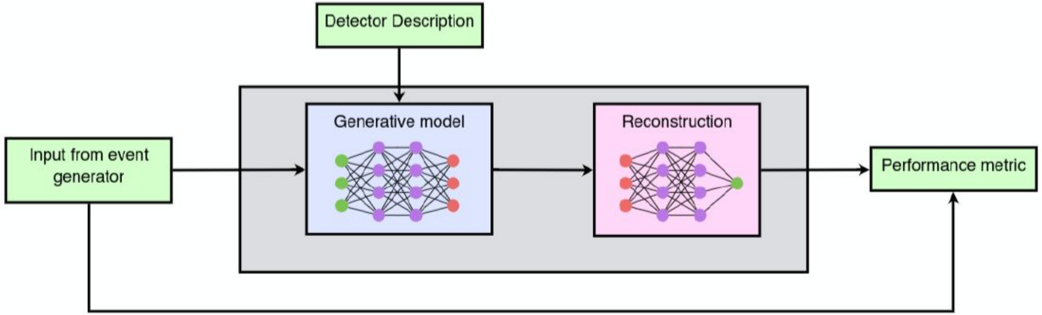
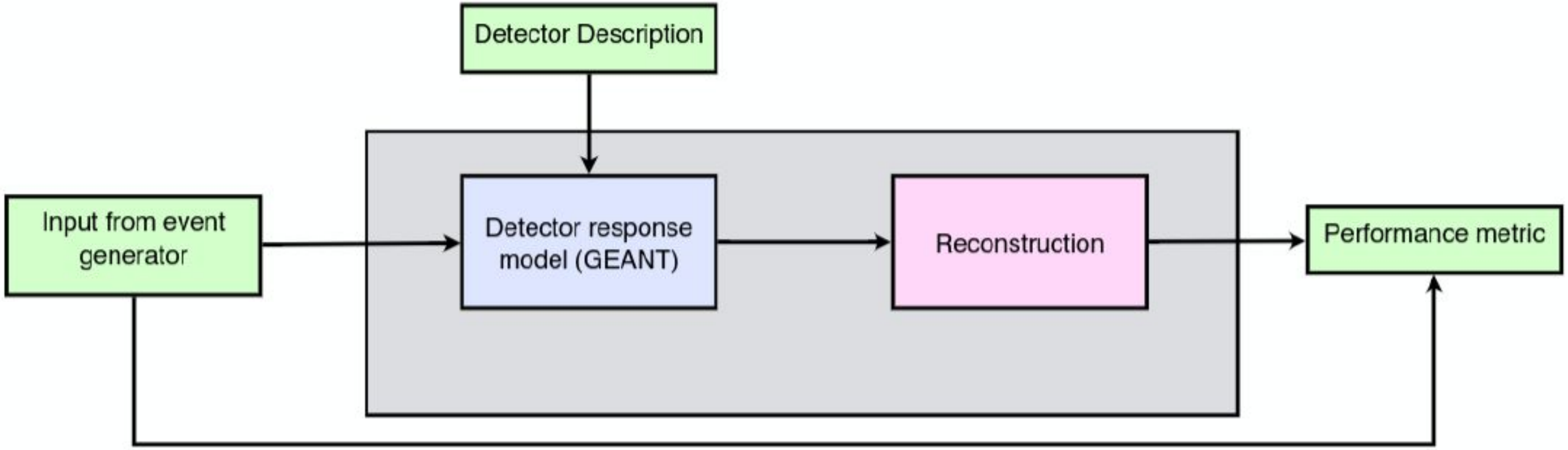
AI-Driven Detector Design for the EIC



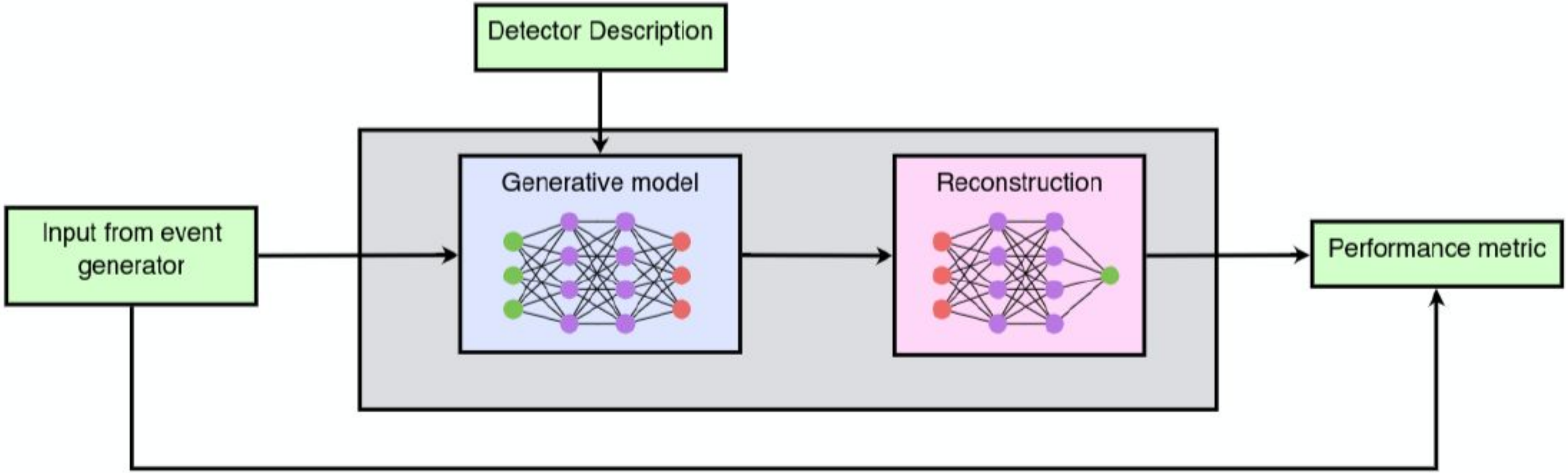
Miguel Arratia (UCR) on behalf of team, which included:
UCR: B. Karki (postdoc), R. Milton, S. Moran-Vasquez (students), K. Barish
LLNL: P. Karande (Staff Researcher, AI/ML expert), S. Sinha (intern) D. Dongwi, A. Angerami, R. Soltz
LBNL: F. Torales-Acosta (postdoc), B. Nachman

Optimizing a Detector Design typically follows:



Challenges:
1) Geant4 is slow, computational expensive and stochastic
2) Reconstruction code typically not optimized for a given Detector Description

AI for detector Optimization

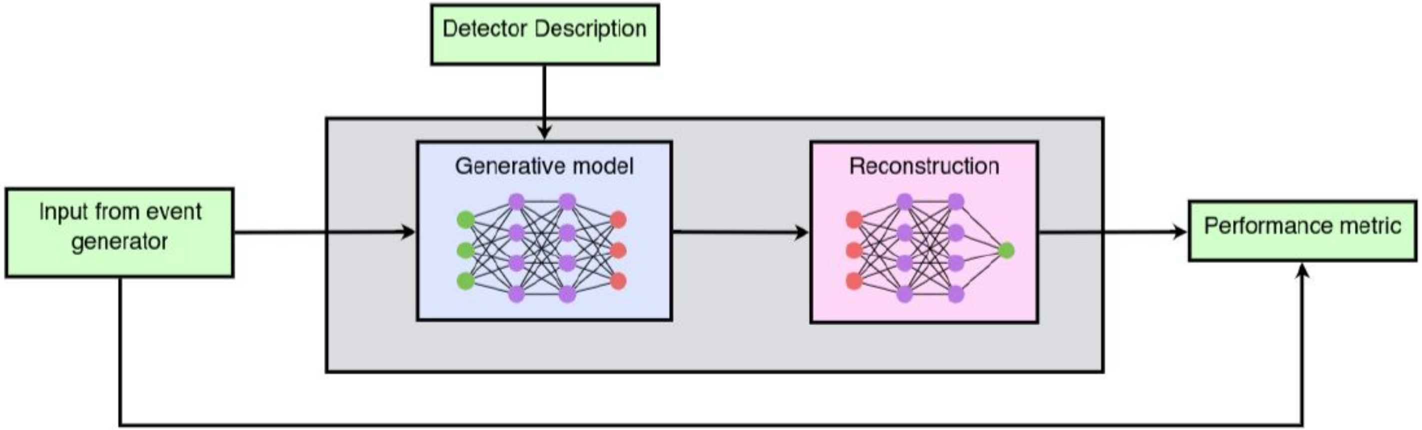


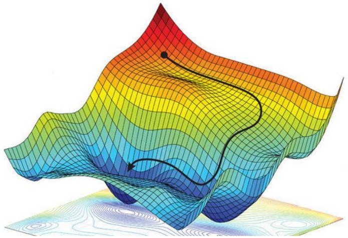
Advantages:

1) DNN generative model is fast and is differentiable
2) DNN based reconstruction is optimal for a given Detector Description

Ultimate Goal

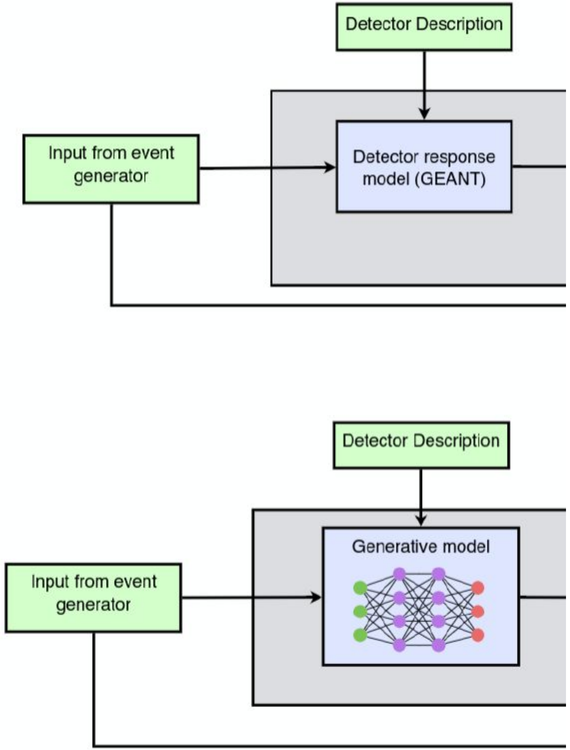
Our ultimate goal is to co-optimize the detector setup and the reconstruction algorithm, a challenging computational task that will be possible if both the generator and reconstruction are differentiable with respect to the detector parameters (*e.g.*, calorimeter segmentation). In this





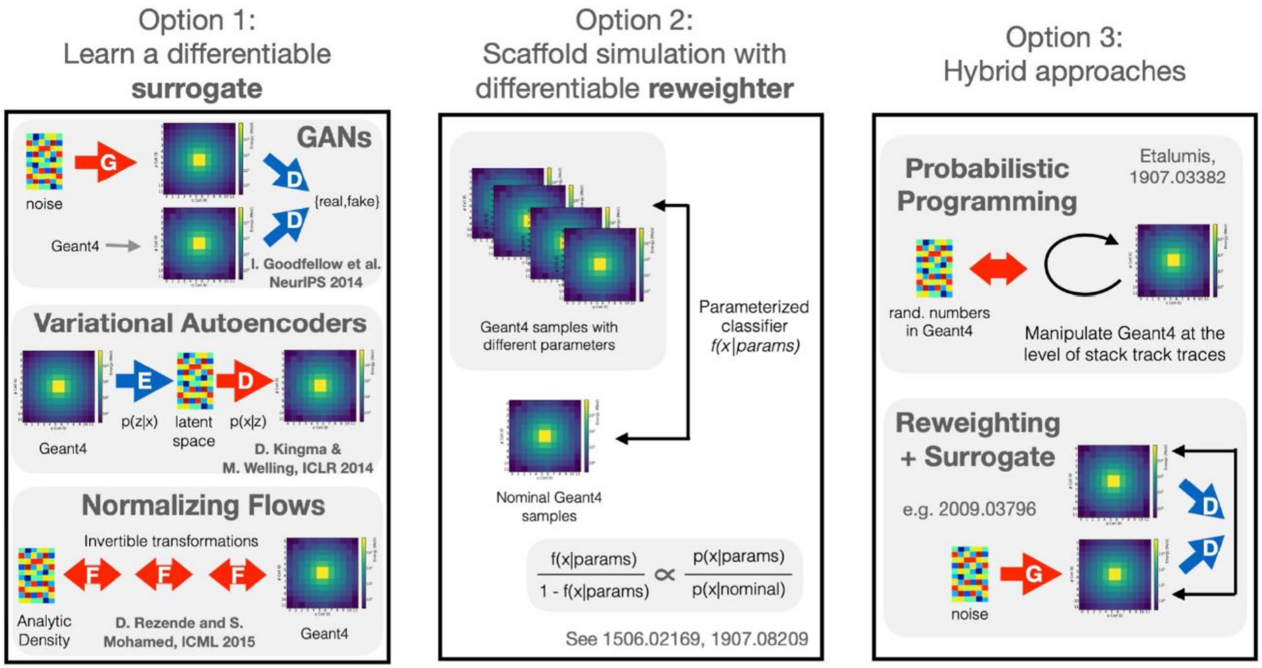
Make gradient-descent optimization possible

Part I. Generative Part



[Perlmutter@ NERSC]

State-of-the-art for generative models at the time of proposal



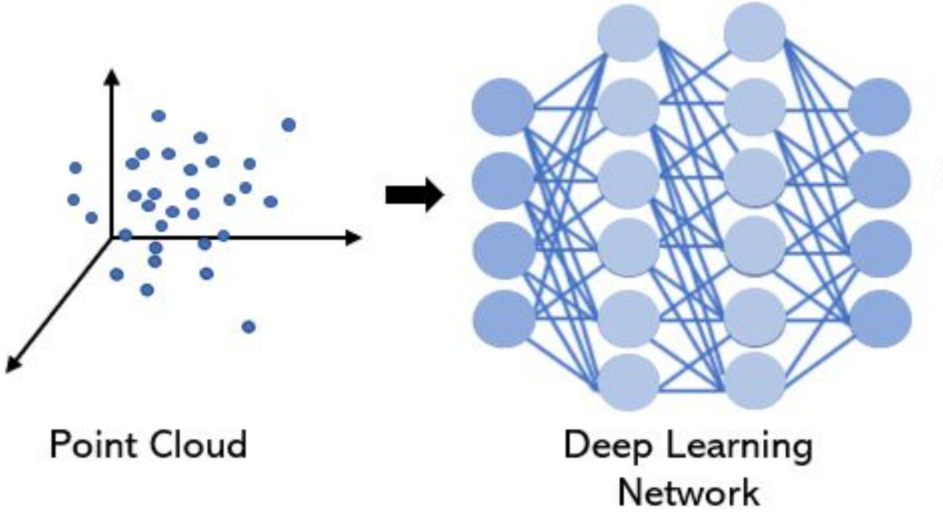
“we have chosen GANs for this proposal”

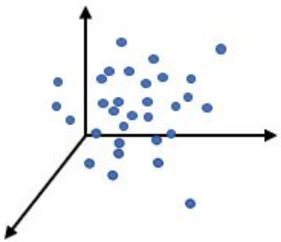
Figure 3. An illustration of three schemes for building a differentiable detector simulation. *The left column represents surrogate*

New emerging paradigm is based on Point Clouds

"Point-Clouds" architectures entered the field in ~2021, as alternative to image based approaches

First-ever use of point-clouds for Calorimeter fast generation was just introduced in 2023 (Buhman et al. arXiv:2305.04847)

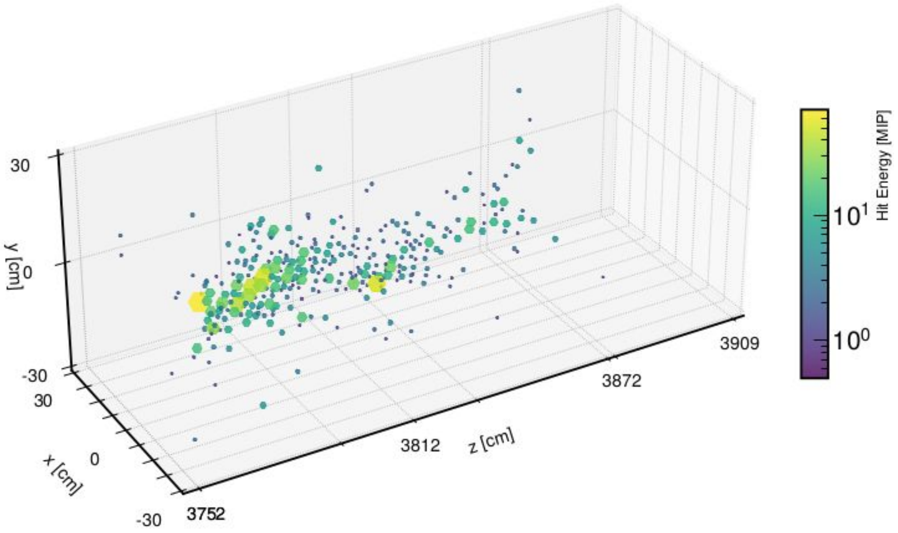


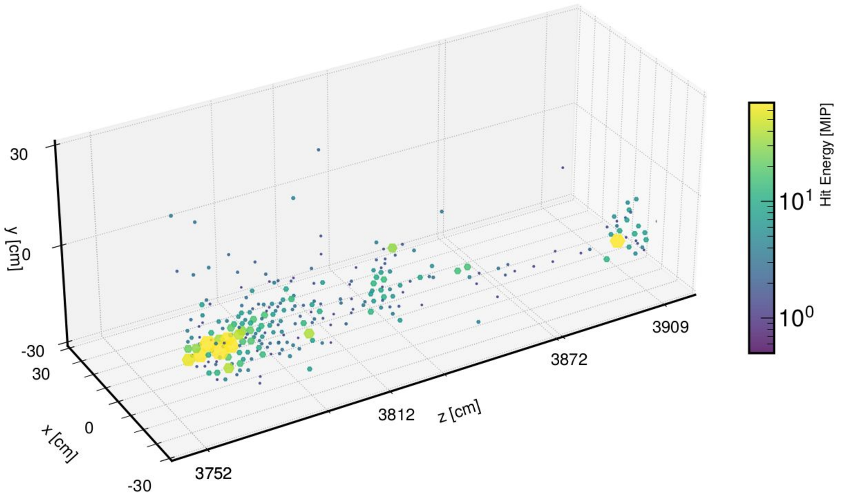


Calorimeter Data is naturally a point cloud!

Data is sparse (image format is wasteful)

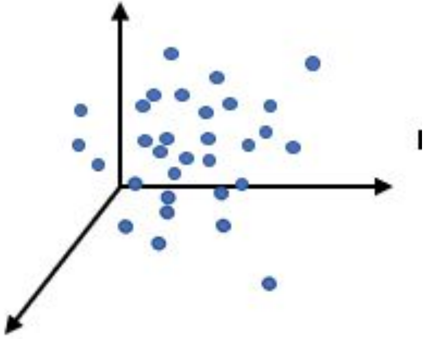
Typically detectors have non-regular grid geometry (images format is constraining)





Example of neutron showers in EIC high-granularity ZDC

How do point-clouds methods compared to traditional image based methods for calorimeter generation?



Point Cloud

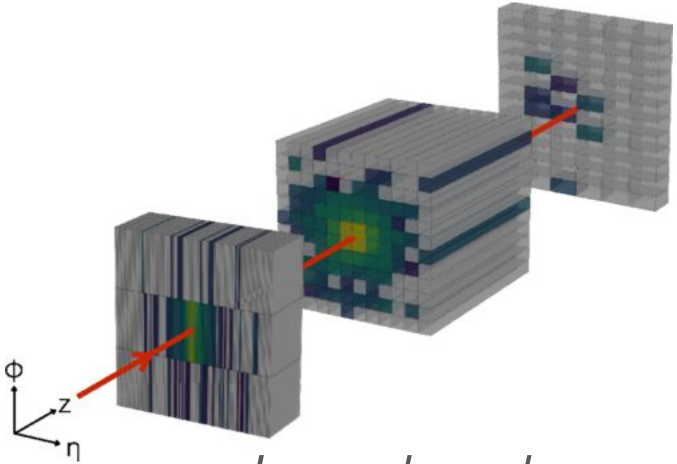


Image-based approach

arXiv:2307.04780v2 [cs.LG] 31 Jul 2023

In review in JINST

Comparison of Point Cloud and Image-based Models for Calorimeter Fast Simulation

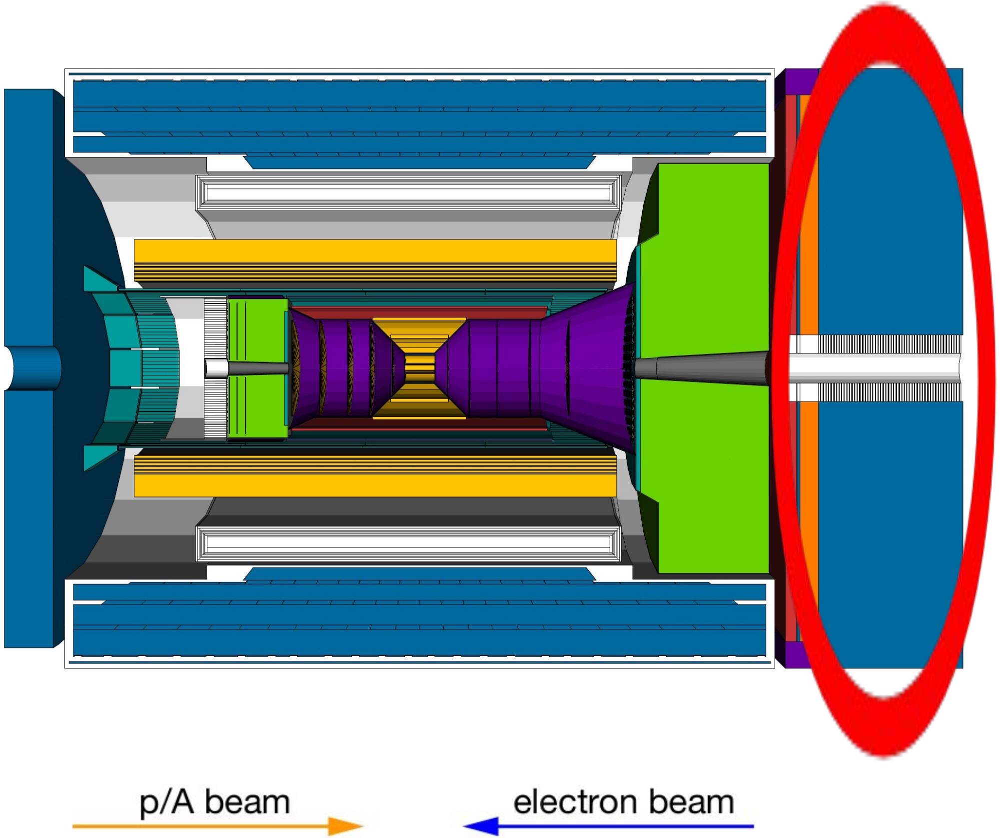
Fernando Torales Acosta,[1,*] Vinicius Mikuni,[2] Benjamin Nachman,[1,3] Miguel Arratia,[4,5] Bishnu Karki,[4] Ryan Milton,[4] Piyush Karande,[6] and Aaron Angerami[7]

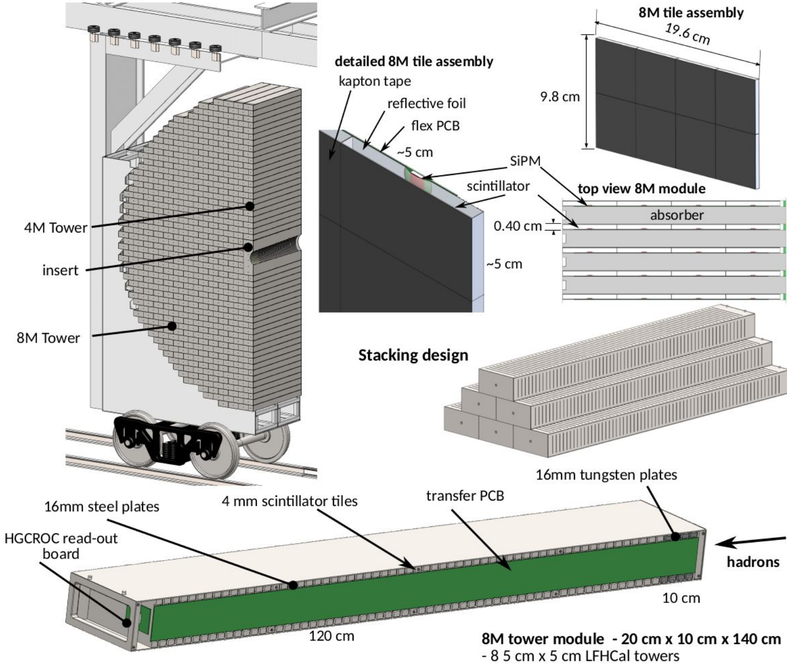
[1] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[2] *National Energy Research Scientific Computing Center, Berkeley Lab, Berkeley, CA 94720, USA*
[3] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*
[4] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[5] *Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA*
[6] *Computational Engineering Division, Lawrence Livermore National Laboratory, Livermore CA 94550*
[7] *Nuclear and Chemical Science Division, Lawrence Livermore National Laboratory, Livermore, CA 94550*

11 citations

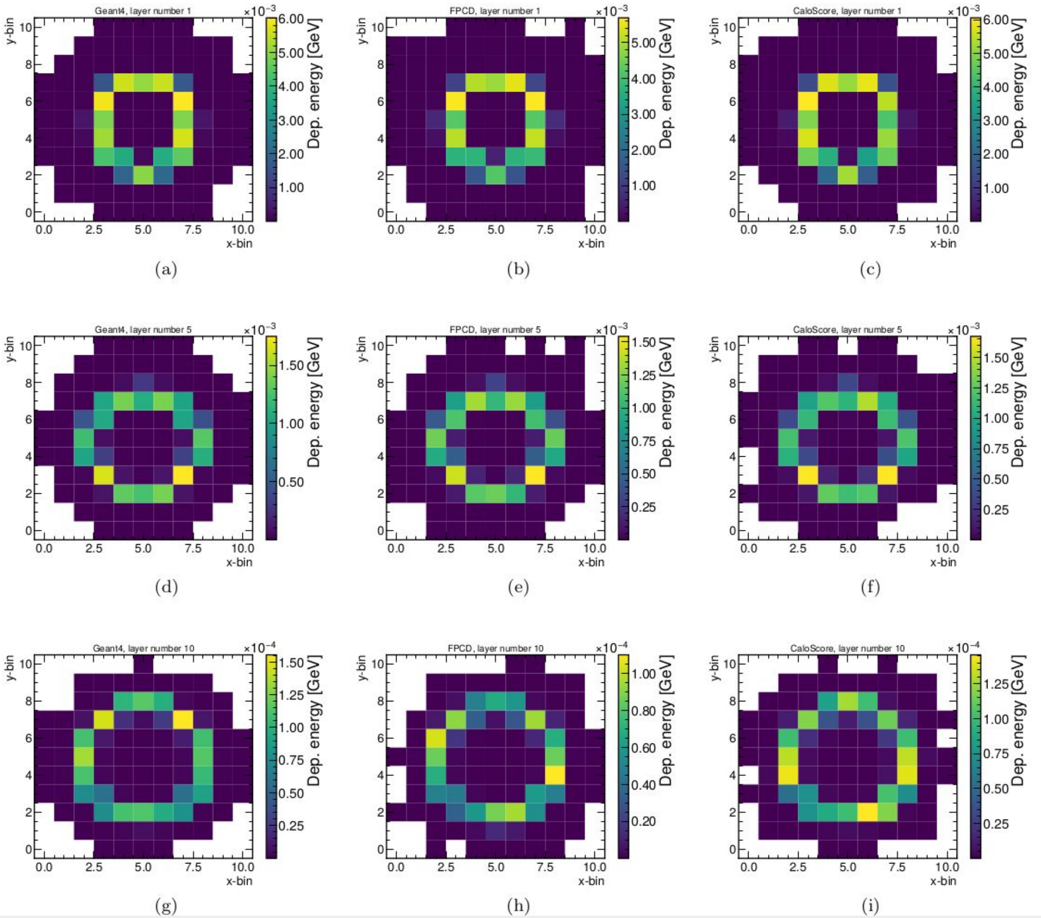
Case Study: EIC Forward Calorimeter

FORWARD HCAL
DSL/DSTC: Friederike Bock (ORNL)
Deputy DSL/DSTC: Miguel Arratia (UCR)





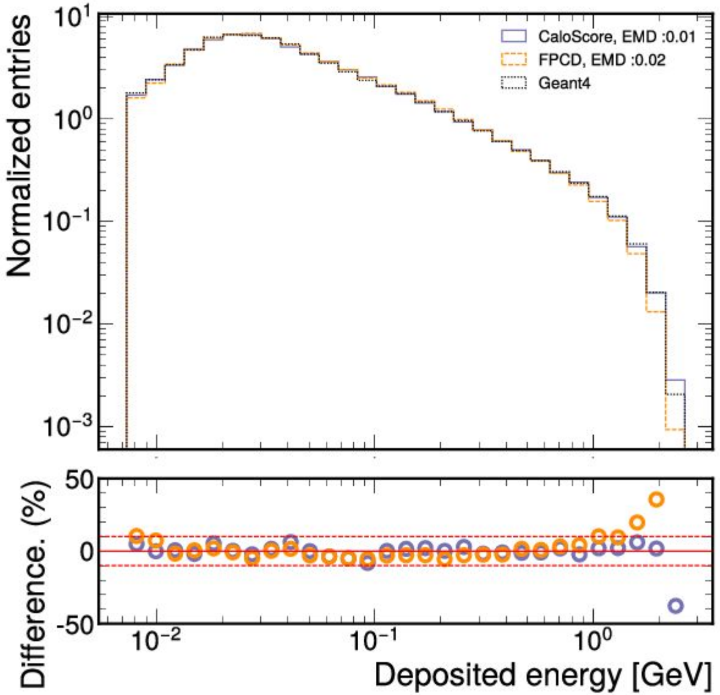
Geant4 vs Point-cloud vs Image

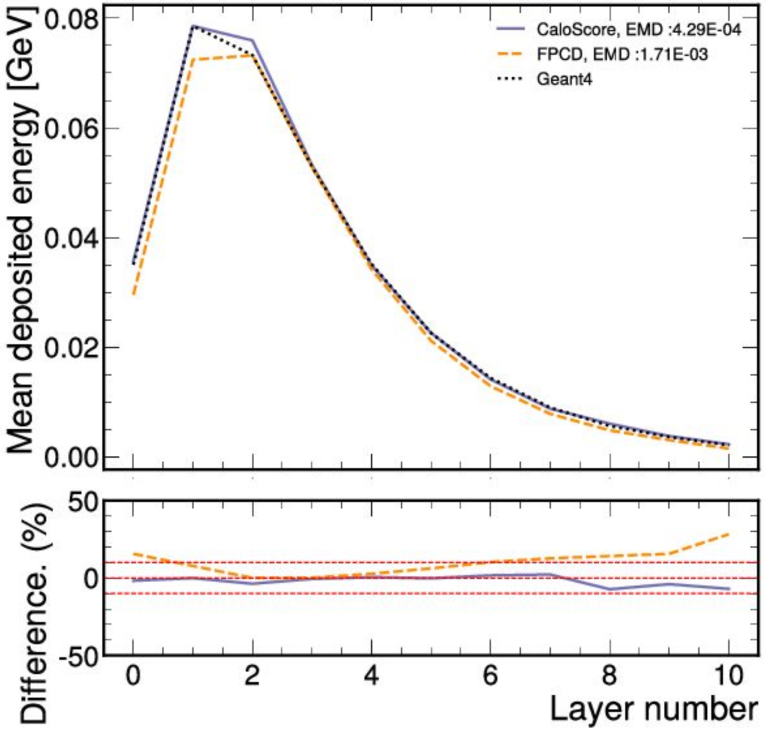


Average shower shape in three different layers

Point-cloud voxelized
to compare to image method

Image vs Point Cloud (voxelized) vs Geant4





Both methods described Geant4 quite well
(with a factor O(10,000) less computing time!)

Performance comparison

Model	# Parameters	Disk Size (Full)	Sample Time	AUC
Image	2,572,161	1016MB (62GB)	8036.19s	0.673
Point Cloud	620,678	509 MB	2631.41s	0.726

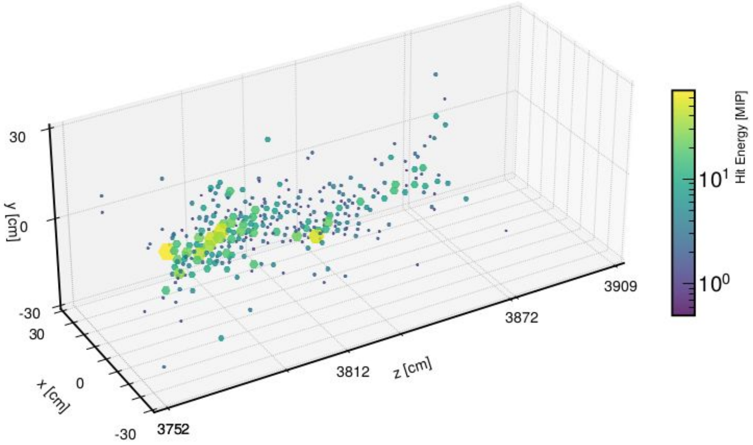
Point cloud is:

- A factor 100 times smaller dataset than image using same zlib compression.
- Samples data 3 times faster.
- Yields ~same performance when emulating Geant4.

arXiv:2307.04780v2 [cs.LG] 31 Jul 2023

"This work establishes a benchmark for future research on generative models, *offering valuable insights into the challenges of modeling hadronic showers in highly granular calorimeters using image-based techniques, while also exploring the potential of point-cloud methods.*

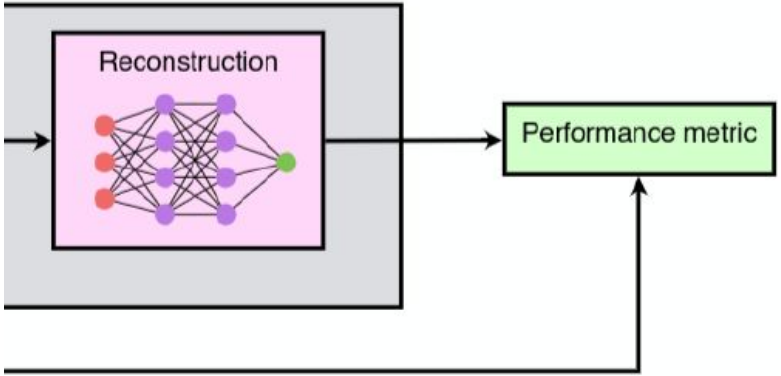
"The current advantages of point clouds, in combination with improvements to close the remaining performance gap described earlier, will likely make point cloud based models a clear choice for highly granular calorimeters"



1st ever AI/ML fast simulation for EIC

Part II. Regression Part

[Lassen @ LLNL]() Leveraging LLNL institutional Computing Grand Challenge program

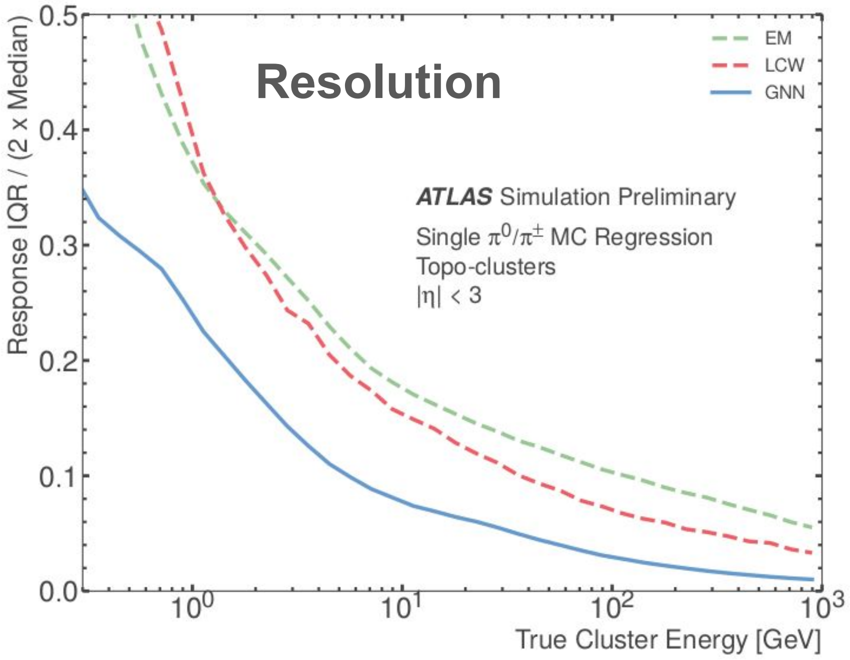


DNN provide us an optimal algorithm for a specific detector configuration, automatically

Some Historical Context for Perspective.

It takes years to fully optimize the software used in calorimeter systems.

For example ATLAS (designed in 1994):



https://cds.cern.ch/record/2825379

EM: (1994-) Basic Baseline

LCW: (2010-). Human-made algo, developed over many years

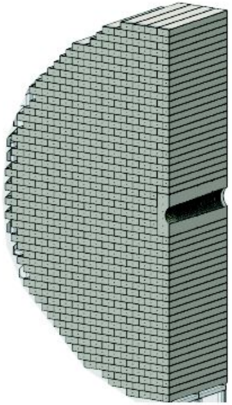
Deep learning 2020-

GNN: (2022-). Graph-Neural Network

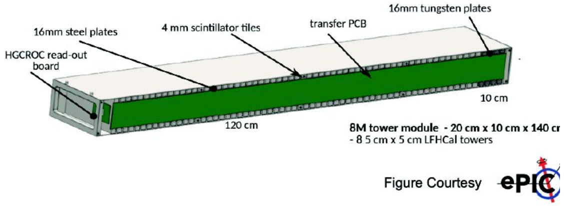
Case study: Longitudinal Segmentation of EIC forward calorimeter

Some of the key questions that our AI-driven optimization approach could answer are:

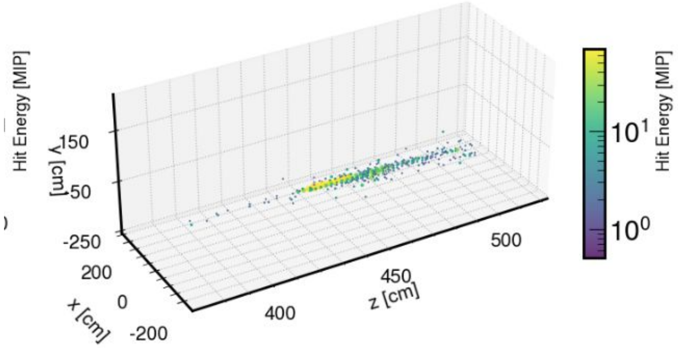
- Given a certain budget, what is the best performance one can expect in longitudinal readout?
- For which angles would a high segmentation have the largest impact?
- Where should the longitudinal layers be placed?



FORWARD HCAL
DSL/DSTC: Friederike Bock (ORNL)
Deputy DSL/DSTC: Miguel Arratia (UCR)



Tower longitudinal granularity is subject to optimization



A shower example at max granularity

arXiv:2310.04442v1 [physics.ins-det] 2 Oct 2023

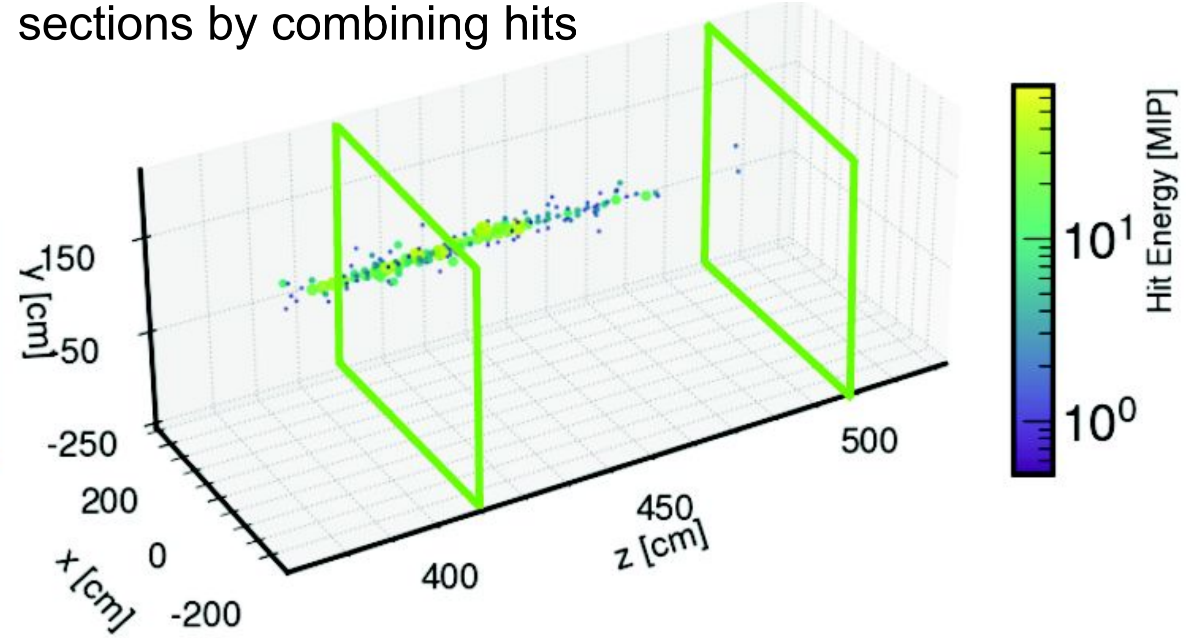
In review in JINST

The Optimal use of Segmentation for Sampling Calorimeters

Fernando Torales Acosta,[1,*] Bishnu Karki,[2,†] Piyush Karande,[3] Aaron Angerami,[4] Miguel Arratia,[2,5]
Kenneth Barish,[2] Ryan Milton,[2] Sebastián Morán,[2] Benjamin Nachman,[1,6] and Anshuman Sinha[3]

Data Process for Models.
We explore various longitudinal granularity scenarios

Example of 3 longitudinal sections by combining hits



This could literally be done in hardware too

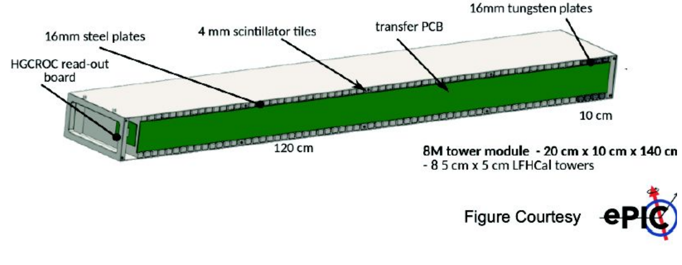
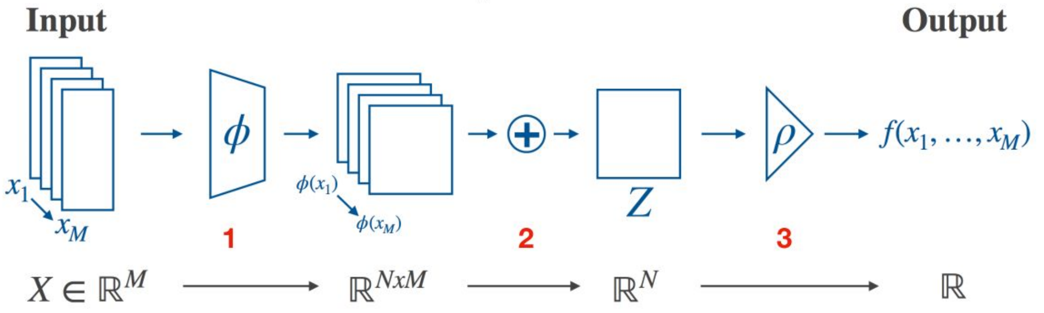


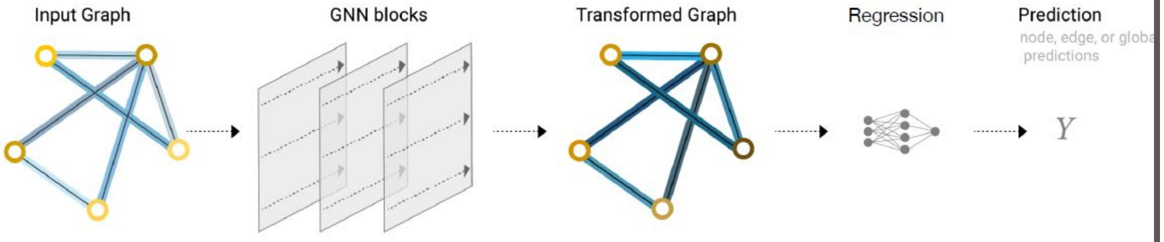
Figure Courtesy ePIC

Two AI/ML architectures explored, with same I/O

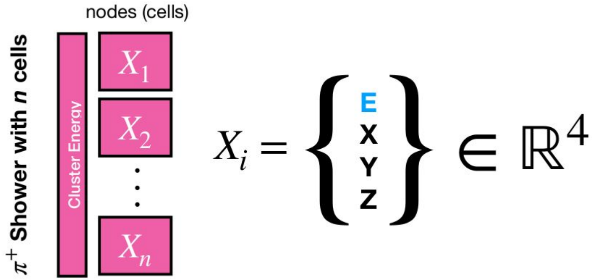
Deep Sets



Graph Neural Network



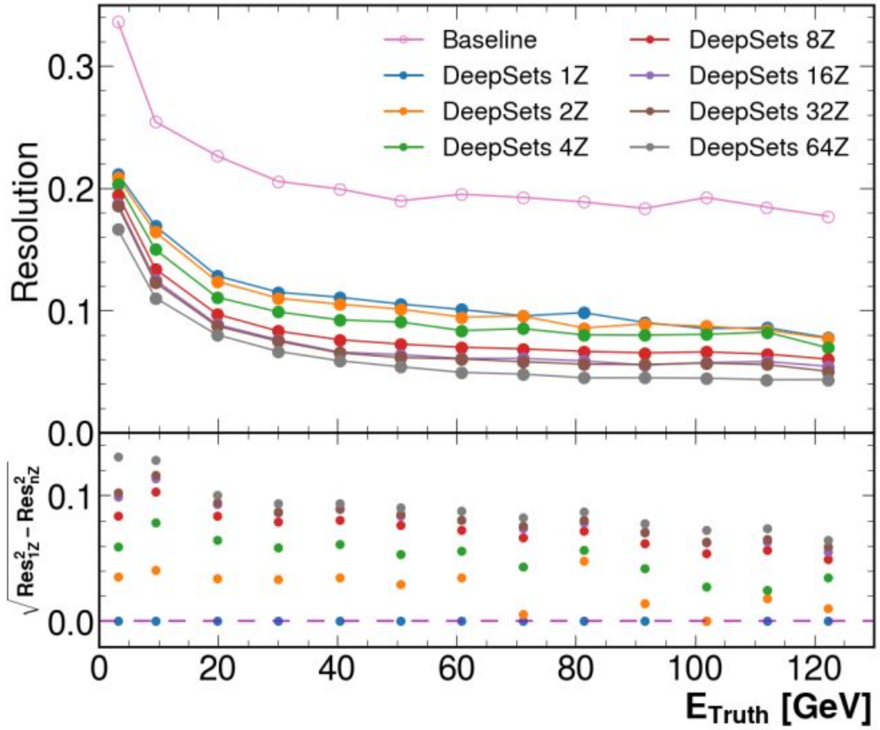
INPUT (Point Cloud):

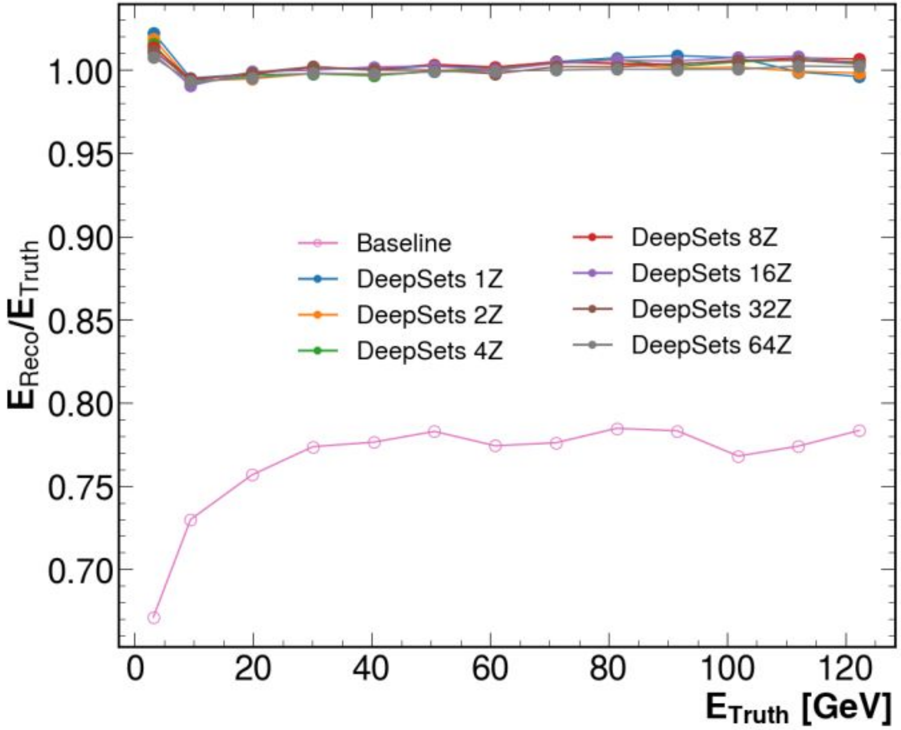


OUTPUT:
Energy and angles

EIC Calo Performance (ECAL + HCAL) with Optimal Reconstruction

under various longitudinal granularity scenarios

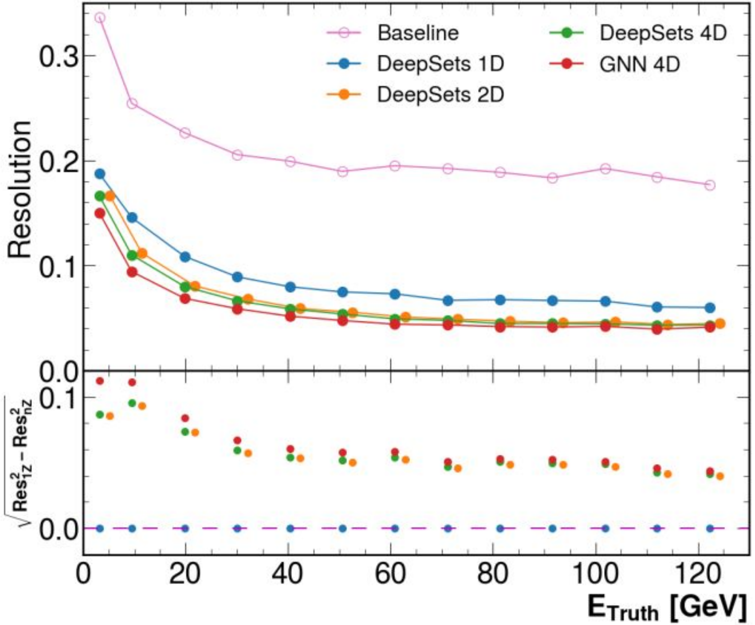


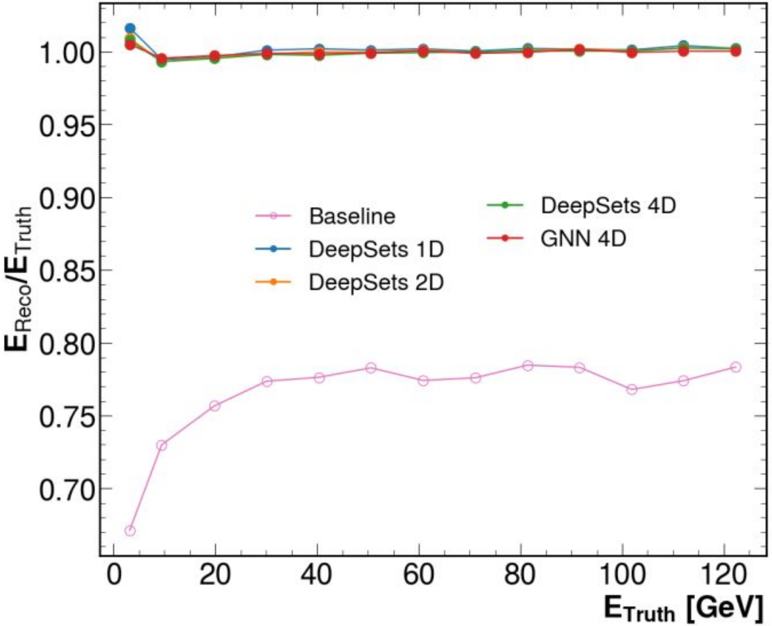


Quantifying potential gains with increased readout

What information drives performance? $X_i = \left\{ \begin{array}{c} E \\ X \\ Y \\ Z \end{array} \right\} \in \mathbb{R}^4$

We train Deepsets models on E, E+Z, E+XYZ (1D, 2D, 4D)





Very timely study given recent key EIC forward calo review

From Sep 2023 review:

2. Are the plans for achieving detector performance and construction sufficiently developed and documented for the present phase of the project? Specifically, are they commensurate with the initiation of the LFHCAL absorber and casing steel procurement?

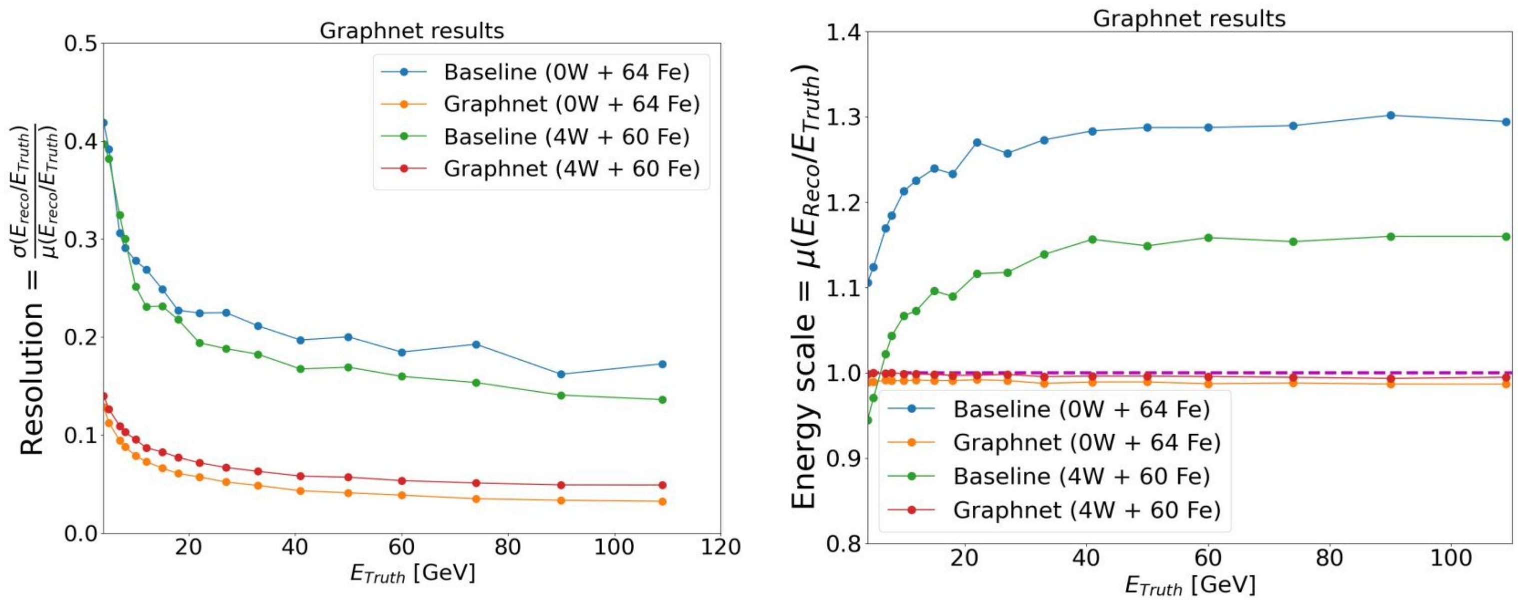
- FB: Yes, advanced detector performance simulations have been shown taking into account the realistic detector geometries. Moreover, detailed construction plans have been presented.
- The baseline questions are well understood but an optimization study should be done.
- Recommendation:
  - Implement software compensation as soon as possible and re-assess the benefits of the tungsten section.

Our project delivered AI/ML tools that significantly enchanged our ability to explore these

“Implement software compensation as soon as possible and re-assess the benefits of the tungsten section.”

Enhanced capacity to perform these with AI/ML!

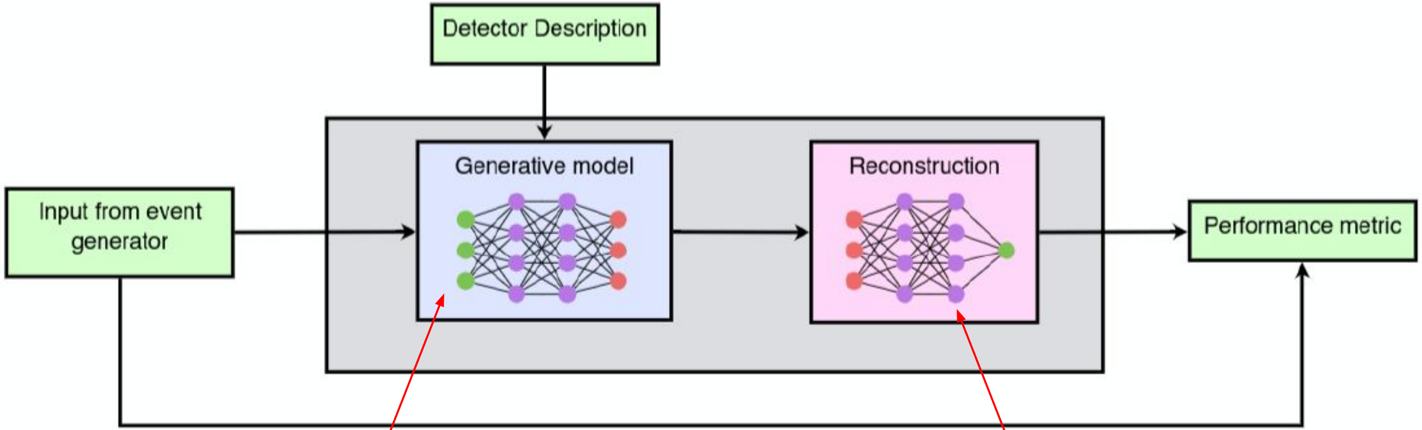
Performance comparison between 0W and 4W results



- We implemented “Software compensation” with AI/ML. Optimal software for each design choice. Note that this would have taken far too long to do otherwise with traditional methods.
- As far as we know, this is the first instance were AI/ML is used to inform a non-trivial design choice in EIC. (tungsten vs no tungsten section in HCAL is a million dollar type question). Links: [one](), two.

Recap: AI-methods have real impact on EIC detector design

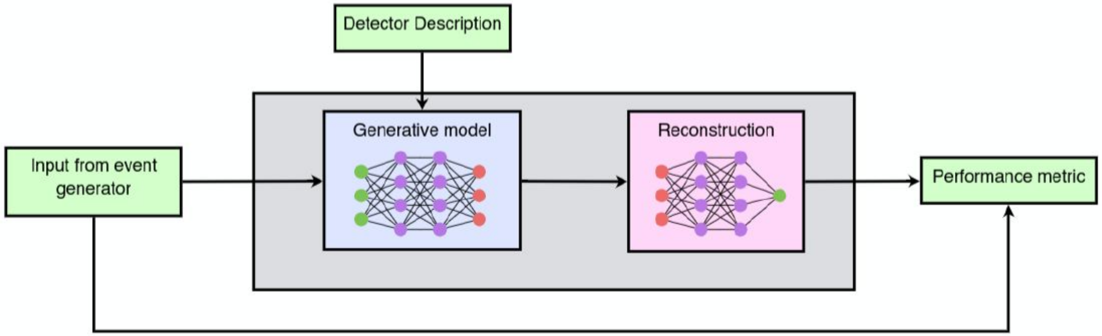
In general, automatic and optimal methods for data-generation and reconstruction save time and enable us to better explore various design options.

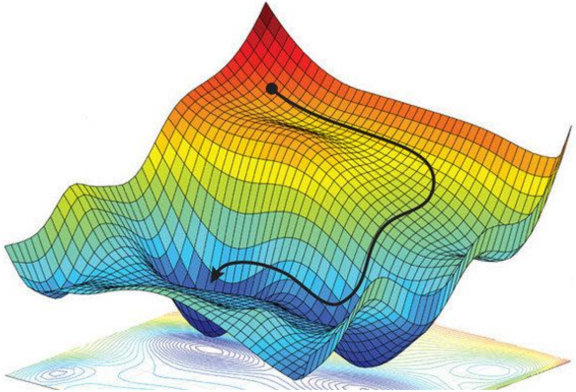


100,000 factor reduction on computational cost of bottleneck of simulations (Geant4 modelling of calorimeter showers)

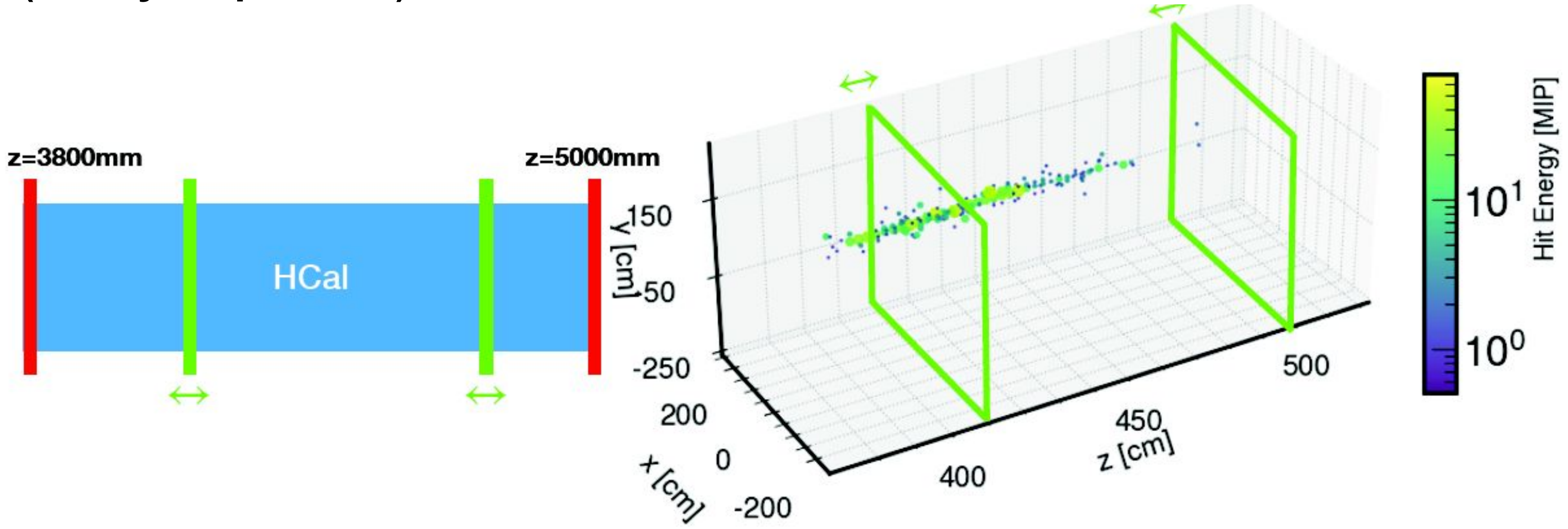
Automatic and optimal software reconstruction, which would take too much time to develop for every detector design option

Part III. Combined Generative model and Regression





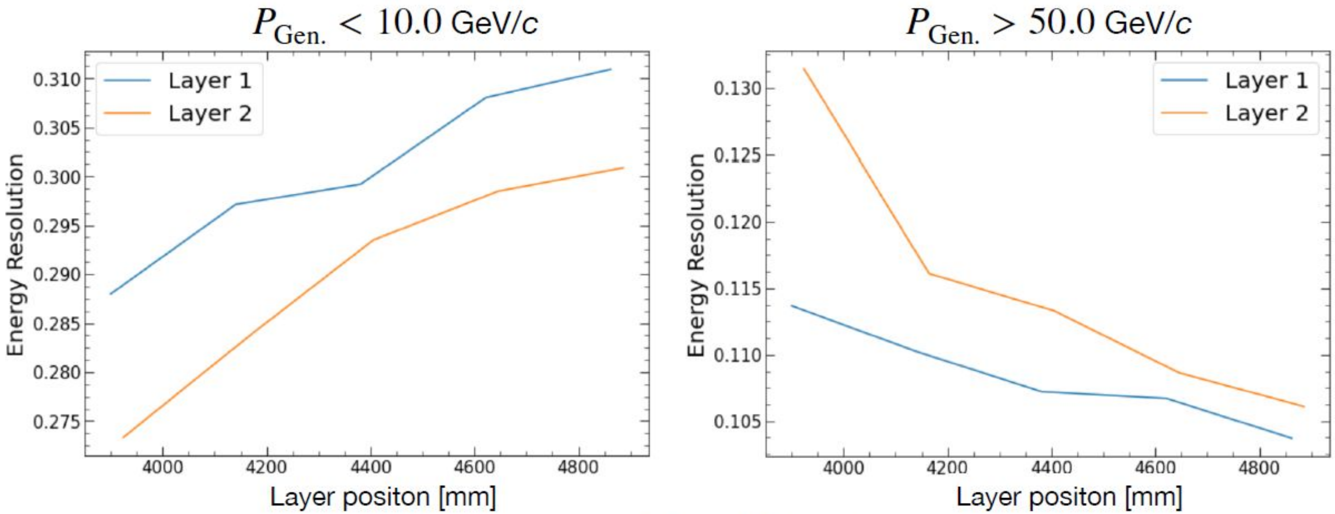
We processed the data to explore conditioning on detector parameters (i.e. layers position).



The network now gets trained on various layer positions, and can later interpolate

Differentiable function of energy resolution conditioned on detector parameters.

$$\sigma_E = f(l_z^1,\ l_z^2, \vec{x})$$



$$\sigma_E = \sigma(E_{\text{pred}}/E_{\text{Gen}})$$

A key step towards automated optimization (gradient descent)

Deliverables & Budget

	FY 2022	FY 2023
a) Funds allocated	490	490
b) Actual costs to date	200	720

Timetable of Activities

Tasks and Deliverables	FY22				FY23				
	1	2	3	4	1	2	3	4	
T1: Implement G4 model of realistic calorimeter and develop training framework									✓
T2: Setup full learning pipeline using simplified setup									✓
D1: Write methods paper									✓
T3: Apply pipeline to EIC G4 simulation									✓
T4: Complete full detector/reconstruction codesign									WORK IN PROGRESS
D2: Write optimization paper									✓ WORK IN PROGRESS
D3: Deliver DNN fast sim and event reconstruction tools to EIC community									✓

Most of our goals (and more) completed.
We made foundational progress to our ultimate goal. Poised for delivering more.

Highlights

“We seek to exploit the opportunities provided by the EIC to achieve the first use of DNNs to design a detector, a milestone in the field.”

Delivered

DNN used in design optimization of forward calorimeter.
- Used to justify & motivate longitudinal granularity
- Inform key (million-dollar type) questions
such as impact of tungsten section.

1st of its kind for EIC

Highlights

“Other outcomes will include a DNN based fast-simulator with high fidelity, and DNN-based reconstruction software.
These will improve the detection capabilities of future EIC experiments,as well as cement the use of AI techniques from an early stage.”

Delivered

- AL/ML reconstruction of combined forward calorimeter system established and used in numerous studies.
  1st of its kind for EIC

- Diffusion generative model for HCAL and ZDC established.
  1st of its kind for EIC

Data and Code Sharing

- Simulations available in zenodo
  - https://doi.org/10.5281/zenodo.8128598
    https://doi.org/10.5281/zenodo.8384822

Dataset | Open

117
DOWNLOADS

- Code available in GitHub
  - https://github.com/eiccodesign/regressiononly
  - https://github.com/ftoralesacosta/GSGM_for_EIC_Cal
    https://github.com/ViniciusMikuni/Calo4EIC

eiccodesign / **regressiononly** Public

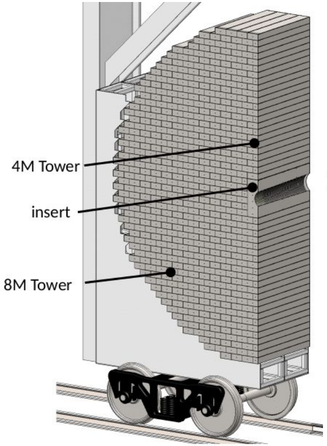
We have ensured that all our work is reproducible and have widely advertised our open datasets and code, enabling the EIC community to build upon our work, which was one of the goals of our project.

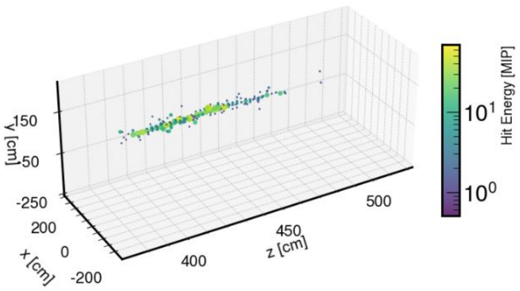
Highlights

"Our studies will be synergistic with the ongoing EIC R&D activities and will seek to inform them"

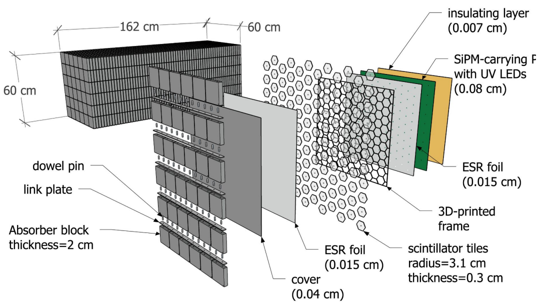
Delivered. *Our AI/ML studies have enhanced and influenced calorimetry R&D at EIC*

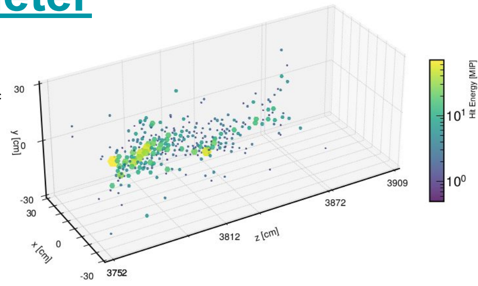
Forward HCAL



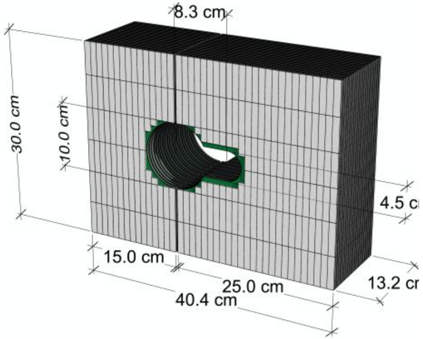


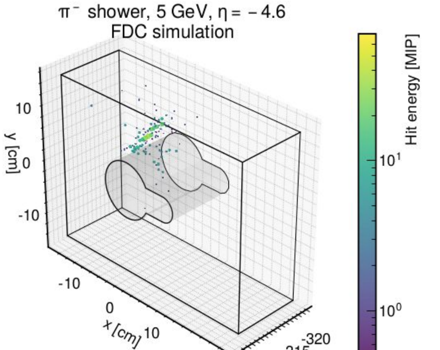
Zero Degree Calorimeter



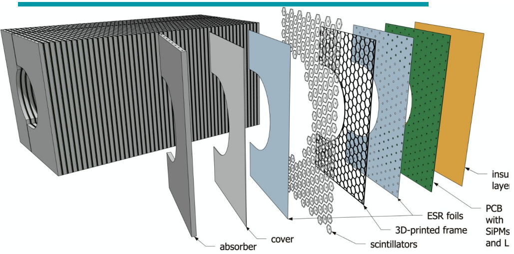


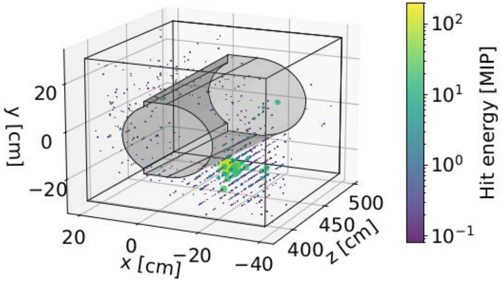
Few Degree Calorimeter





Calorimeter Insert





Highlights

“To help train the “AI workforce” of the future, we will engage students at the graduate and undergraduate levels”

Delivered

- 3 postdoc, 2 graduate students and 2 undergraduates were initiated into state-of-the-art AI/ML methodologies, along with the use of HPC resources at LLNL and NERSC

- Over the entire summer 2023, UCR team visited LLNL and LBNL: Ryan Milton, Sebastian Moran-Vasquez, JiaJun Huang, and Chase Owen <u>(50% of whom are first gen)</u>

- Inspired by their project experience, UCR graduates Ryan and Sebastian plan to take graduate-level AI/ML classes, integrating them into their thesis.

Summary

- Our project established a new collaboration between teams at LLNL, LBNL with expertise in AI/ML, and UC Riverside team with expertise in detector R&D.

- We have set the foundations of a framework for combining generative models and regression models for co-optimization, reaching milestones results along the way. Most of our milestones were completed.

- 2 journal articles submitted for publication + 2 conf. Proceedings. 7 conference presentations.

- Pivotal AI/ML developments for EIC calorimetry, already producing real-world impact.

- 3 postdoc, 2 graduate students, 2 undergraduate trained, all with limited previous experience, were trained on state-of-the-art AI/ML methods and utilization of HPC resources at LLNL and NERSC.