling following MN^2 and convergence to true front
- ~1s/call with 10^4 size!
- For 11 variables and 3 objectives needs ~ 10000 evaluations to converge

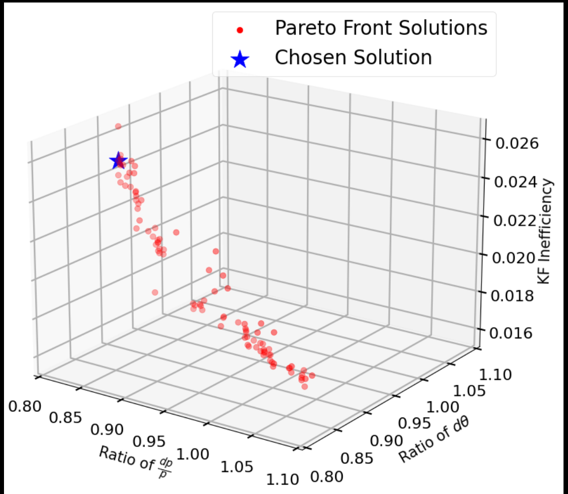
~10k CPU hours

Optimal Detector Design Solutions



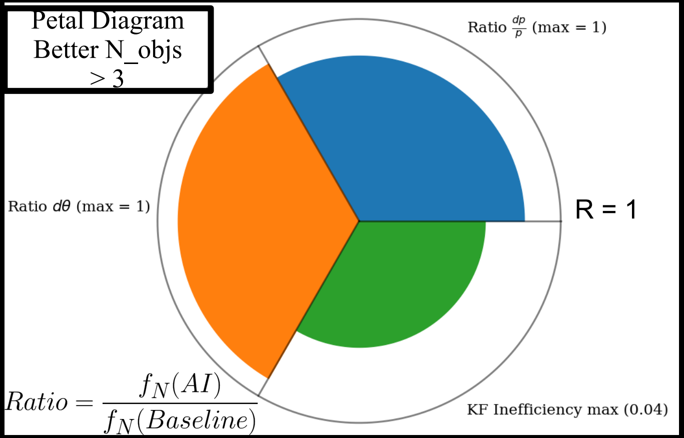


Inferring solutions at any stage of optimization

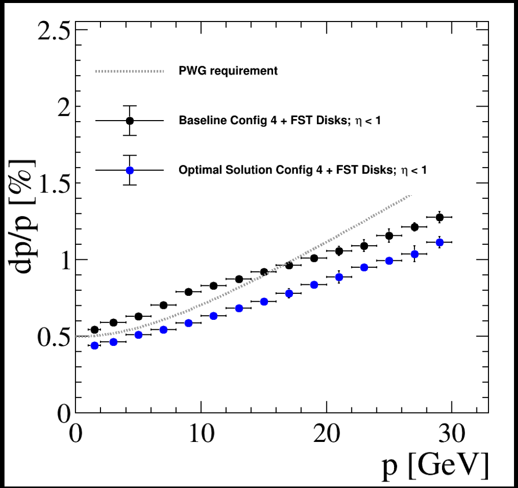


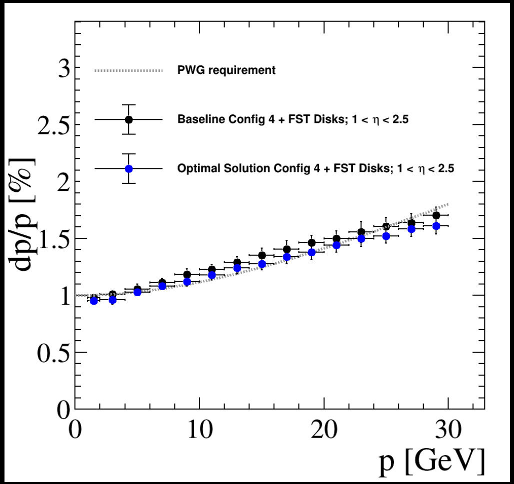
Example of solution

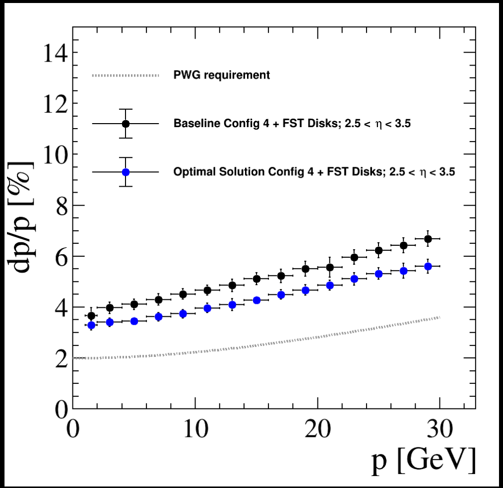
Performance of the chosen Solution

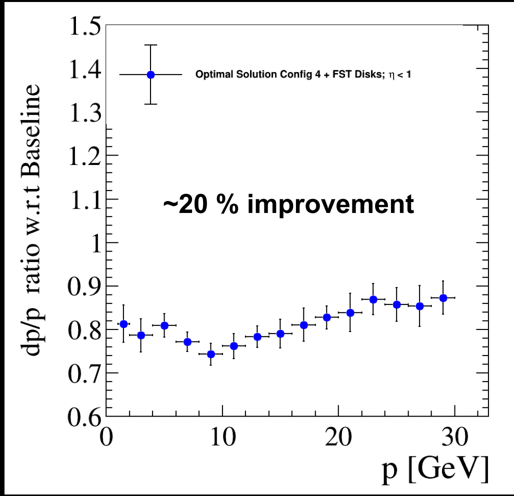


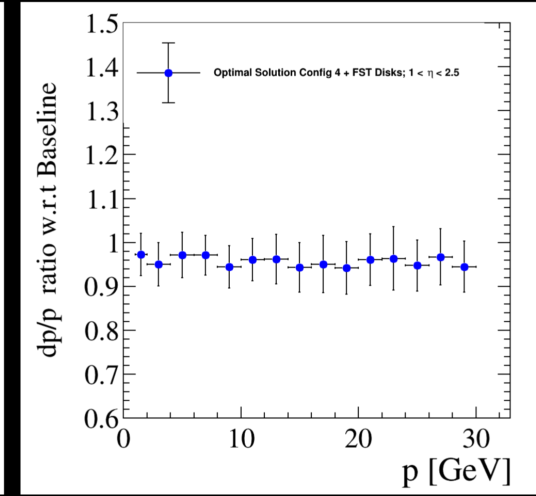
Momentum Resolution

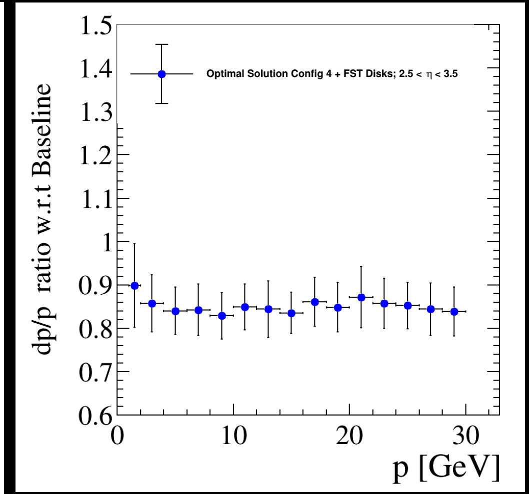






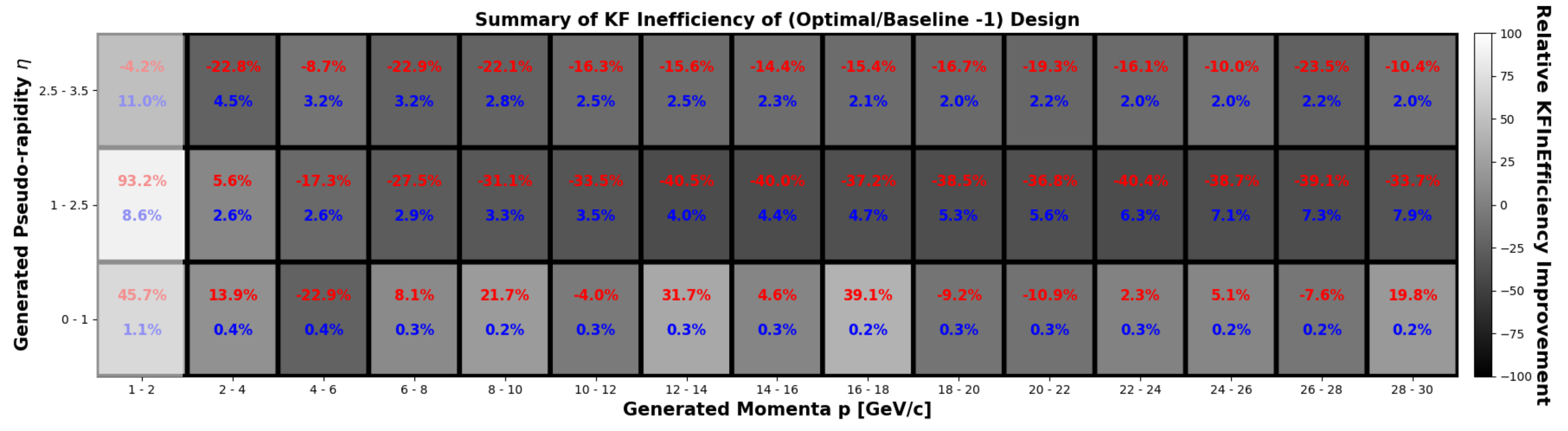




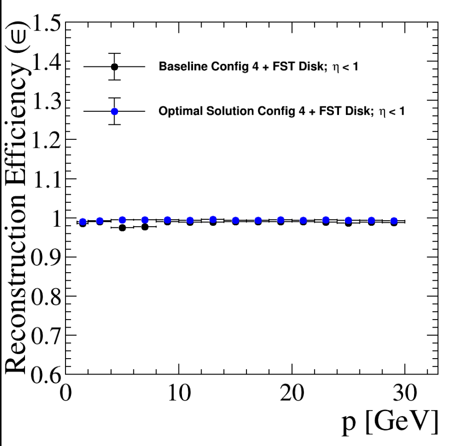


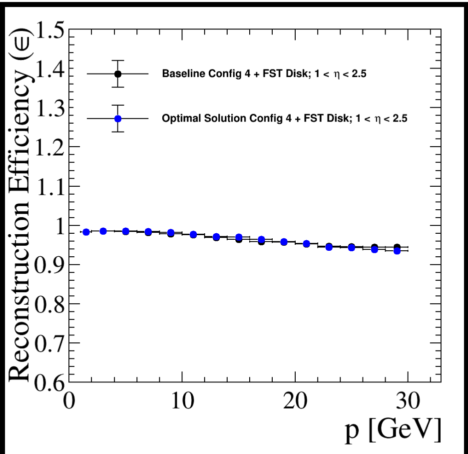
KF Inefficiency Improvement

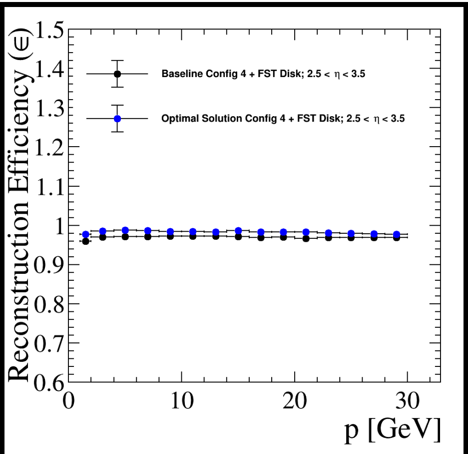
- Optimal/baseline -1
- Baseline Ineff



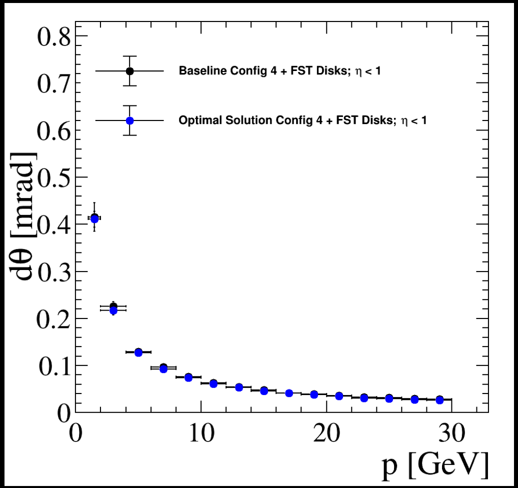
Validation Reconstruction Efficiency

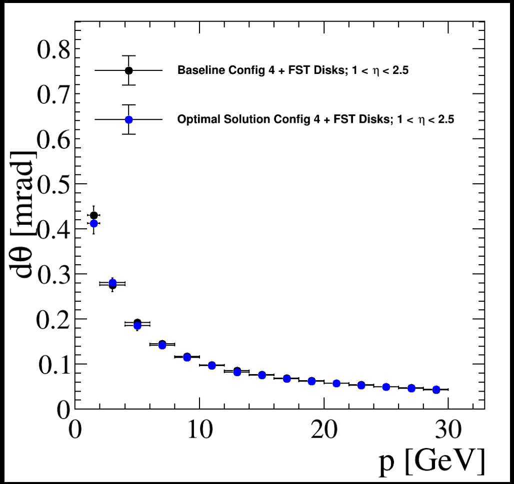


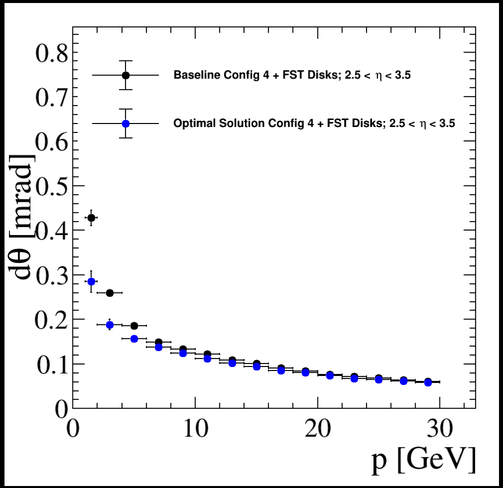


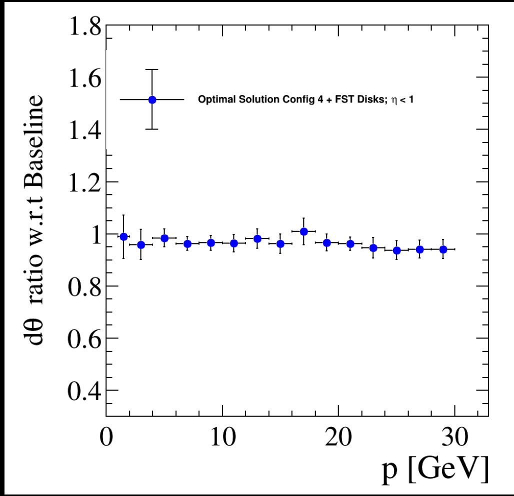


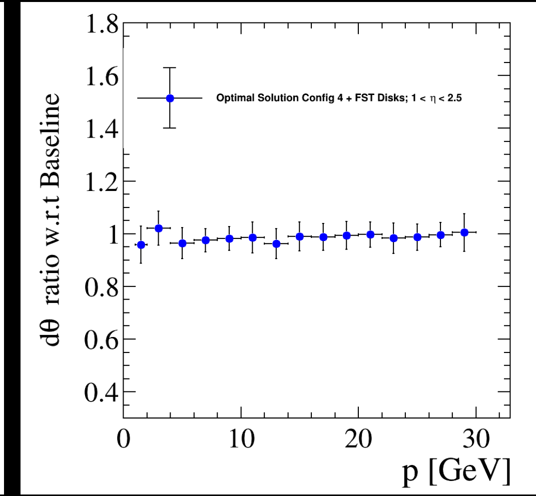
θ Resolution [mrad]

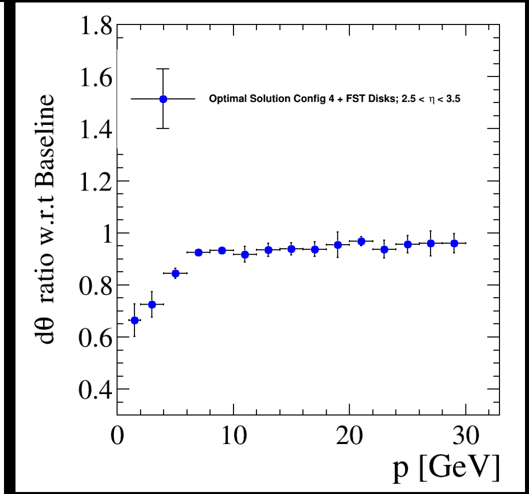




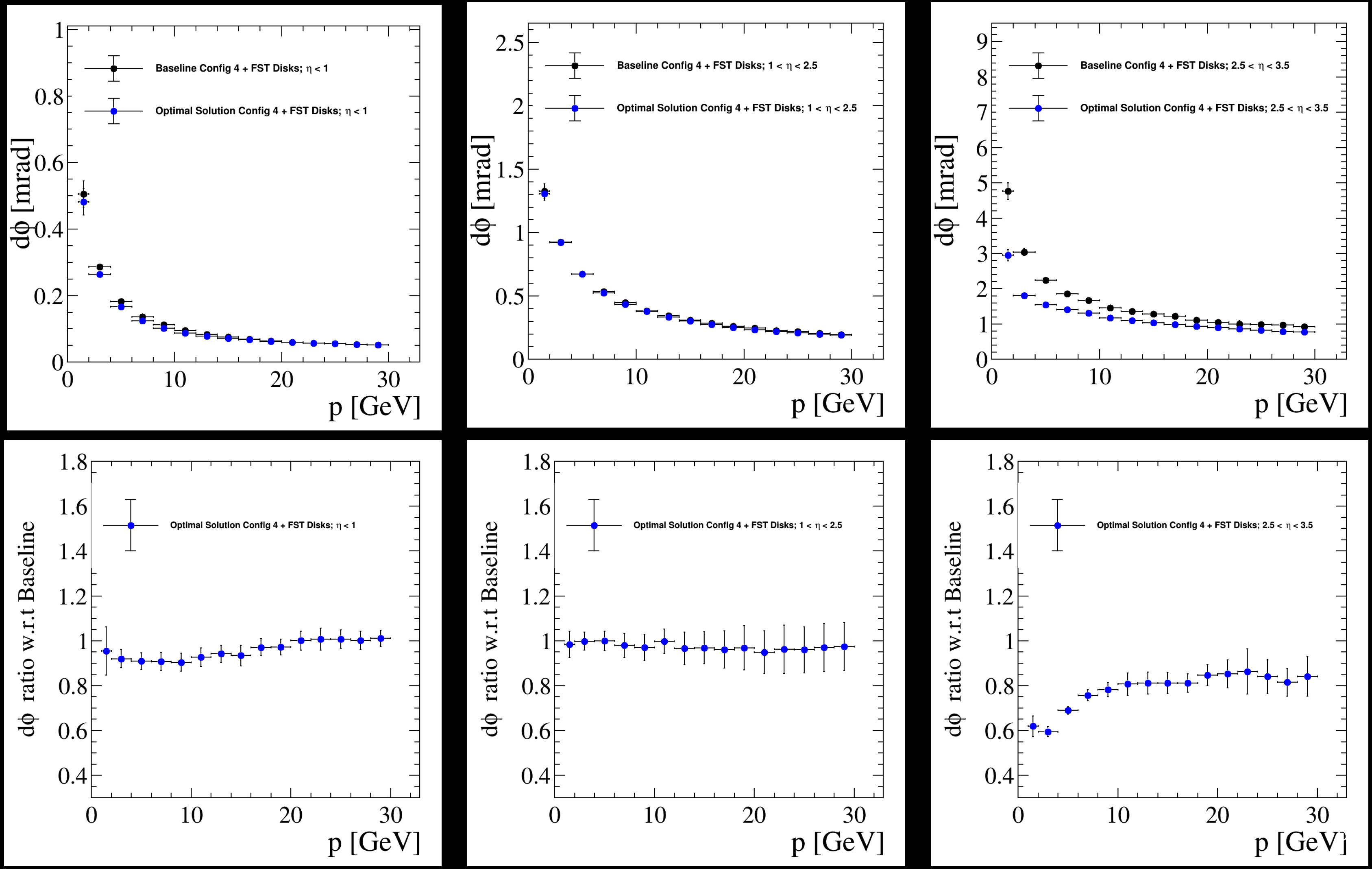






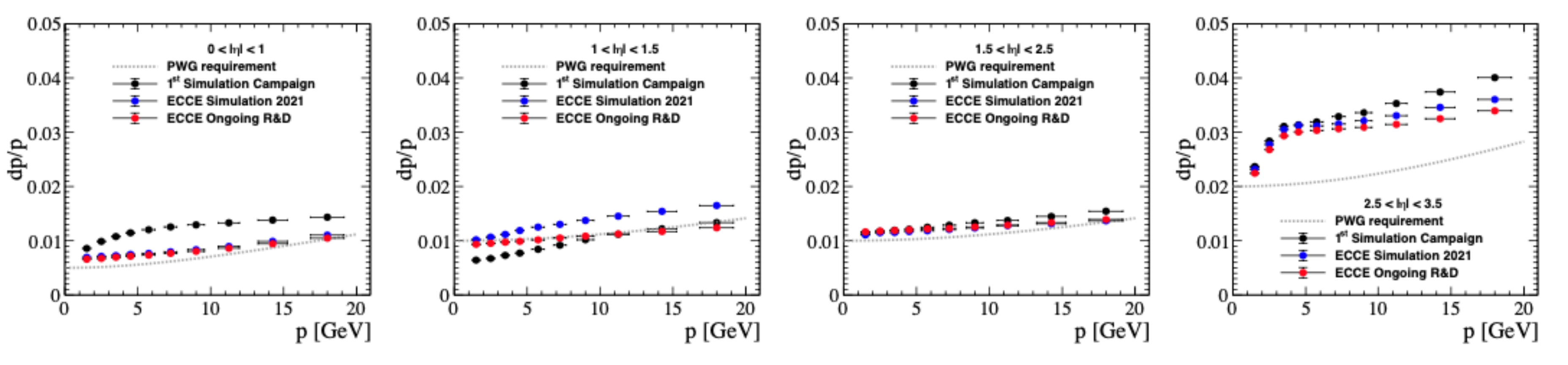


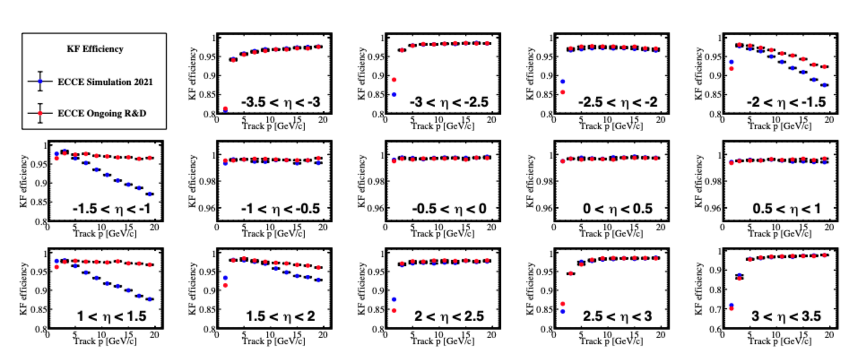
Validation φ Resolution [mrad]



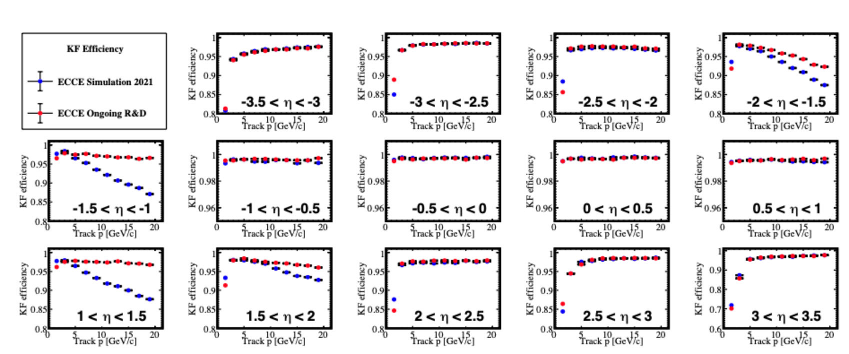
Evolution of Detector Performance (ECCE)

\



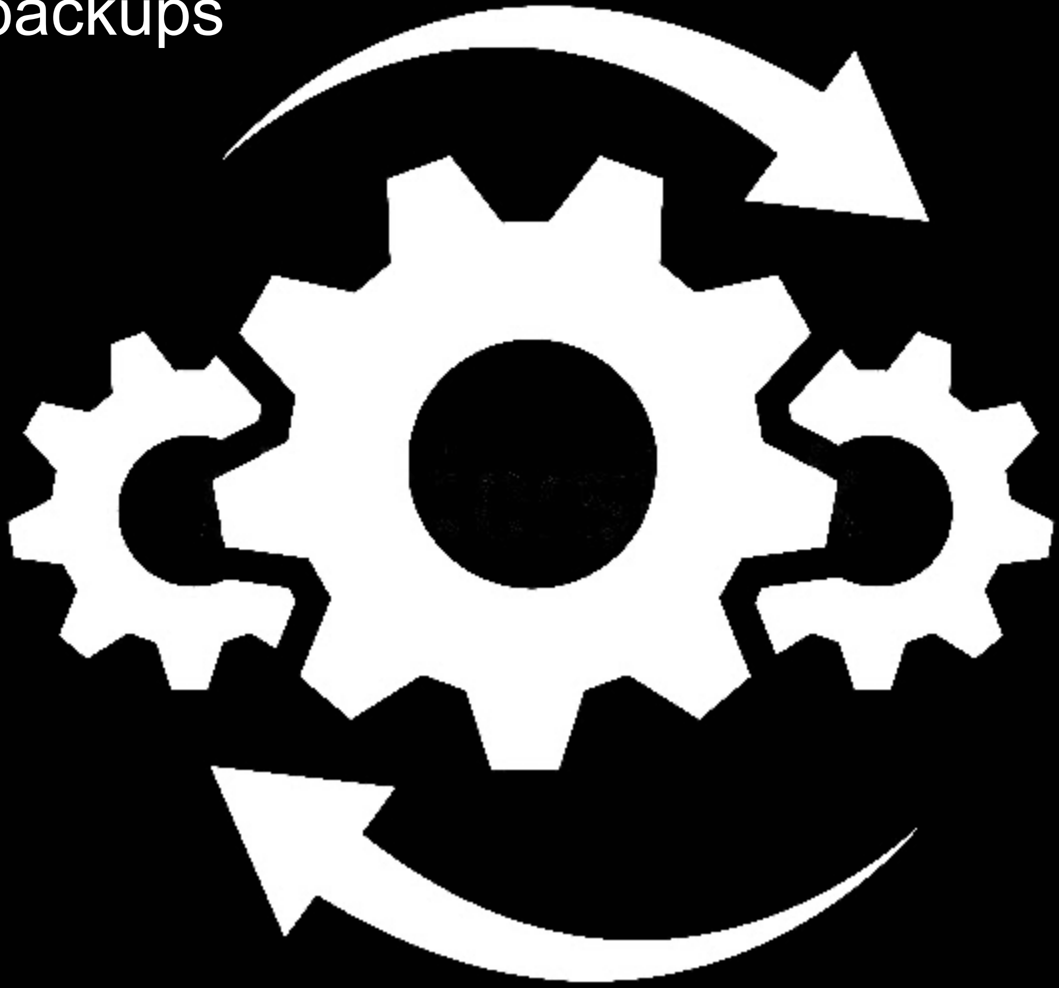


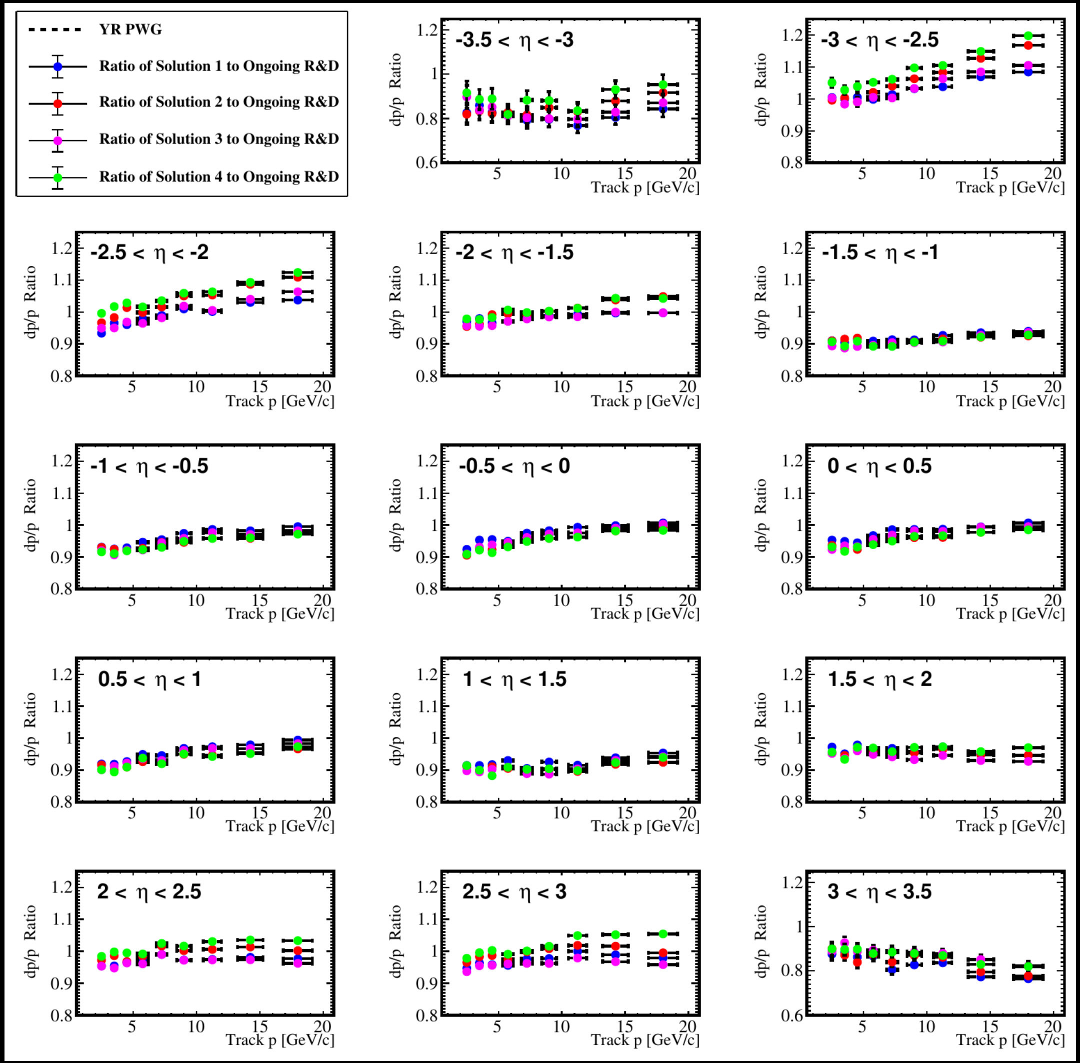
KF Efficiency



Polar Angular Reso

Second tier backups



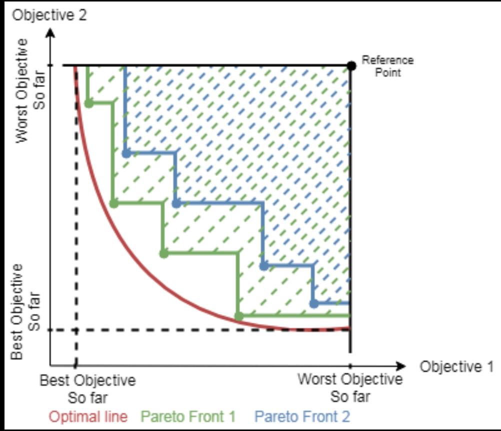


Highlights of this optimisation

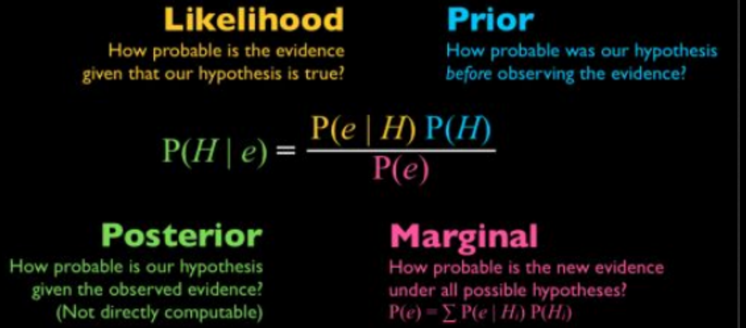
- Finer eta bins and momentum bins
- Includes almost all of the tracker subsystems for optimisation
- Includes optimisation of the support structures too
- Baseline detector setup corresponds to a projective design which itself is a result of previous phases of optimisation
- More optimisations with Bayesian based approaches are also carried out currently.

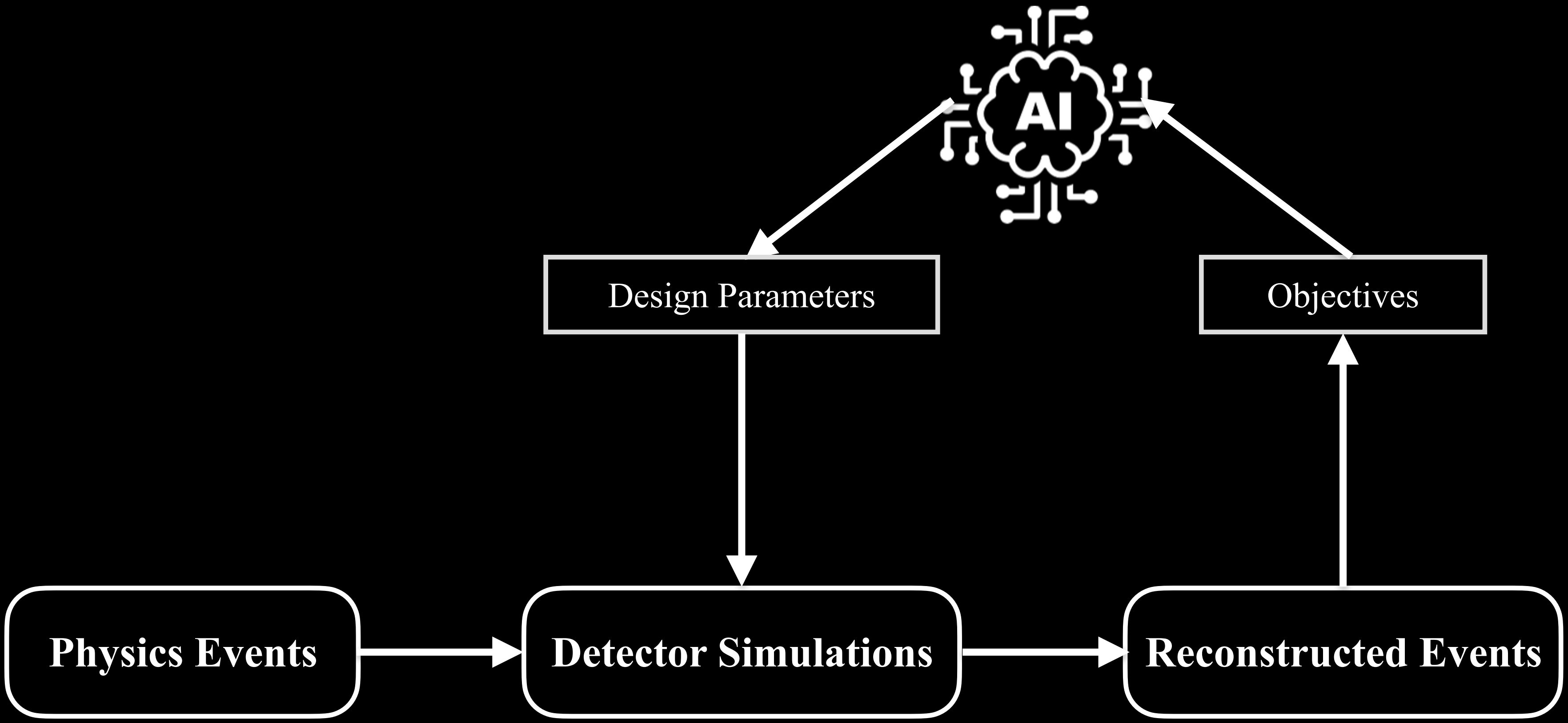
MOO details

- **Validating convergence.**
  - Look in the design space for improvements in the last few calls
  - Look into objective space. And perform cluster analysis on them
  - Make a custom metric to analyse convergence.
- **Hypervolume**
  - The volume of the First front w.r.t a reference point
- **Bayesian Optimization**
  - Used When the evaluation of each point is resource intensive.



Hyper volume definition




AI
Design Parameters
Objectives
Physics Events
Detector Simulations
Reconstructed Events