ZDC Simulation with ML

Wen-Chen Chang

2025/1/9

* Date: Jan 9 (Thursday), 2025

* Time: 11:00-12:00 AM at Taiwan (GMT+8)

* Zoom link:

https://cern.zoom.us/j/66342263280?pwd=DBemHUOnO6QIiQyU5y2WbeaEaBGcyT.1

Current Participants

郭家銘（中央）
林伯儒（中央）
呂昀儒（中研院）
周欣毅（中研院）
謝佳諭（中研院）
姚錫泓（中研院）
陳煒炫 Alan（卓越領航計畫、中研院）
林志勳（中研院）
章文箴（中研院）

CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation

https://arxiv.org/abs/2410.21611

Table 1: Models submitted to the CaloChallange.

Approach	Model	Code	Dataset				Section
			$1-\gamma$	$1-\pi$	2	3	
GAN	CaloShowerGAN [21]	[22]	✓	✓			3.1
	MDMA [23, 24]	[25]			✓	✓	3.2
	BoloGAN [26]	[27]	✓	✓			3.3
	DeepTree [28, 29]	[30]			✓		3.4
NF	L2LFlows [31, 32]	[33]			✓	✓	4.1
	CaloFlow [34, 35]	[36, 37]	✓	✓	✓	✓	4.2
	CaloINN [38]	[39]	✓	✓	✓		4.3
	SuperCalo [40]	[41]			✓		4.4
	CaloPointFlow [42]	[43]			✓	✓	4.5
Diffusion	CaloDiffusion [44]	[45]	✓	✓	✓	✓	5.1
	CaloClouds [46, 47]	[48, 49]				✓	5.2
	CaloScore [50, 51]	[52, 53]	✓		✓	✓	5.3
	CaloGraph [54]	[55]	✓	✓			5.4
	CaloDiT [56]	[57]			✓		5.5
VAE	Calo-VQ [58]	[59]	✓	✓	✓	✓	6.1
	CaloMan [60]	[61]	✓	✓			6.2
	DNNCaloSim [62, 63]	[64]		✓			6.3
	Geant4-Transformer [65]	[66]				✓	6.4
	CaloVAE+INN [38]	[39]	✓	✓	✓	✓	6.5
	CaloLatent [67]	[68]			✓		6.6
CFM	CaloDREAM [69]	[70]			✓	✓	7.1
	CaloForest [71]	[72]	✓	✓			7.2

Metrics (Sec. 8)

1. High-level Features (Histograms)
2. Pearson Correlation Coefficient (PCC)
3. Classifier-based Metrics
4. Computer Science-inspired Metrics
5. Manifold-based Metrics
6. Generation Timings
7. Memory Requirements

Classifier-based Metrics

Classifiers offer a way to perform a two-sample test that is sensitive to the full distribution, including correlations between features.

Here we focus on two different classifier tests. The first one, a binary classification task, compares each submission with the Geant4 test set. The second one, a multiclass classification task, compares all submissions with each other. For each, we consider two different neural network architectures.

Binary classification

The binary classification test evaluates how well the underlying distribution was learned and therefore how close the generated distribution is to the reference. It relies on the Neyman-Pearson Lemma, stating that the most powerful classifier to distinguish two samples is their likelihood ratio.

If a well-trained classifier is unable to distinguish submitted samples from the Geant4 test set, we conclude that the submission replicates the Geant4 distribution well. The result of this test, however, depends on the preprocessing that was applied to the data.

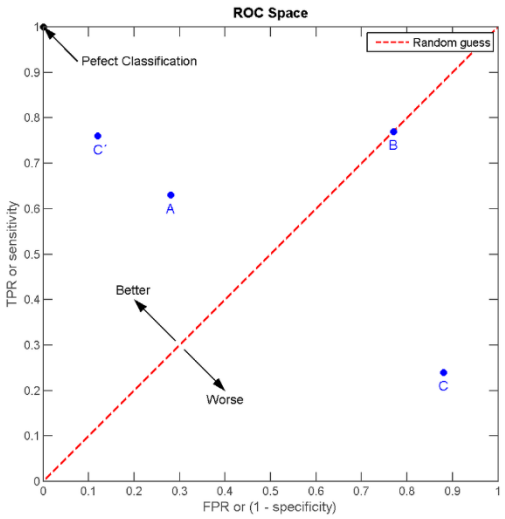
Binary classification

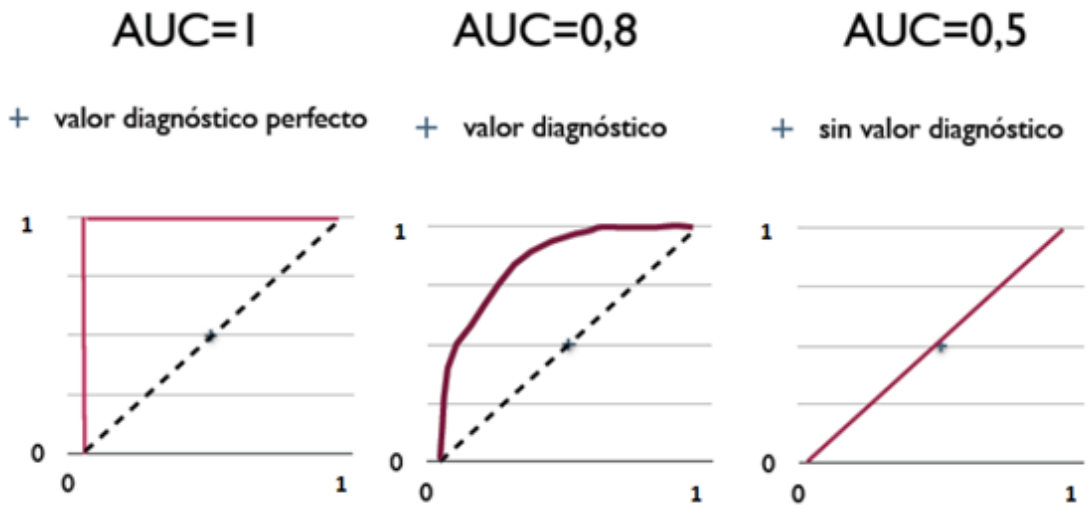
the submission replicates the GEANT4 distribution well [108, 185]. The result of this test, however, depends on the preprocessing that was applied to the data. Using the calorimeter showers in the physical space lets the classifier focus on the brightest voxels only, since energy depositions in them are orders of magnitude above the low-energy depositions. Applying a logarithm or logit transformation, enhances the sensitivity to mismodeling in them. While this gives a better understanding on whether or not the entire distribution was learned well, it might be that the difference is only in features and correlations that are irrelevant for the down-stream physics analysis. For that reason, we consider two different sets of input features. The first one are the energy depositions in the voxels (called "low-level" observables), the second one are the observables we introduced in Section 8.1 (called "high-level" observables).

- The energy deposition in each voxel: $\mathcal{I}_{ia}$.
- The energy depositions in each layer of the calorimeter, as the sum over all voxels in that layer: $E_i = \sum_a \mathcal{I}_{ia}$.
- The total energy deposition in the shower, as sum over all voxels, normalized to the incident energy: $E_{dep}/E_{inc} = \sum_{a,i} \mathcal{I}_{ia}/E_{inc}$.
- The centers of energy in η, ϕ, and r direction, defined via $\sum_a l_a \mathcal{I}_{ia} / \sum_a \mathcal{I}_{ia}$. The locations l_a are either $\phi_a = r_a \sin\alpha_a$, $\eta_a = r_a \cos\alpha_a$ or r_a, where r_a and α_a are the centers of the voxels in α and r. These are taken as the mean of the voxel boundary values defined in the `binning.xml` files. The sum goes over all voxels a in a given layer.
- The width of the center of energy distributions in η, ϕ, r direction: $\sqrt{\frac{\sum_a l_a^2 \mathcal{I}_{ia}}{\sum_a \mathcal{I}_{ia}} - \left(\frac{\sum_a l_a \mathcal{I}_{ia}}{\sum_a \mathcal{I}_{ia}}\right)^2}$
- The sparsity, defined as 1 minus the activity, with the activity being the fraction

Binary classification: Figure of merit AUC

The figure of merit in this setup is the AUC, the area under the receiver operating characteristic (ROC) curve. The ROC curve shows the true positive rate (TPR) as a function of the false positive rate (FPR). In a random classifier, the TPR will grow linearly with the FPR giving a AUC of 0.5. In a classifier that can separate the two datasets perfectly well, the ROC curve will become a step function, so the AUC becomes 1. We train ten classifiers with different random initialization and average the AUCs when reporting the results.





https://zh.wikipedia.org/zh-tw/ROC%E6%9B%B2%E7%BA%BF

Binary classification

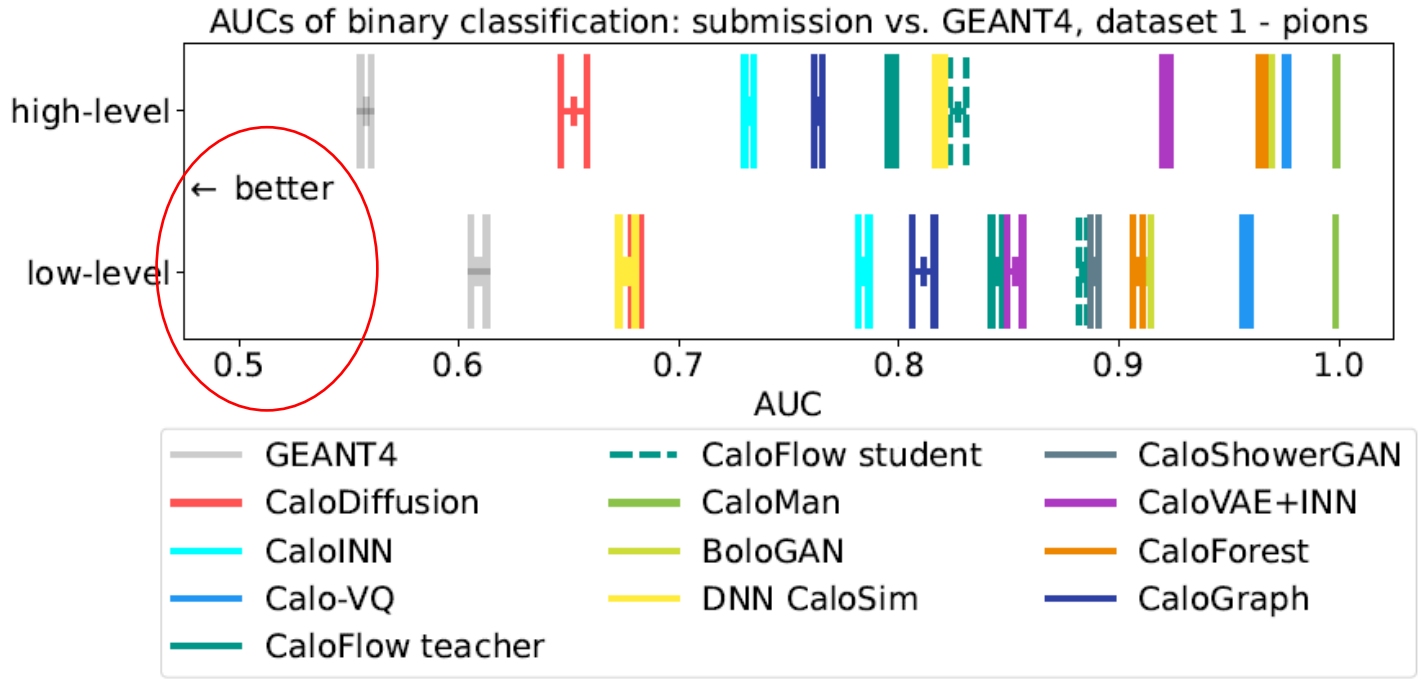


Figure 52: Low-level and high-level AUCs for evaluating GEANT4 vs. submission of ds $1-\pi^+$, averaged over 10 independent evaluation runs. For the precise numbers, see Table C8.

Moving on to classifier-based metrics, we find the AUCs of high- and low-level observables in figure 52 (and table C8). Here we observe several things. First, the AUC for separating the training and test Geant4 samples is larger than the expected value of 0.5. This is due to the fact that two slightly different versions of the ATLAS software were used due to technical problems in generating high statistics with the old version used for the ATLAS internal training. The differences were expected and deemed small enough to be irrelevant for physics applications. The AUC from the generative models will have this value as the maximum achievable separation instead of the usual 0.5.

Second, we see very low AUCs for CaloDiffusion, which was already indicated by the separation powers of the observables before.

Third, we see a low AUC for DNNCaloSim in the low-level observables which is, however, not reflected in the AUC of the high-level observables. This latter fact also correlates with the separation powers seen before. Other than that, we see overall good scores from diffusion and normalizing flow-based models, whereas GAN and VAE-based models show AUCs worse than 0.9.

Multiclass classification

With the multiclass classification setup, we try to assess which of the submissions is closest to Geant4. The method was introduced in [193] in the context of comparing hydrodynamical galaxy simulations, and subsequently applied to high-energy physics scenarios in [31, 194]. It relies on training a single classifier with cross entropy loss on the task "submission 1 vs. submission 2 vs. . . . vs. submission n".

When evaluating the trained classifier on a Geant4-based test set, we can read off which submission the Geant4 sample is closest to.

Multiclass classification

As figure of merit, we consider the average of the log posterior [193]. It is defined as

$$LP(\text{model } i|\text{samples } j) = \frac{1}{N}\sum_{x_k \in j} \log p_{\text{model } i}(x_k), \tag{41}$$

the average logarithm of the probability that samples j come from the model (submission) i. Here, the index k goes over all N samples in the set j. As a cross check of the quality of the trained multiclass classifier, we look at its performance in identifying the held-out test sets of each submission. A well-trained classifier will be able to distinguish the individual submissions from each other, so

$$LP(\text{model } i|\text{samples } j = i) > LP(\text{model } i|\text{samples } j \neq i). \tag{42}$$

arXiv:2412.12346

Feasibility Study of Measuring $\Lambda^0 \to n\pi^0$ Using a High-Granularity Zero-Degree Calorimeter at the Future Electron-Ion Collider

Sebouh J. Paul[a], Ryan Milton[a], Sebastián Morán[a], Barak Schmookler[a], Miguel Arratia[a,1]

[a]*Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*

Abstract

Key measurements at the future Electron-Ion Collider (EIC), including first-of-their-kind studies of kaon structure, require the detection of Λ^0 at forward angles. We present a feasibility study of $\Lambda^0 \to n\pi^0$ measurements using a high-granularity Zero Degree Calorimeter to be located about 35 m from the interaction point. We introduce a method to address the unprecedented challenge of identifying Λ^0s with energy $O(100)$ GeV that produce displaced vertices of $O(10)$ m. In addition, we present a reconstruction approach using graph neural networks. We find that the energy and angle resolution for Λ^0 is similar to that for neutrons, both of which meet the requirements outlined in the EIC Yellow Report. Furthermore, we estimate performance for measuring the neutron's direction in the Λ^0 rest frame, which reflects the Λ^0 spin polarization. We estimate that the neutral-decay channel $\Lambda^0 \to n\pi^0$ will greatly extend the measurable energy range for the charged-decay channel $\Lambda^0 \to p\pi^-$, which is limited by the location of small-angle trackers and the accelerator magnets. This work paves the way for EIC studies of kaon structure and spin phenomena.

HEXPLIT algorithm

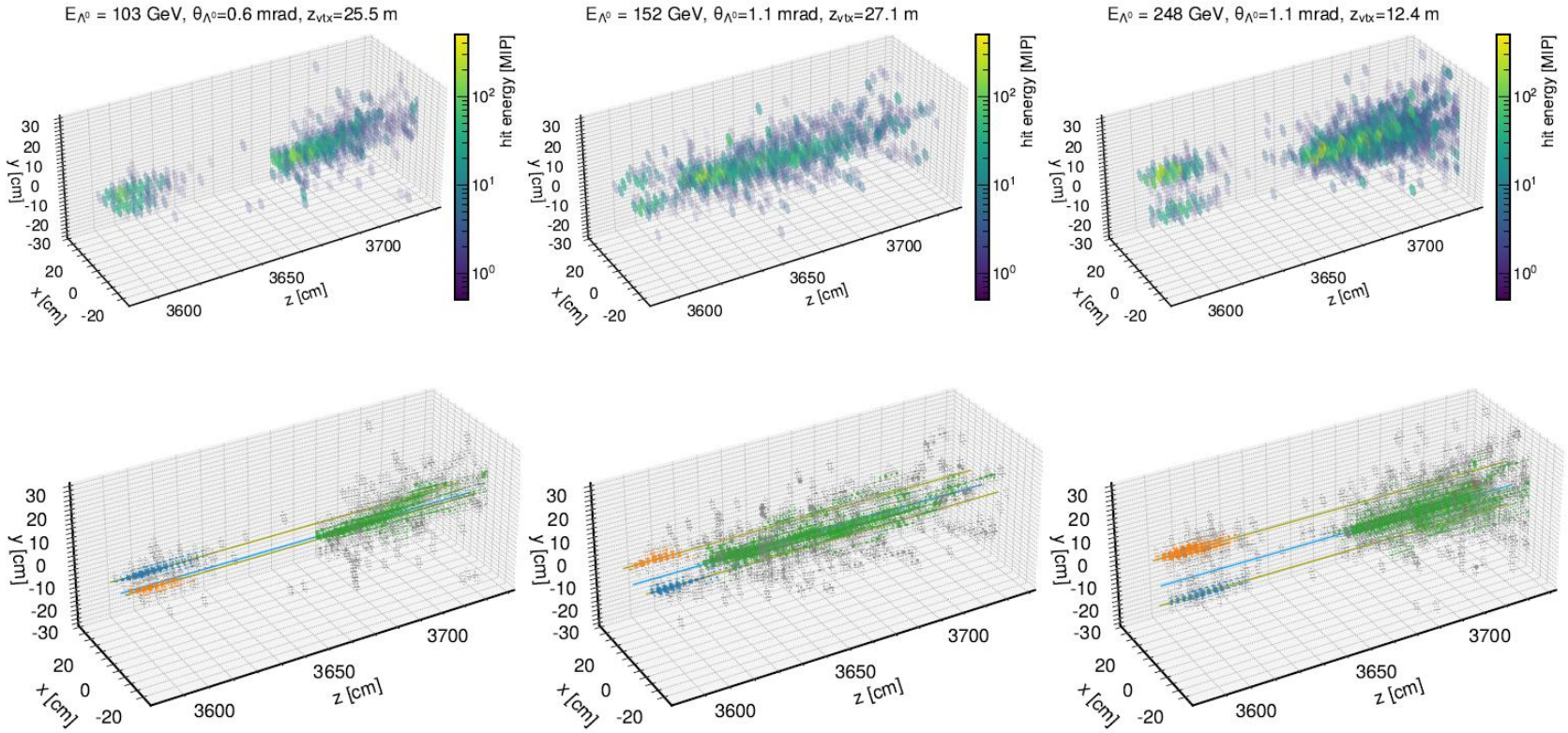


Figure 2: Top row: event display for example events, with the hits color coded by energy deposited, in units of MIPs. Bottom row: the same events, with subcell hits from HEXPLIT algorithm color-coded by which cluster the subcell hits have been assigned to by the 3D topological cluster described in the text.