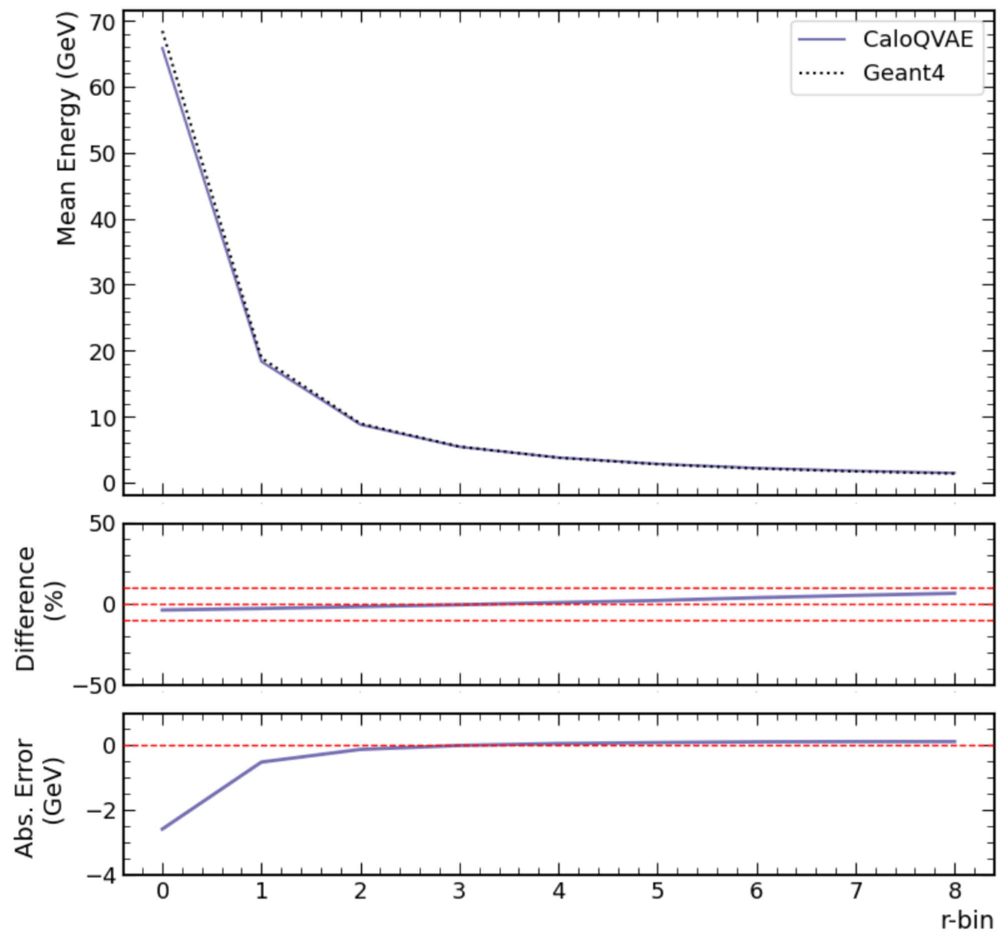
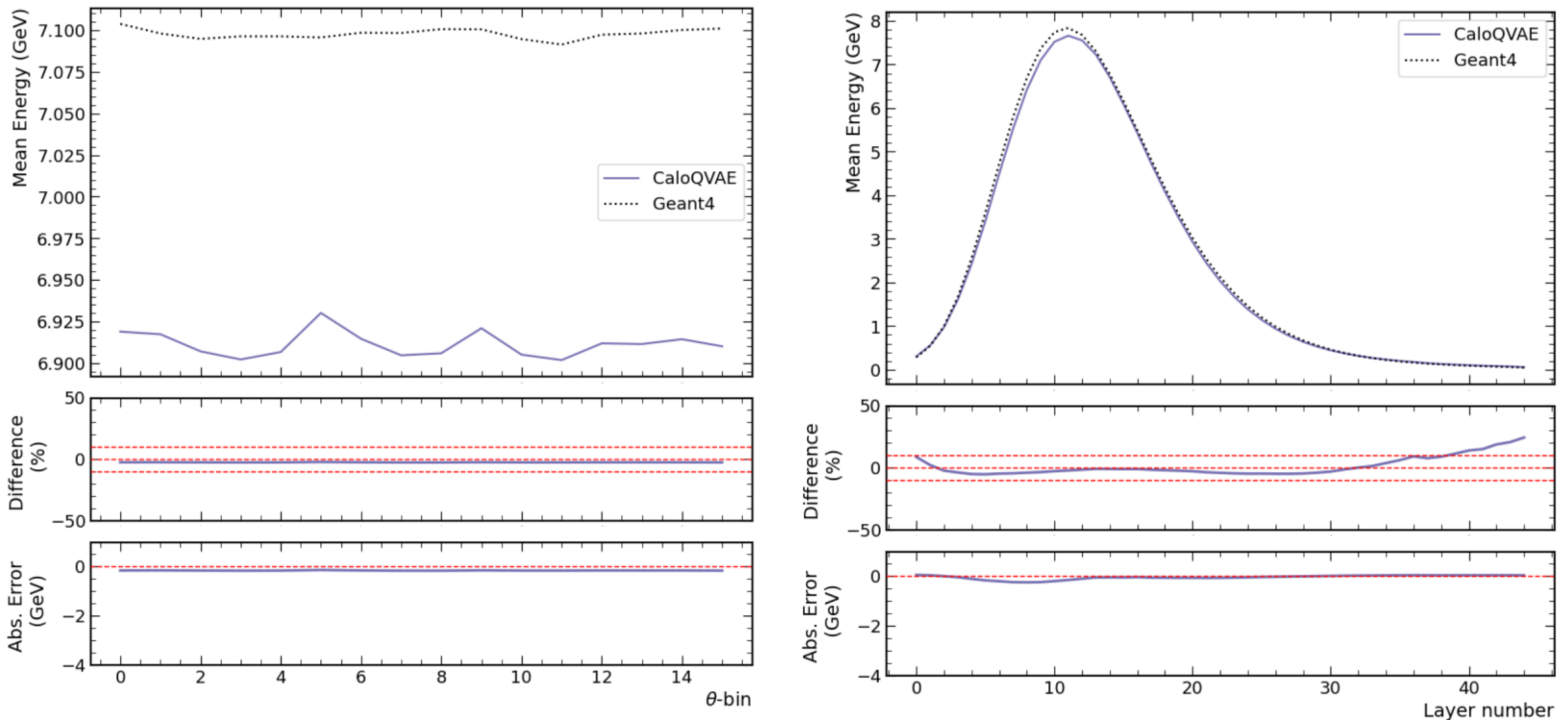
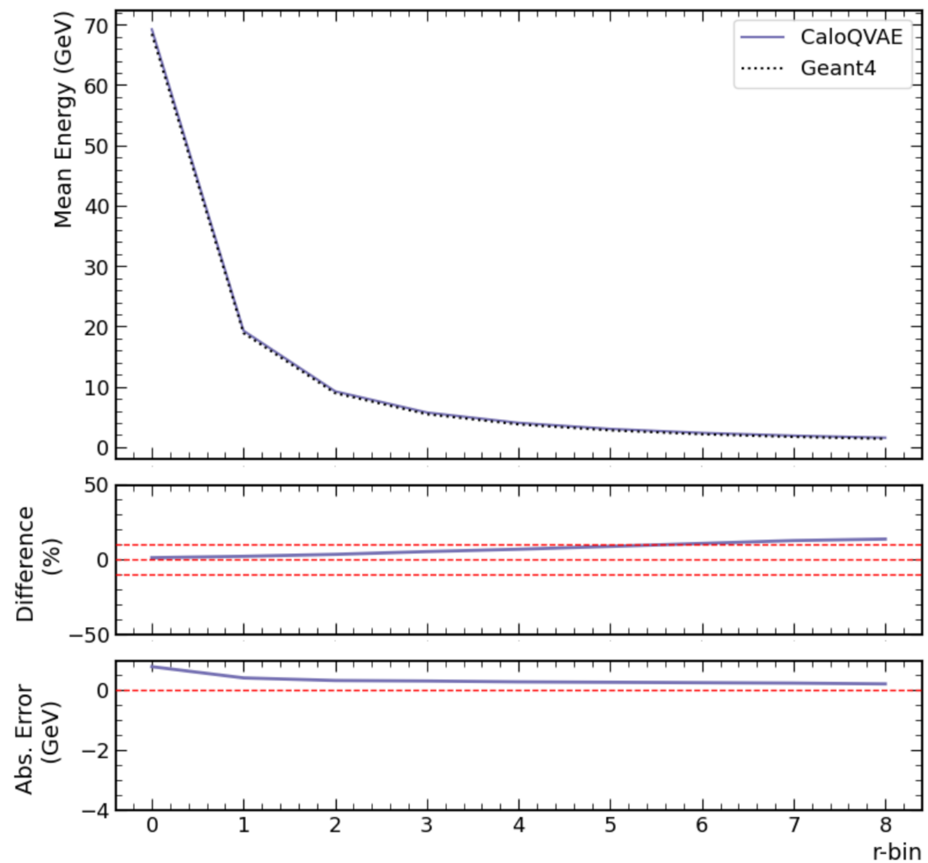
Synthetic_Uncond_Enc



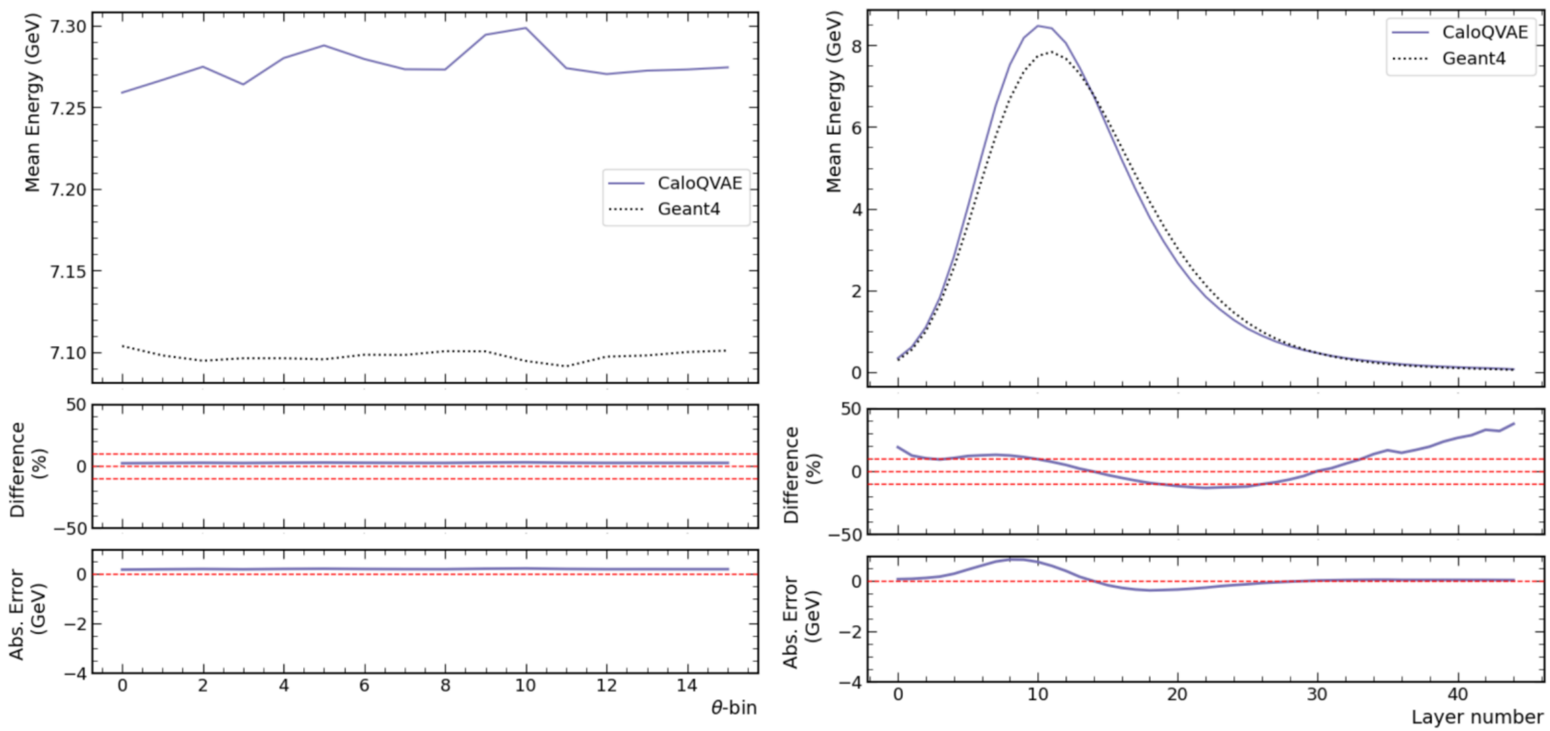
Synthetic_Uncond_Enc



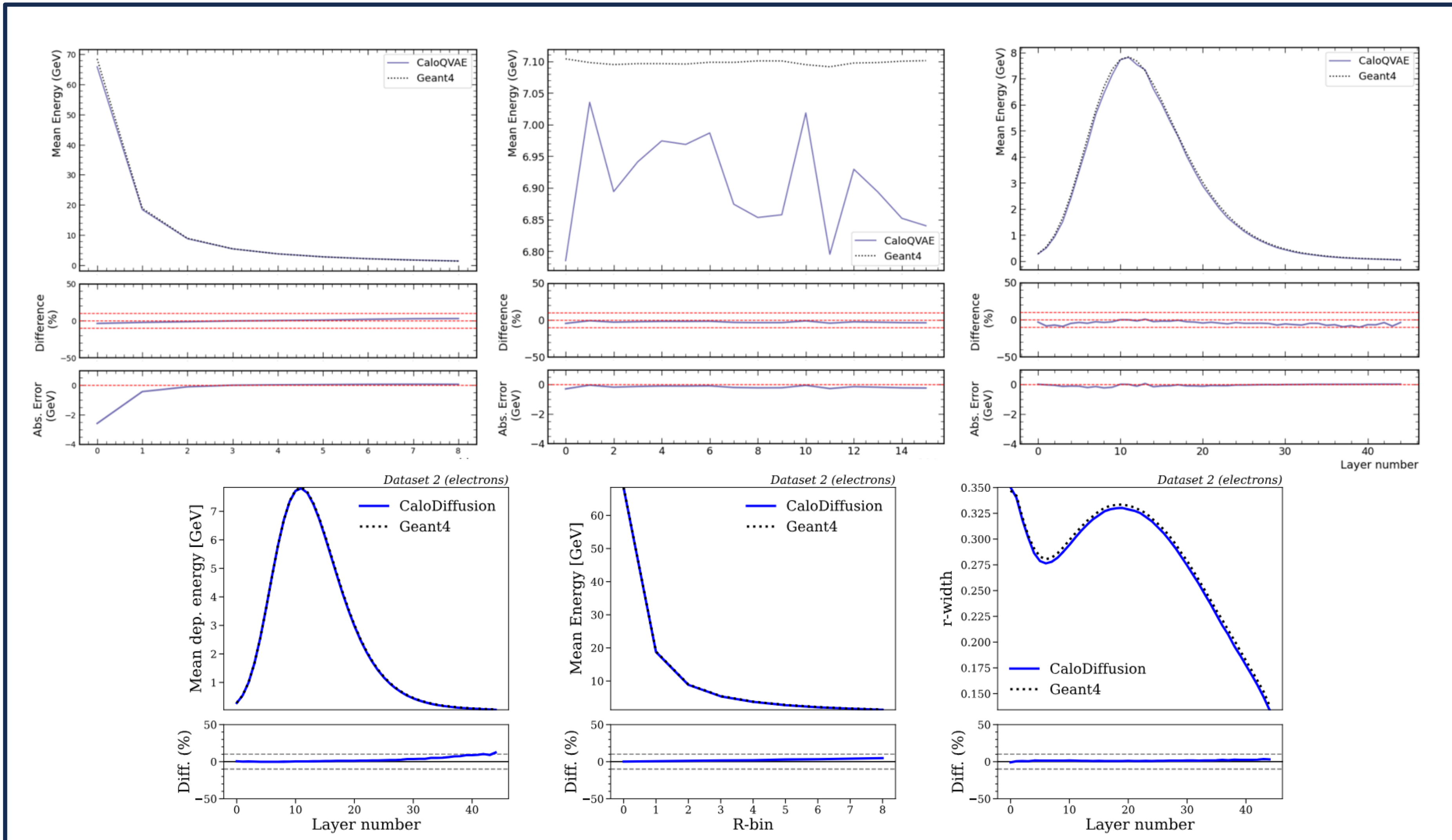
Model: Wise-Tree-1429



Model: Wise-Tree-1429

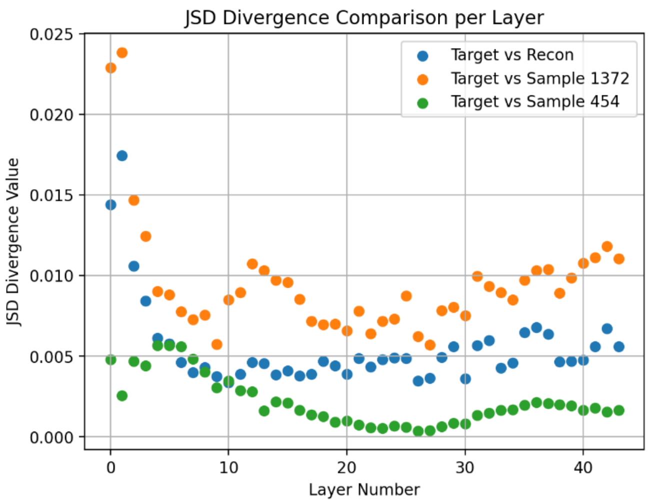


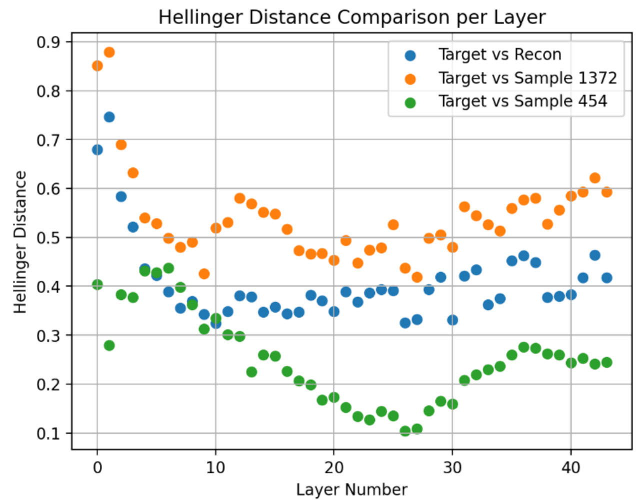
Model: Honest-Hill-454



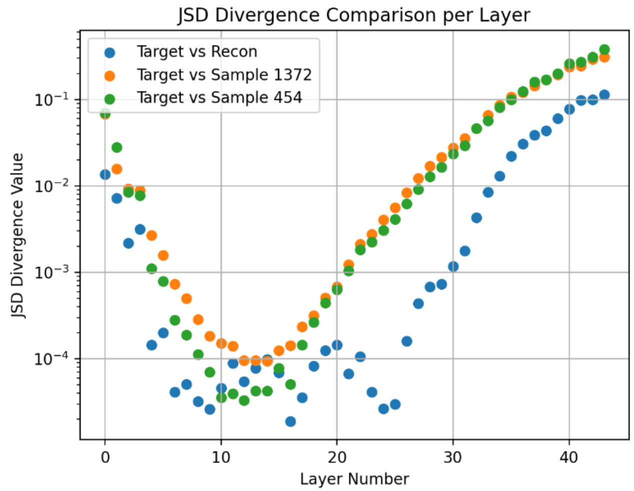
Model: Honest-Hill-454

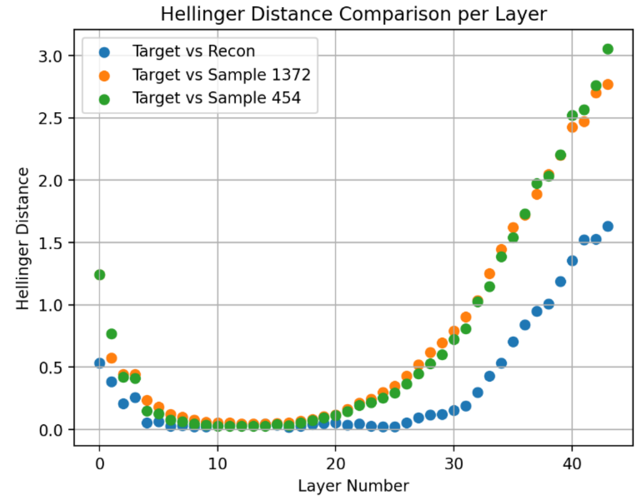
JSD





HLD





Evaluation Metrics for the Model

Evaluating generative models in high energy physics

Raghav Kansal,* Anni Li, and Javier Duarte
University of California San Diego, La Jolla, CA 92093, USA

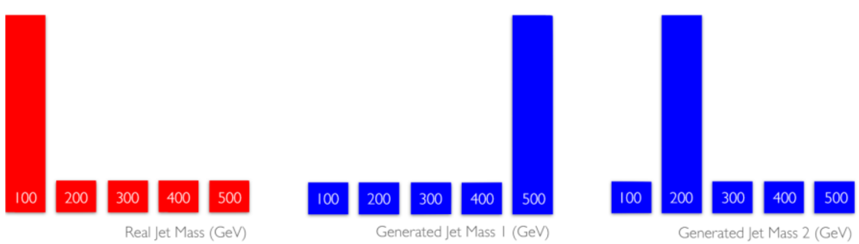
Nadezda Chernyavskaya, Maurizio Pierini
European Center for Nuclear Research (CERN), 1211 Geneva 23, Switzerland

Breno Orzari, Thiago Tomei
Universidade Estadual Paulista, São Paulo/SP, CEP 01049-010, Brazil

(Received 18 November 2022; Accepted 8 March 2023; Published 21 April 2023)

There has been a recent explosion in research into machine-learning-based generative modeling to tackle computational challenges for simulations in high energy physics (HEP). In order to use such alternative simulators in practice, we need well-defined metrics to compare different generative models and evaluate their discrepancy from the true distributions. We present the first systematic review and investigation into evaluation metrics and their sensitivity to failure modes of generative models, using the framework of two-sample goodness-of-fit testing, and their relevance and viability for HEP. Inspired by previous work in both physics and computer vision, we propose two new metrics, the Fréchet and kernel physics distances (FPD and KPD, respectively), and perform a variety of experiments measuring their performance on simple Gaussian-distributed, and simulated high energy jet datasets. We find FPD, in particular, to be the most sensitive metric to all alternative jet distributions tested and recommend its adoption, along with the KPD and Wasserstein distances between individual feature distributions, for evaluating generative models in HEP. We finally demonstrate the efficacy of these proposed metrics in evaluating and comparing a novel attention-based generative adversarial particle transformer to the state-of-the-art message-passing generative adversarial network jet simulation model. The code for our proposed metrics is provided in the open source JetNet Python library.

DOI: 10.1103/PhysRevD.107.076017



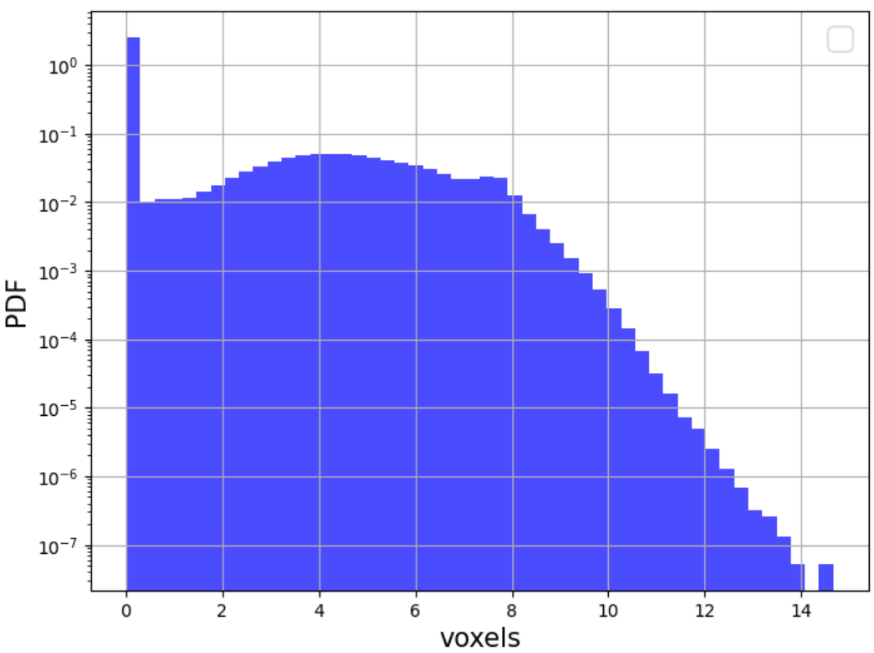
CaloDiffusion Metrics

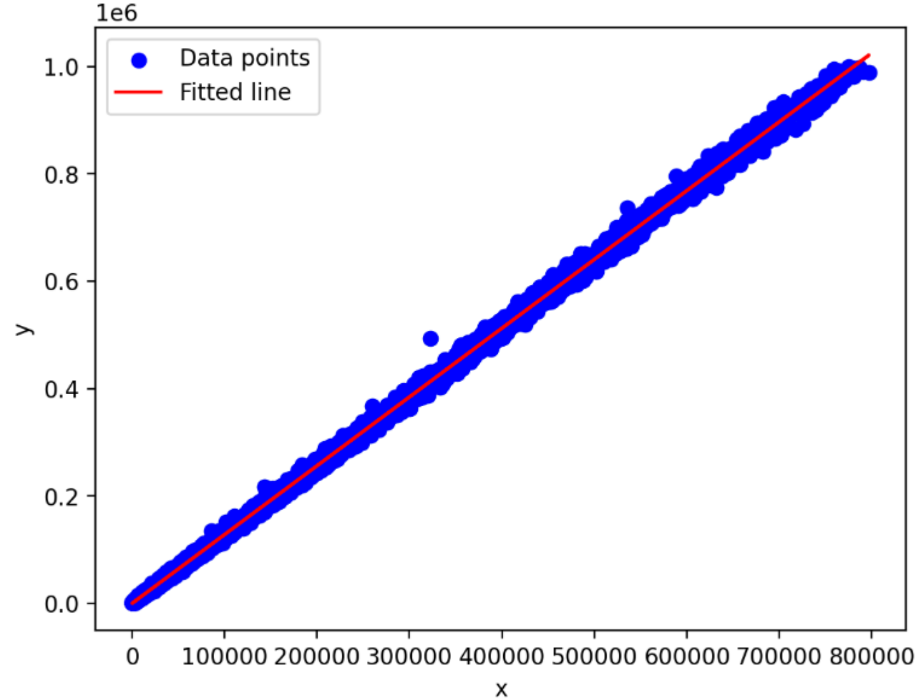
Dataset	Classifier AUC (low / high) `CaloDiffusion`	`CaloFlow`	`CaloScore v2`
1 (photons)	**0.62** / 0.62	0.70 / **0.55**	0.76 / 0.59
1 (pions)	**0.65** / **0.65**	0.78 / 0.70	- / -
2 (electrons)	**0.56** / **0.56**	0.80 / 0.80	0.60 / 0.62
3 (electrons)	**0.56** / **0.57**	0.91 / 0.95	0.67 / 0.85

Dataset	FPD	KPD
1 (photons)	0.014(1)	0.004(1)
1 (pions)	0.029(1)	0.004(1)
2 (electrons)	0.043(2)	0.0001(2)
3 (electrons)	0.031(2)	0.0001(1)

TABLE II. Additional metrics comparing the agreement between showers generated with `Geant4` and `CaloDiffusion`. The number in parentheses is the uncertainty in the last significant digit as evaluated with the `JETNET` library.

Incident Energy Mapping





$$\left(\frac{\log(E_{inc}) \ - \log(\min(E_{inc}))}{\log(\max(E_{inc})) \ - \log(\min(E_{inc}))}\right) * (\max(u_i) \ * m(\frac{E_{inc}}{\sum E_i}))$$

```
Slope: 1.2812657528661318
Intercept: 230.97830116318073
R-squared: 0.9994513756687647
P-value: 0.0
Standard error: 0.00030021981220185786
```

Partition Connections Graph

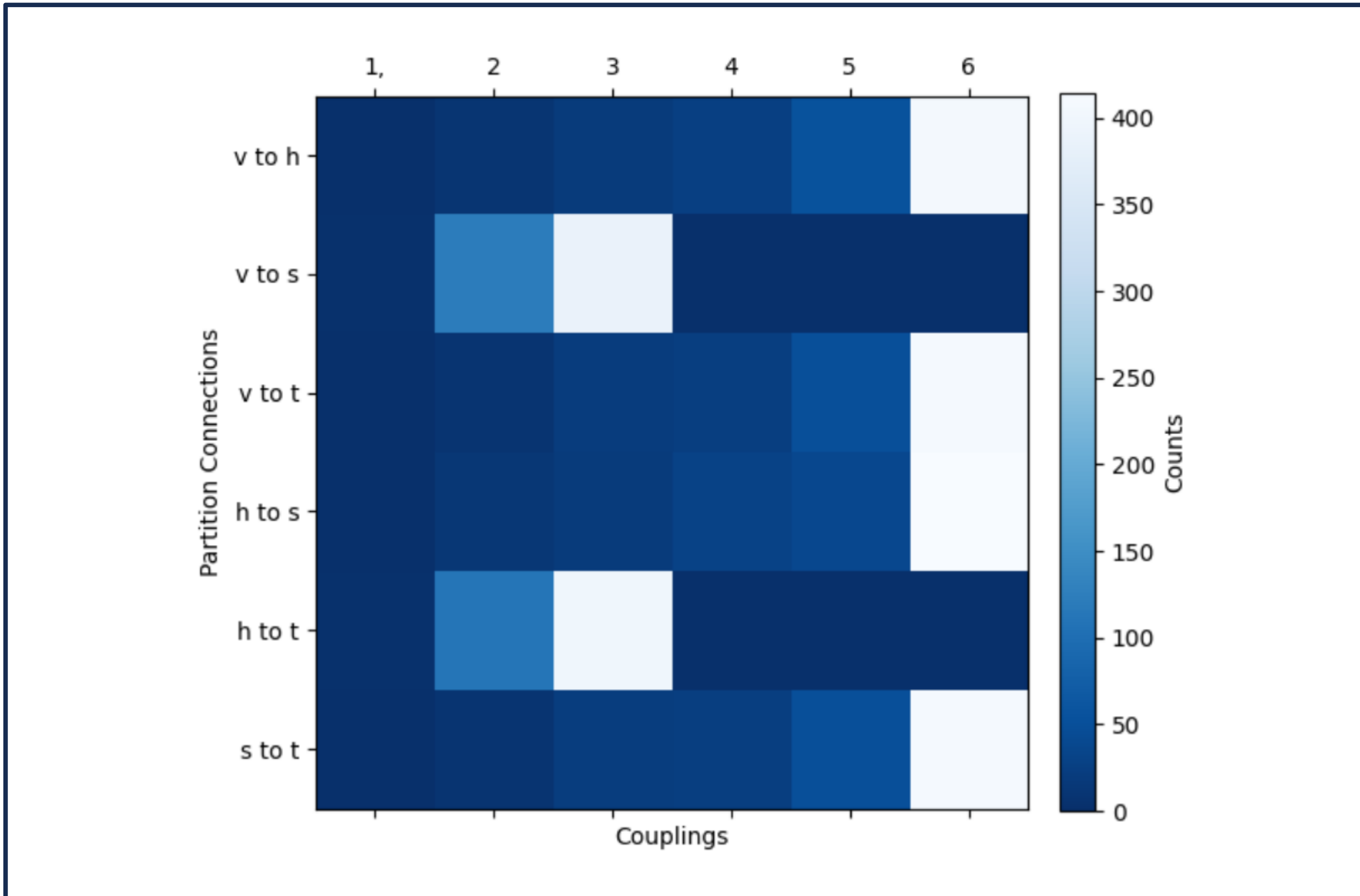
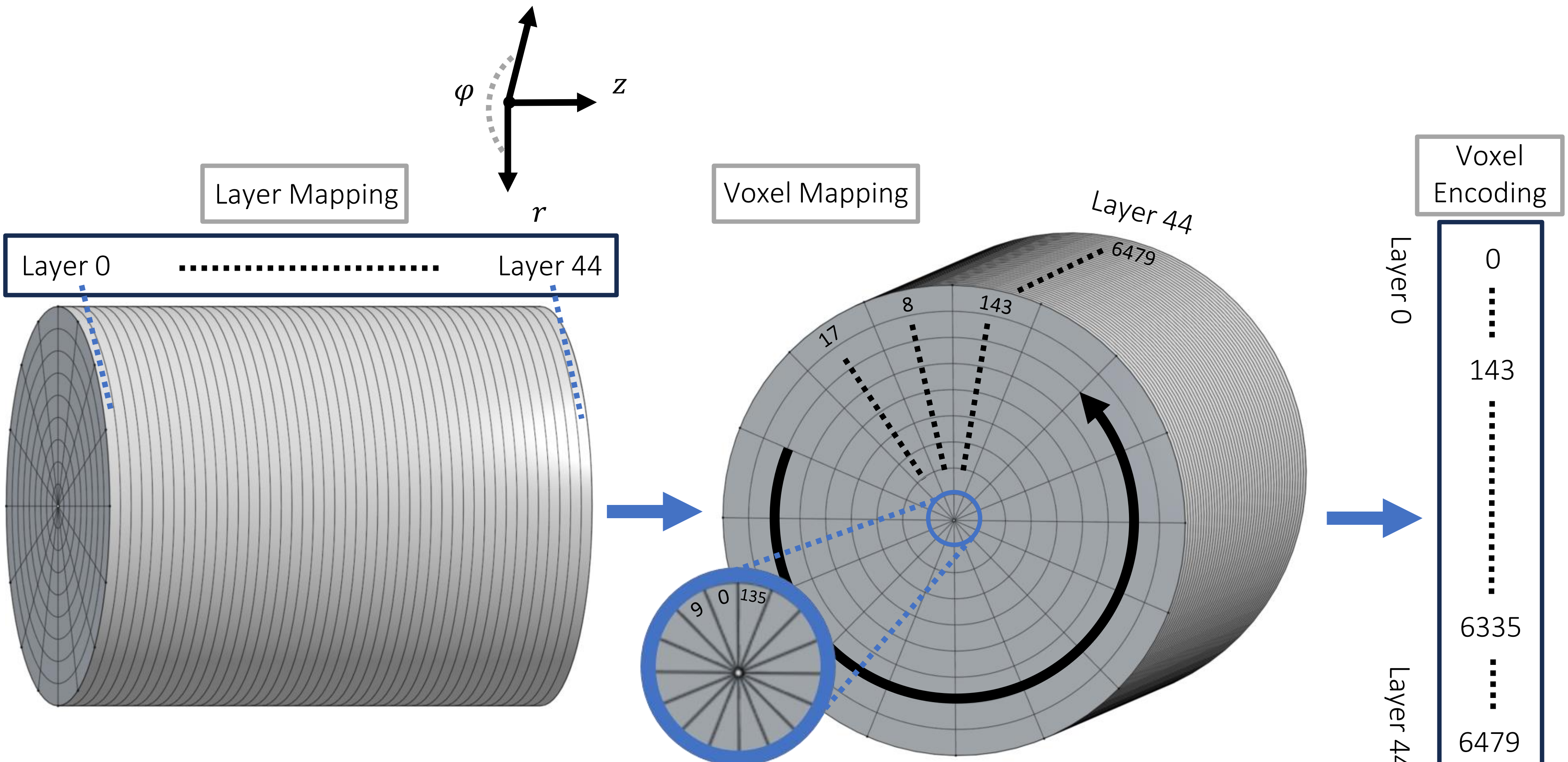


Diagram CAD Models of Calorimeter Cylinders



(a)