The Fast Calorimeter Simulation Challenge 2022

— ML4Jets at DESY Hamburg, Germany —

Claudius Krause

Institute of High Energy Physics (HEPHY), Austrian Academy of Sciences (OeAW)

November 9, 2023

with Michele Faucci Giannelli, Gregor Kasieczka, Ben Nachman, Dalila Salamani, David Shih, and Anna Zaborowska

`https://calochallenge.github.io/homepage/`

The Fast Calorimeter Challenge 2022

In February 2022, we introduced 4 different calorimeter datasets to

⇒ trigger development of new generative models.

⇒ evaluate existing models on common datasets.

⇒ improve our understanding of common struggles, advantages, disadvantages, and scaling behavior.

ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

The Fast Calorimeter Challenge 2022

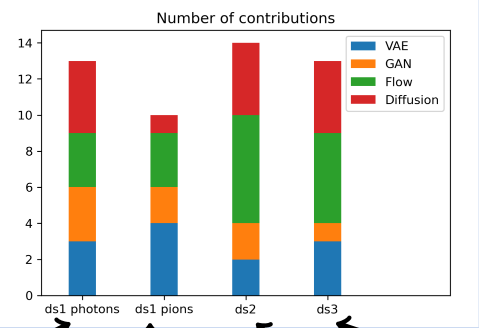
In February 2022, we introduced 4 different calorimeter datasets to

- ⇒ trigger development of new generative models.
- ⇒ evaluate existing models on common datasets.
- ⇒ improve our understanding of common struggles, advantages, disadvantages, and scaling behavior.

May 30th, 2023: CaloChallenge Workshop in Frascati, IT
https://agenda.infn.it/event/34036/

A total of 50 samples have been submitted

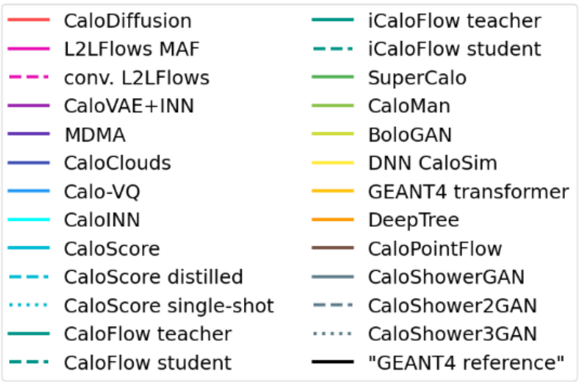
- 50 samples have been submitted.
- Nicely distributed over different DGMs.



368-dim. 533-dim. 6480-dim. 40500-dim.

Many models are discussed this week!

K. Pedro, Tue 2.30pm →
T. Buss, Mon 11am →
F. Ernst, Mon 11.15am →
B. Käch, Mon 11.30am →
E. Buhmann, Tue 3pm →
Q. Liu, Mon 11.45am →
F. Ernst, Mon 11.15am →
V. Mikuni ML4Jets 22 →
V. Mikuni, Tue 2.15pm →
I. Pang ML4Jets 22 →



← I. Pang ML4Jets 22
← I. Pang, Mon 2.30pm
← J. Cresswell ML4Jets 22
← M. Scham, Tue 12.15pm
← S. Schnake, Tue 11.45am
← M. Faucci-Giannelli ML4Jets 22

The main evaluations presented today*

⇒ The *separation power* of high-level feature histograms:

$$S(h_1, h_2) = \frac{1}{2} \sum_{i=1}^{n_{\text{bins}}} \frac{(h_{1,i} - h_{2,i})^2}{h_{1,i} + h_{2,i}}$$

Diefenbacher et al. [2009.03796, JINST]

⇒ A multi-class classifier based on voxels:
Train on submission 1 vs. submission 2 vs. ... vs. submission n
and evaluate the *log posterior*:

$$L = \langle \log \left(p(x_{\in \text{class } i} | x_{\text{taken from } j}) \right) \rangle \qquad j \in \{\text{submission } k, \text{GEANT4}\}$$

⇒ The *generation time*.

* There'll be more in the final document.

The main evaluations presented today*

⇒ The *separation power* of high-level feat

$$S(h_1, h_2) = \frac{1}{2} \sum_{i=1}^{n_{\text{bins}}} \frac{(h_{1,i} - h_{2,i})^2}{h_{1,i} + h_{2,i}}$$

⇒ A multi-class classifier
Train on submission 1 vs. submission
and evaluate the *log posterior*:

$$L = \langle \log \left(p(x_{\in \text{class } i} | x_{\text{taken from } j}) \right) \rangle$$

$j \in \{$ $, \text{GEANT4}\}$

- Simple DNN with 2 hidden layer of 2048 neurons*.
- features*: $log_{10} E_{inc}, \mathcal{I}_a / E_{inc}$
- Cross Entropy loss,
  ADAM optimizer,
  25 epochs (val loss min around 15)

⇒ The *generation time*.

* There'll be more in the final document.

The main evaluations presented today*

$\Rightarrow$ The *separation power* of high-level feature histograms:

$$S(h_1, h_2) = \frac{1}{2} \sum_{i=1}^{n_{\text{bins}}} \frac{(h_{1,i} - h_{2,i})^2}{h_{1,i} + h_{2,i}}$$

Diefenbacher et al. [2009.03796, JINST]

$\Rightarrow$ A multi-class classifier based on voxels:
Train on submission 1 vs. submission 2 vs.
and evaluate the *log posterior*:

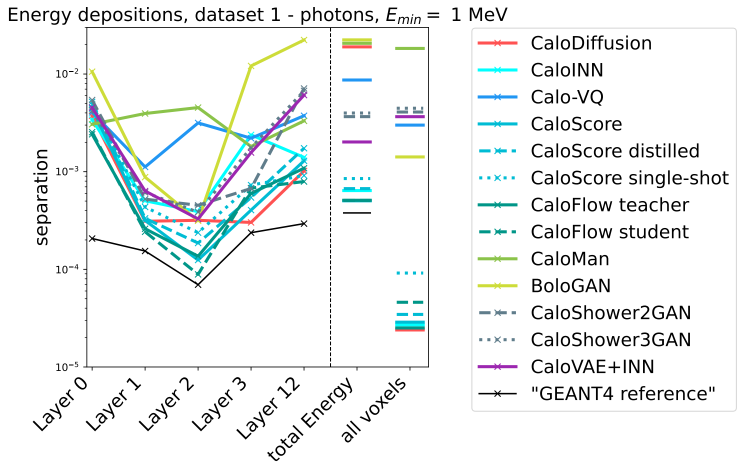
$$L = \langle \log \left(p(x_{\in \text{class } i} | x_{\text{taken from } j}) \right) \rangle$$

$\Rightarrow$ The *generation time*.

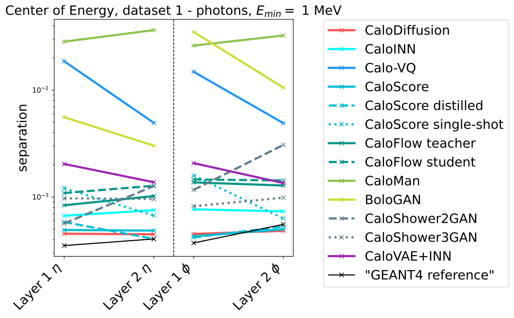
- start singularity container
- load model weights + biases
- generate samples
- save them to `.hdf5`

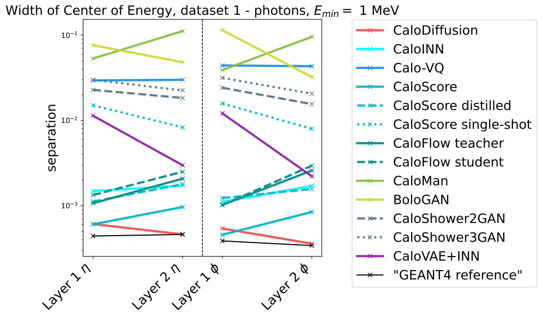
* There'll be more in the final document.

Histogram separation power ds1 photons

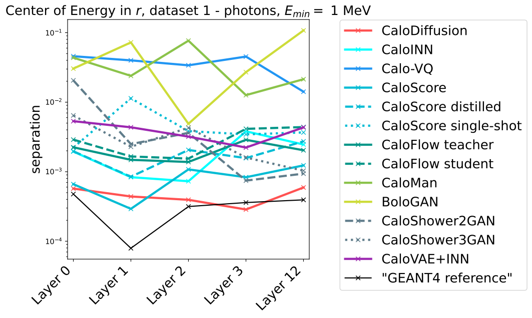


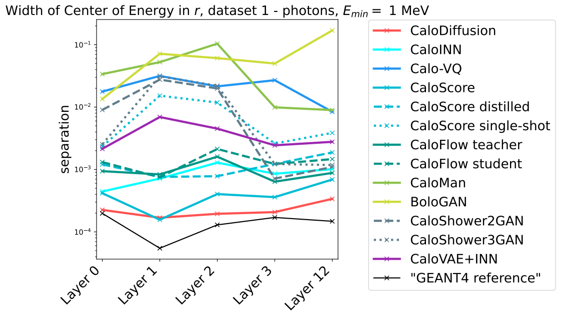
Histogram separation power ds1 photons



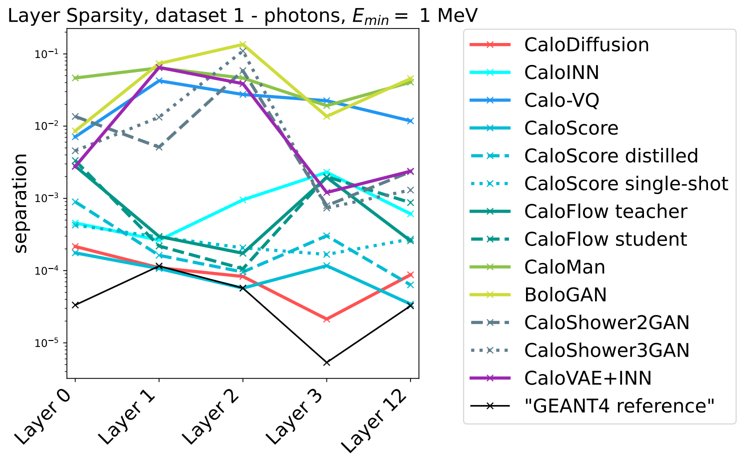


Histogram separation power ds1 photons





Histogram separation power ds1 photons

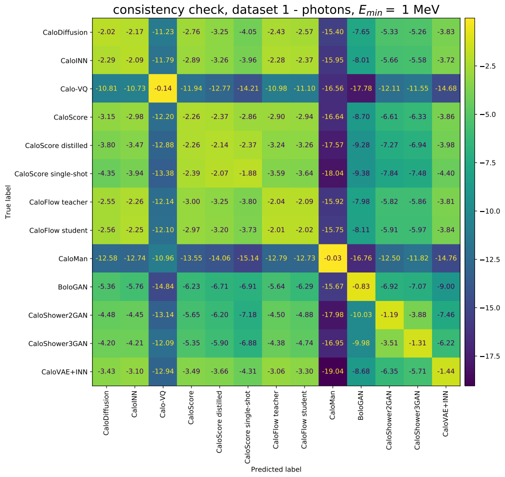


Log posterior ds1 photons

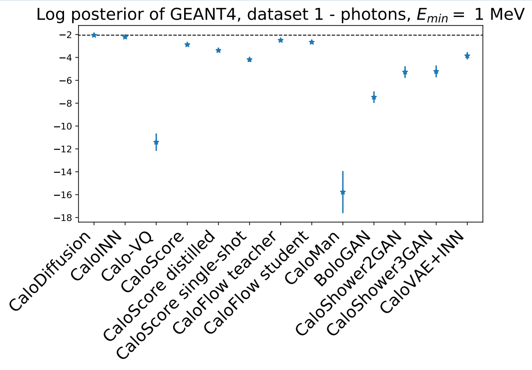
- submission vs submission:
  what we see: $\langle \log p(\text{model}|data) \rangle$

$\Rightarrow$ each sample is correctly identified (diagonal is largest entry per row)

(mean of 10 independent runs.)



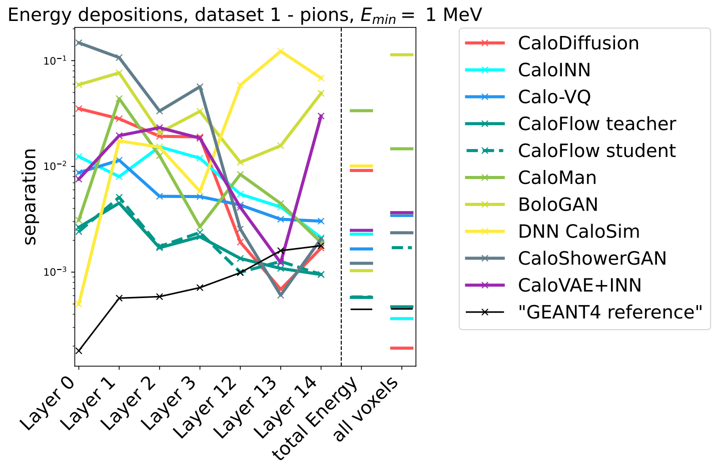
Log posterior ds1 photons



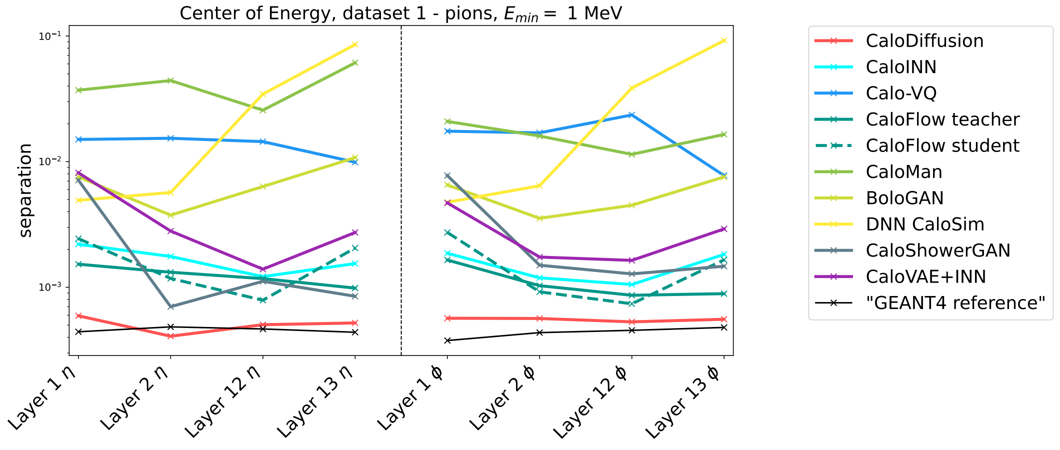
(based on 10 independent runs.)

CaloDiffusion	-2.0506 $\pm$ 0.0637	1.
CaloINN	-2.2104 $\pm$ 0.0707	2.
Calo-VQ	-11.4103 $\pm$ 0.7629	12.
CaloScore	-2.8781 $\pm$ 0.1104	5.
CaloScore distilled	-3.3831 $\pm$ 0.1476	6.
CaloScore single-shot	-4.1859 $\pm$ 0.1956	8.
CaloFlow teacher	-2.4987 $\pm$ 0.1318	3.
CaloFlow student	-2.6619 $\pm$ 0.1431	4.
CaloMan	-15.7763 $\pm$ 1.8428	13.
BoloGAN	-7.4741 $\pm$ 0.5061	11.
CaloShower2GAN	-5.2747 $\pm$ 0.5135	10.
CaloShower3GAN	-5.2186 $\pm$ 0.5214	9.
CaloVAE+INN	-3.8530 $\pm$ 0.3185	7.

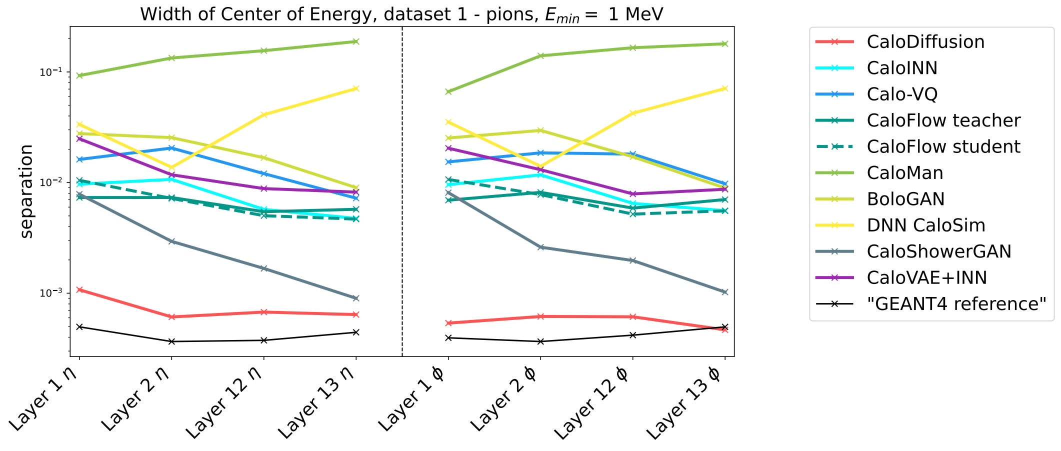
Histogram separation power ds1 pions



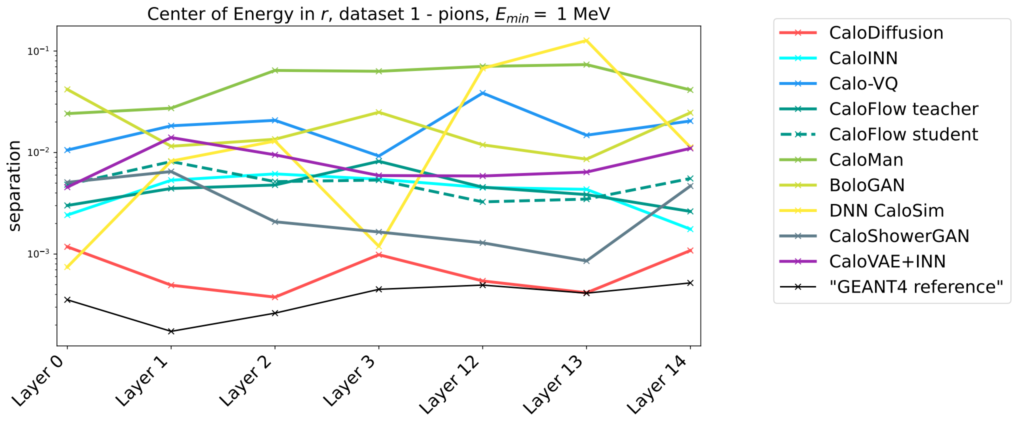
Histogram separation power ds1 pions



Histogram separation power ds1 pions



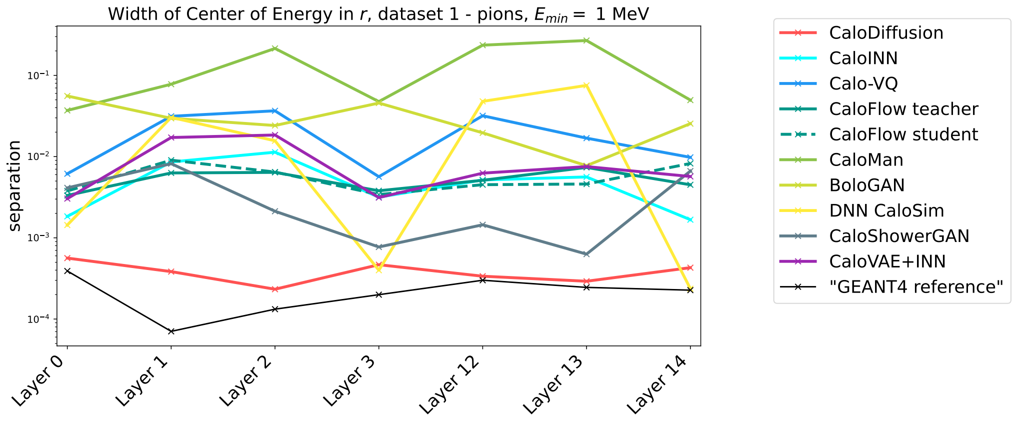
Histogram separation power ds1 pions



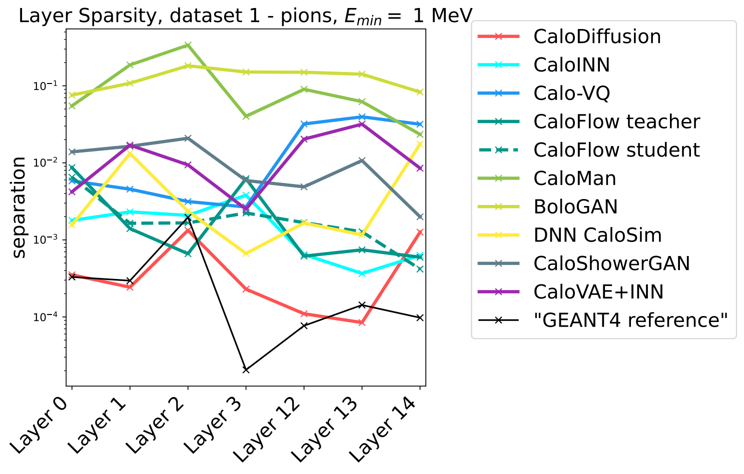
ÖAW AUSTRIAN ACADEMY OF SCIENCES

HEPHY INSTITUTE OF HIGH ENERGY PHYSICS

Histogram separation power ds1 pions



Histogram separation power ds1 pions

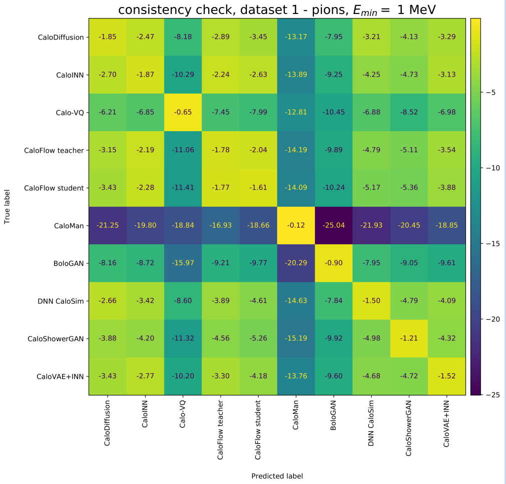


Log posterior ds1 pions

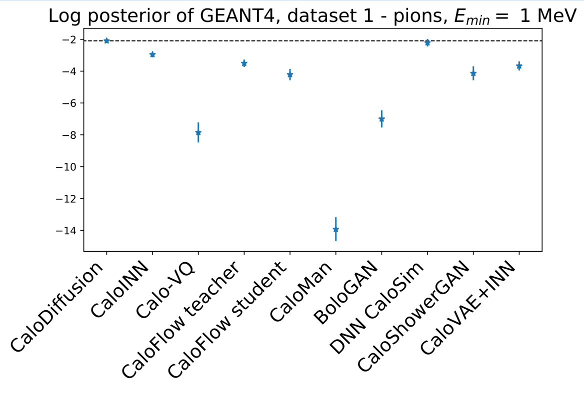
- submission vs submission:
  what we see: $\langle \log p(\text{model}|\mathit{data})\rangle$

$\Rightarrow$ each sample is correctly identified (diagonal is largest entry per row)

(mean of 10 independent runs.)



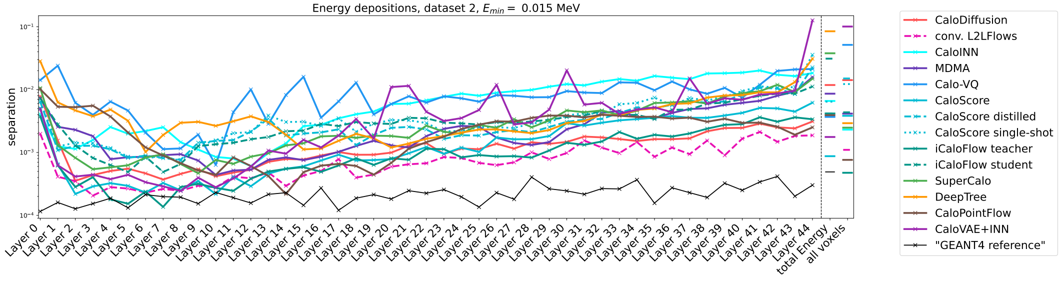
Log posterior ds1 pions



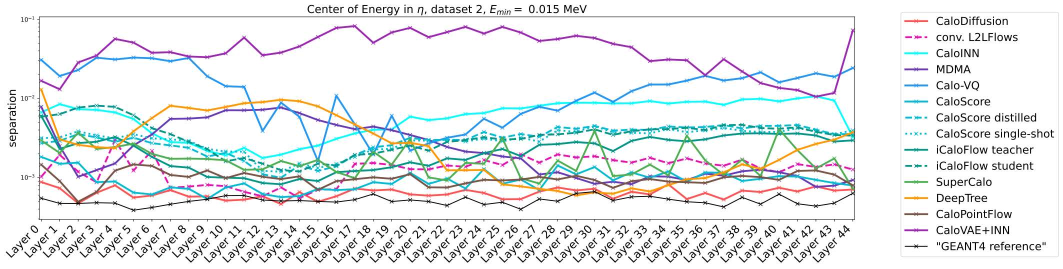
CaloDiffusion	-2.0939 $\pm$ 0.1443	1.
CaloINN	-2.9518 $\pm$ 0.1768	3.
Calo-VQ	-7.8493 $\pm$ 0.6290	9.
CaloFlow teacher	-3.5017 $\pm$ 0.2313	4.
CaloFlow student	-4.2131 $\pm$ 0.3583	7.
CaloMan	-13.9270 $\pm$ 0.7541	10.
BoloGAN	-6.9995 $\pm$ 0.5386	8.
DNN CaloSim	-2.2117 $\pm$ 0.2388	2.
CaloShowerGAN	-4.1314 $\pm$ 0.4400	6.
CaloVAE+INN	-3.6720 $\pm$ 0.2886	5.

(based on 10 independent runs.)

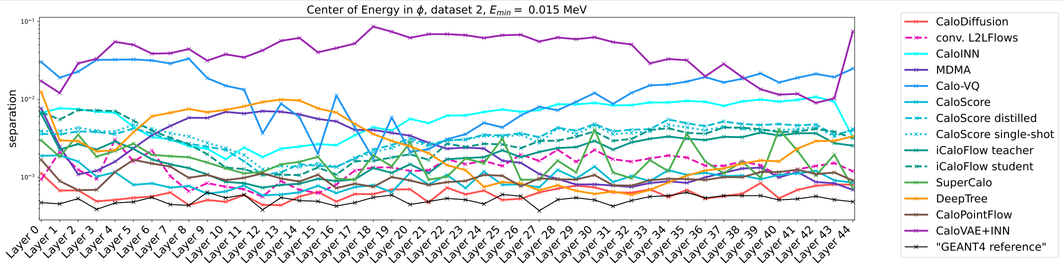
Histogram separation power ds2



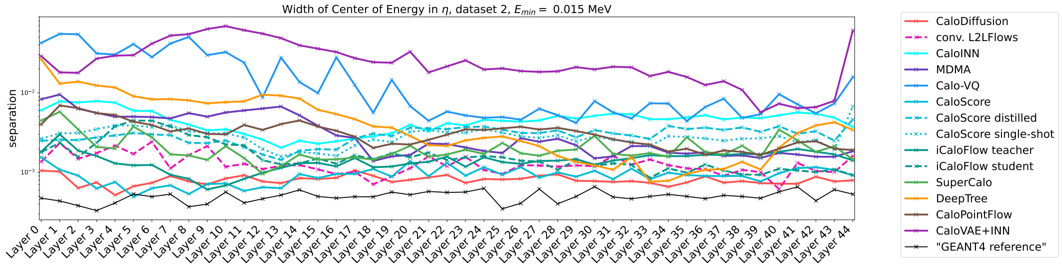
Histogram separation power ds2



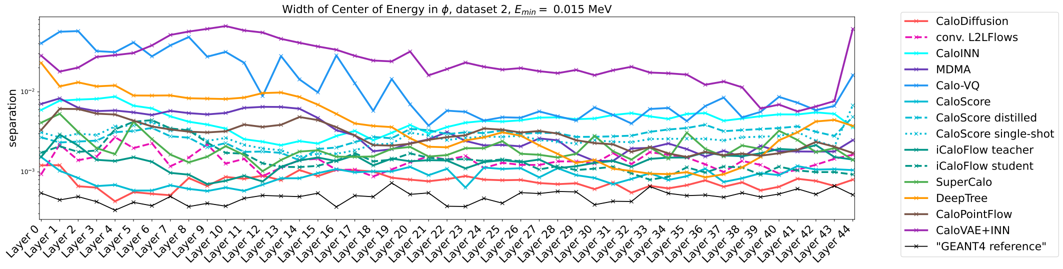
Histogram separation power ds2



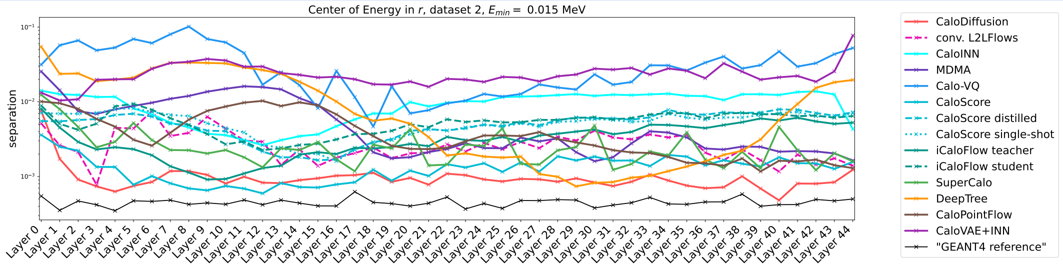
Histogram separation power ds2



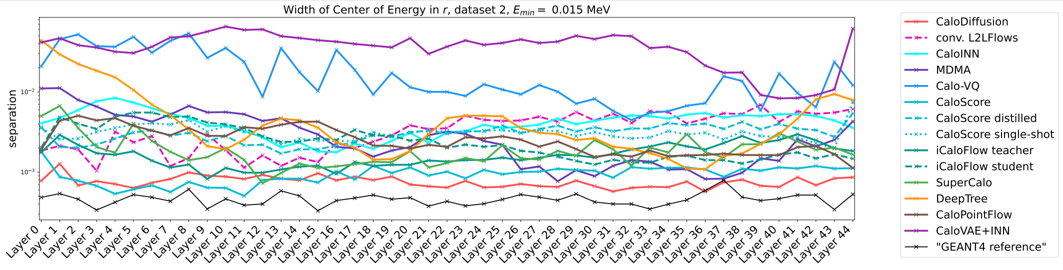
Histogram separation power ds2



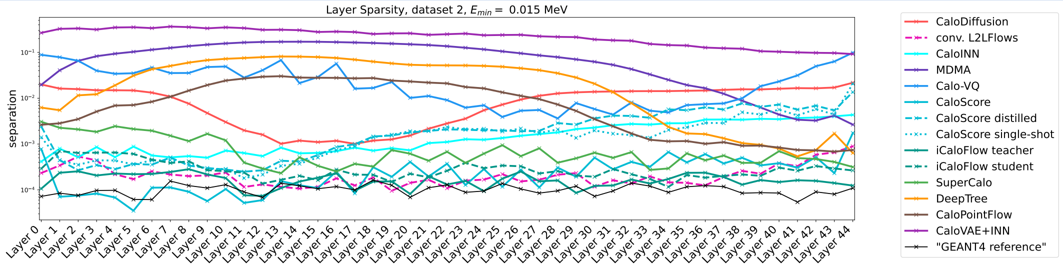
Histogram separation power ds2



Histogram separation power ds2



Histogram separation power ds2

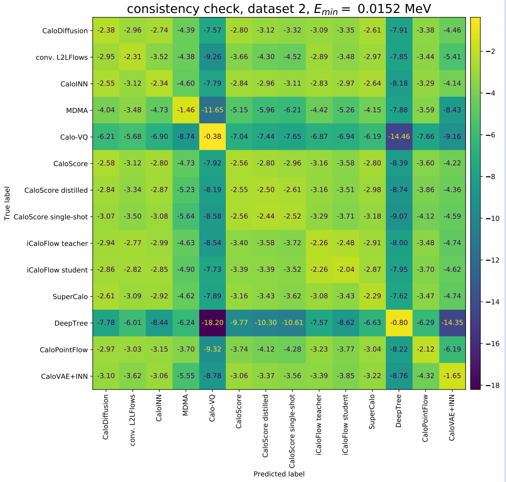


Log posterior ds2

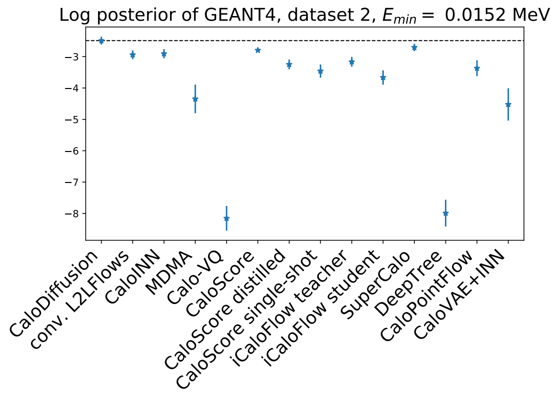
- submission vs submission:
  what we see: $\langle \log p(\text{model}|\mathit{data}) \rangle$

$\Rightarrow$ each sample is correctly identified
(diagonal is largest entry per row)

(mean of 10 independent runs.)



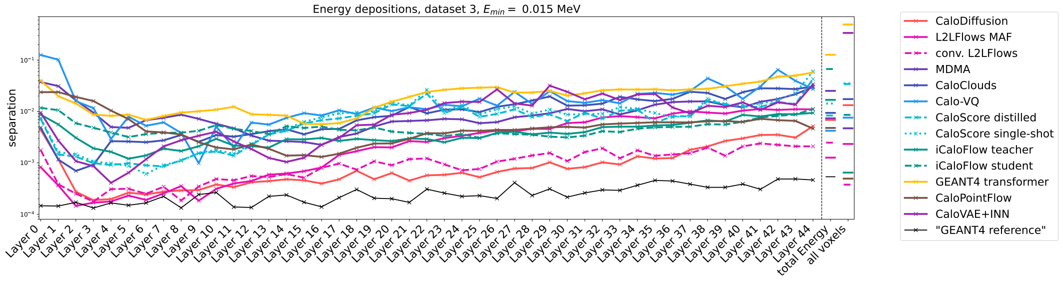
Log posterior ds2



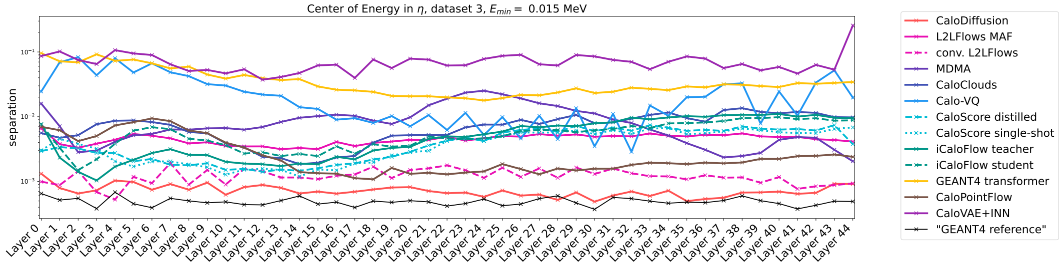
(based on 10 independent runs.)

CaloDiffusion	-2.4895 $\pm$ 0.1255	1.
conv. L2LFlows	-2.9404 $\pm$ 0.1424	5.
CaloINN	-2.9061 $\pm$ 0.1459	4.
MDMA	-4.3483 $\pm$ 0.4573	11.
Calo-VQ	-8.1542 $\pm$ 0.3924	14.
CaloScore	-2.7967 $\pm$ 0.0569	3.
CaloScore distilled	-3.2482 $\pm$ 0.1576	7.
CaloScore single-shot	-3.4612 $\pm$ 0.2108	9.
iCaloFlow teacher	-3.1657 $\pm$ 0.1583	6.
iCaloFlow student	-3.6658 $\pm$ 0.2291	10.
SuperCalo	-2.7070 $\pm$ 0.1180	2.
DeepTree	-7.9900 $\pm$ 0.4257	13.
CaloPointFlow	-3.3695 $\pm$ 0.2519	8.
CaloVAE+INN	-4.5236 $\pm$ 0.5153	12.

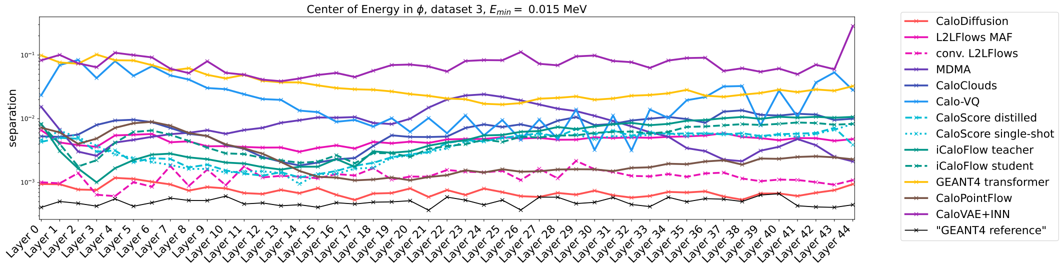
Histogram separation power ds3



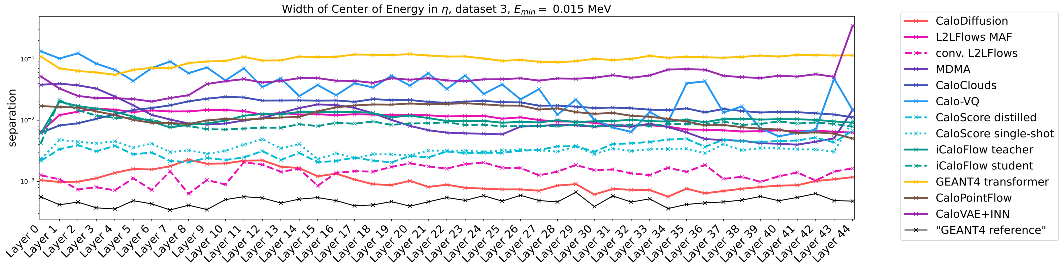
Histogram separation power ds3



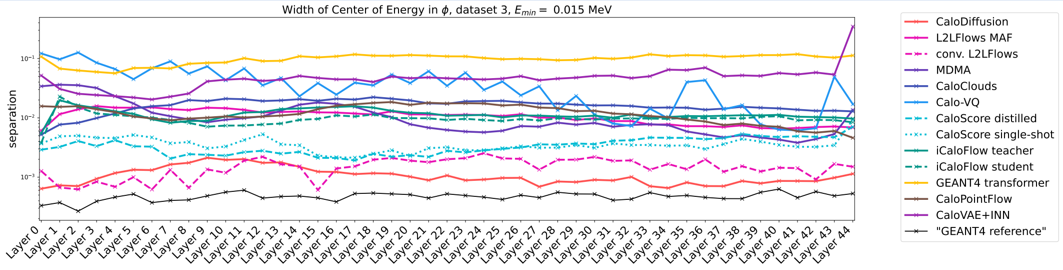
Histogram separation power ds3



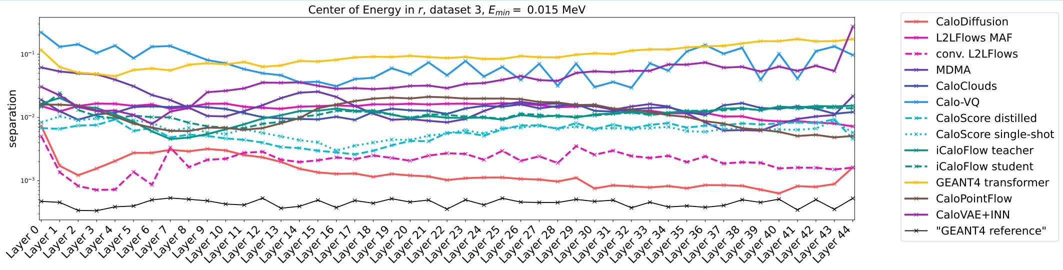
Histogram separation power ds3



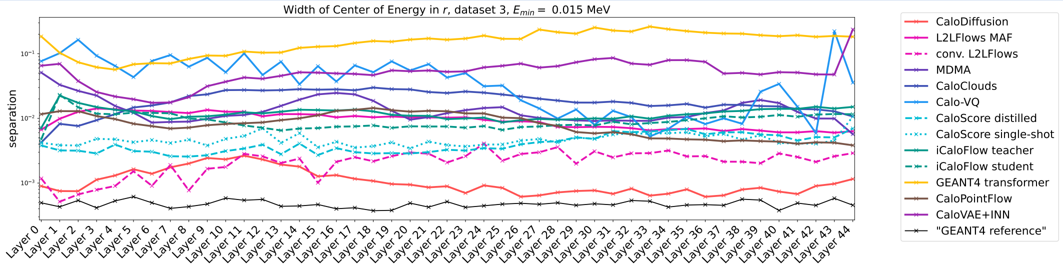
Histogram separation power ds3



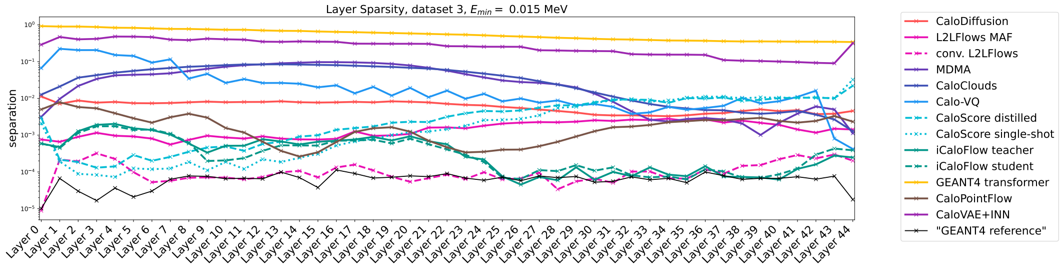
Histogram separation power ds3



Histogram separation power ds3



Histogram separation power ds3

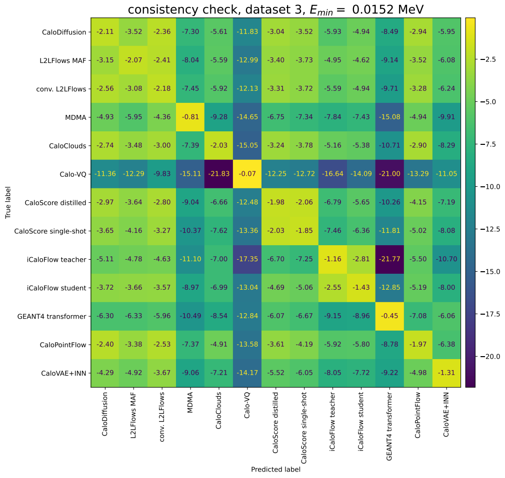


Log posterior ds3

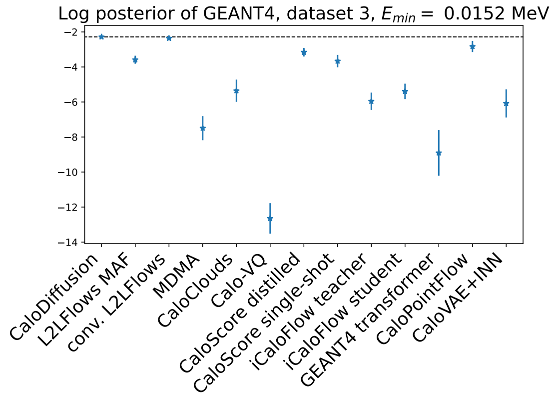
- submission vs submission:
  what we see: $\langle \log p(\text{model}|\textit{data}) \rangle$

$\Rightarrow$ each sample is correctly identified (diagonal is largest entry per row)

(mean of 10 independent runs.)



Log posterior ds3



CaloDiffusion	-2.2818 $\pm$ 0.0759	1.
L2LFlows MAF	-3.5881 $\pm$ 0.2290	5.
conv. L2LFlows	-2.3636 $\pm$ 0.0640	2.
MDMA	-7.4930 $\pm$ 0.6880	11.
CaloClouds	-5.3565 $\pm$ 0.6361	7.
Calo-VQ	-12.6454 $\pm$ 0.8734	13.
CaloScore distilled	-3.1603 $\pm$ 0.2414	4.
CaloScore single-shot	-3.6653 $\pm$ 0.3523	6.
iCaloFlow teacher	-5.9583 $\pm$ 0.4948	9.
iCaloFlow student	-5.3947 $\pm$ 0.4400	8.
GEANT4 transformer	-8.9068 $\pm$ 1.3047	12.
CaloPointFlow	-2.8335 $\pm$ 0.3178	3.
CaloVAE+INN	-6.0804 $\pm$ 0.8054	10.

(based on 10 independent runs.)

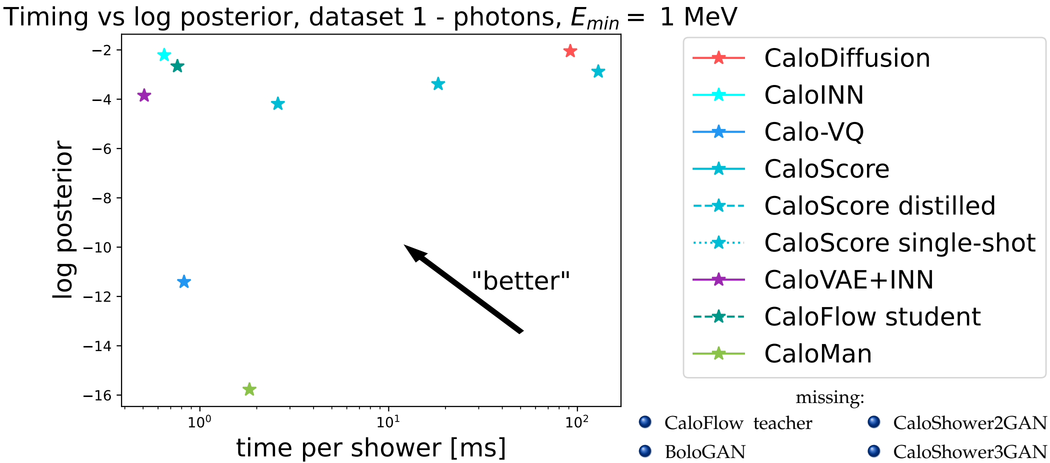
Generation Times (preliminary!)

These numbers are preliminary (pending verification)!

- I run on an *NVIDIA TITAN V** at Rutgers.
- Configuration: batch size of (mostly) 100*, single run*
- generating the full dataset (100k), but for some models fewer (10k) samples
- Only about 3/4 of the models are included here*

* There'll be more in the final document.

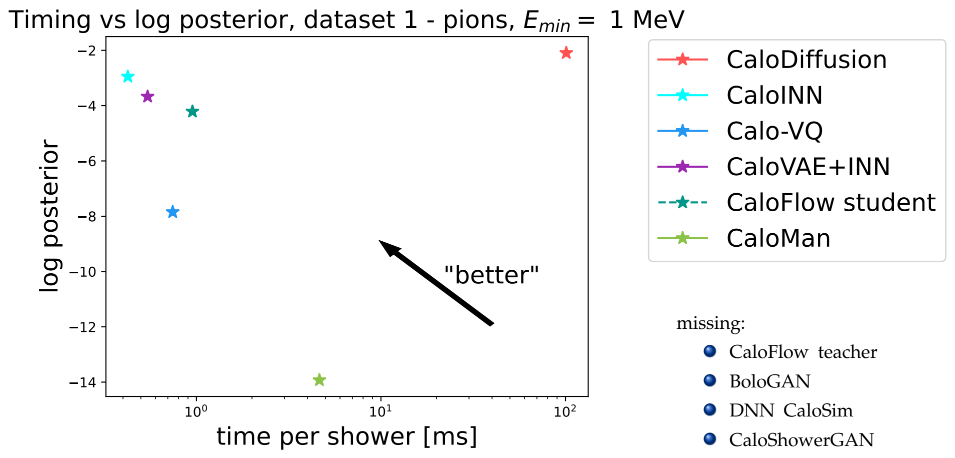
Generation Times (preliminary!) ds1 photons



missing:

- CaloFlow teacher
- CaloShower2GAN
- BoloGAN
- CaloShower3GAN

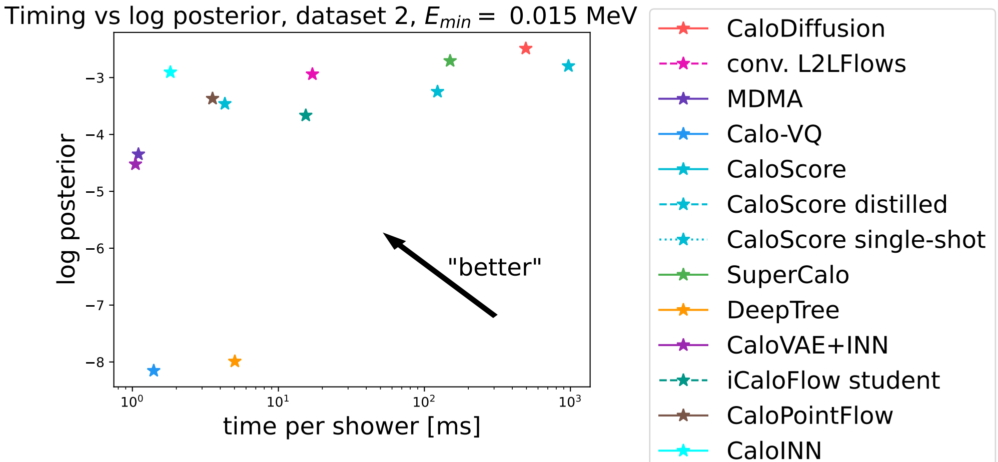
Generation Times (preliminary!) ds1 pions



missing:

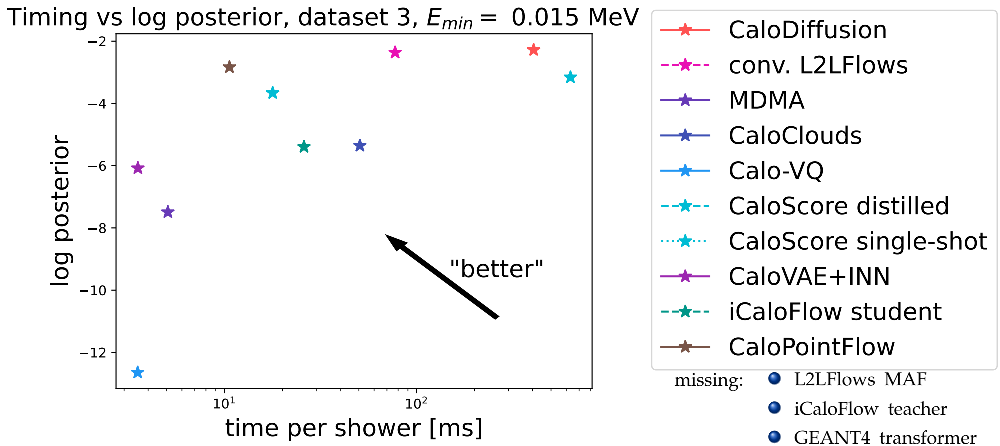
- CaloFlow teacher
- BoloGAN
- DNN CaloSim
- CaloShowerGAN

Generation Times (preliminary!) ds2



missing: iCaloFlow teacher

Generation Times (preliminary!) ds3



missing:
- L2LFlows MAF
- iCaloFlow teacher
- GEANT4 transformer

The Fast Calorimeter Challenge 2022

⇒ The CaloChallenge was very successful:

- 10+ talks in multiple sessions here at ML4Jets 2023
- 5 talks at ML4Jets 2022 in Rutgers
- 10+ papers on arXiv and some more "in the making"

⇒ the write-up will follow soon, but I want to include

- new submissions from Tuesday ;-)
- many more metrics.

"Calorimeter Simulation"
via midjourney

A *Big Thank You* to everyone who participated with comments, discussions, presentations, metrics, and submissions!