Advancing Particle Physics with Point Cloud-Based Generative Models

A Generative Machine Learning Group Perspective

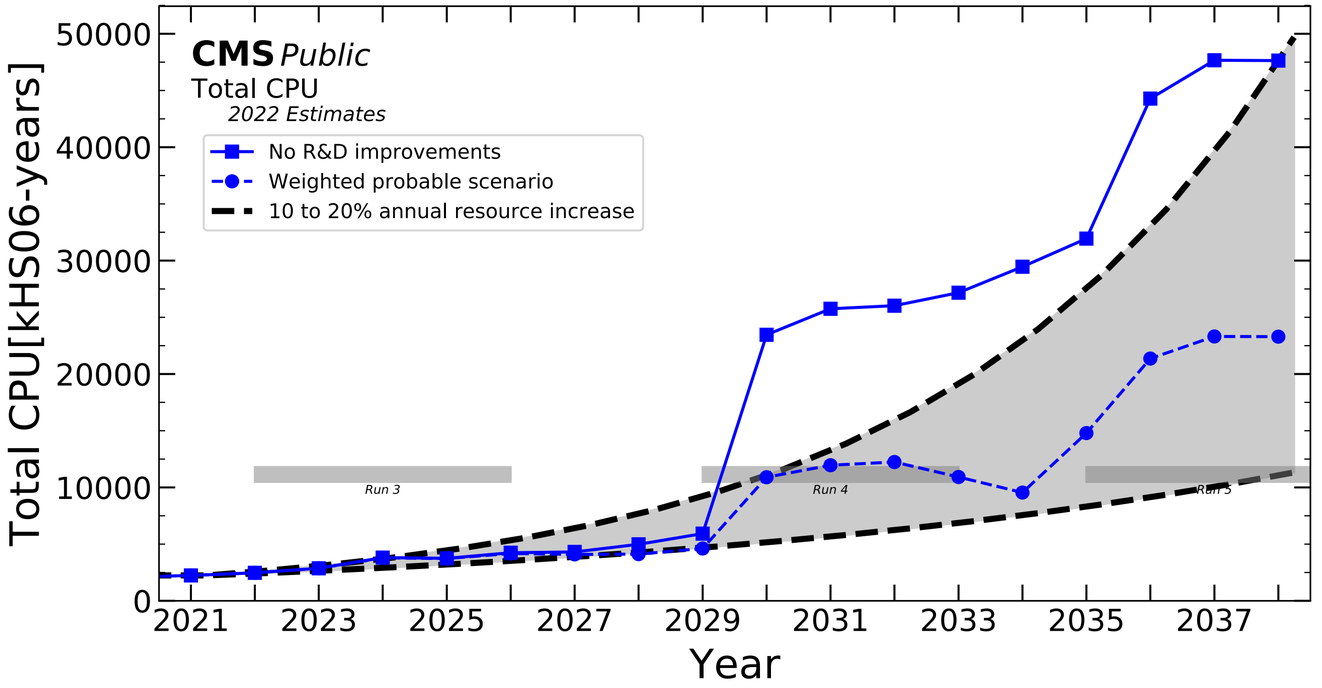
DESY CMS: Machine Learning Group
Kerstin Borras, Dirk Krücker, Benno Käch[1], Isabell Melzer-Pellmann, Moritz Scham[2], Simon Schnake
01.07.24
FH SciComp Workshop

[1]Funded through Helmholtz AI grant, number ZT-I-PF-5-064
[2]Funded through Helmholtz AI grant, number ZT-I-PF-5-3

CMS Experiment

Computing Challenge

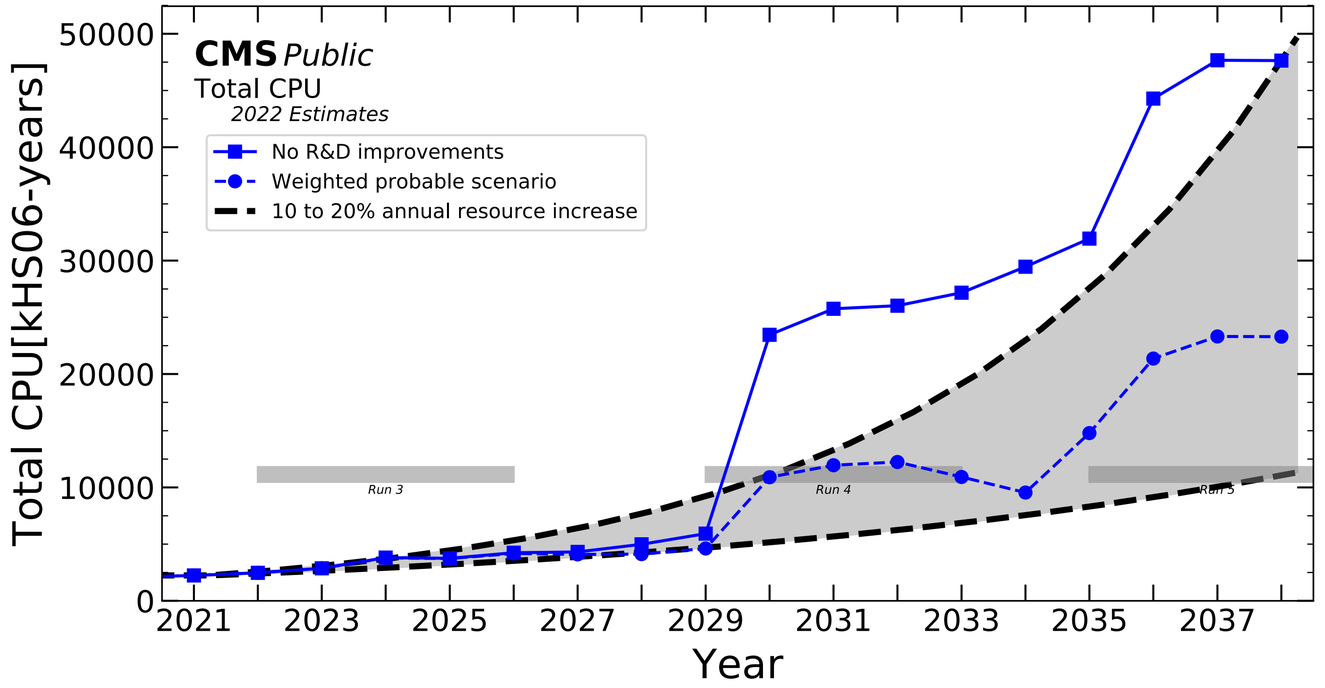
- The projected computation need is larger than the available resources
- Monte Carlo simulation are the most computing intense part
- Already > 50 % of computing budget
- Coming High Luminosity upgrades makes MC approach challenging



CMS Experiment

Computing Challenge

- The projected computation need is larger than the available resources
- Monte Carlo simulation are the most computing intense part
- Already > 50 % of computing budget
- Coming High Luminosity upgrades makes MC approach challenging



There is a need for replacing simulations by fast generative models

Point Clouds

Jets & Calorimeter Showers

Calorimeter Showers

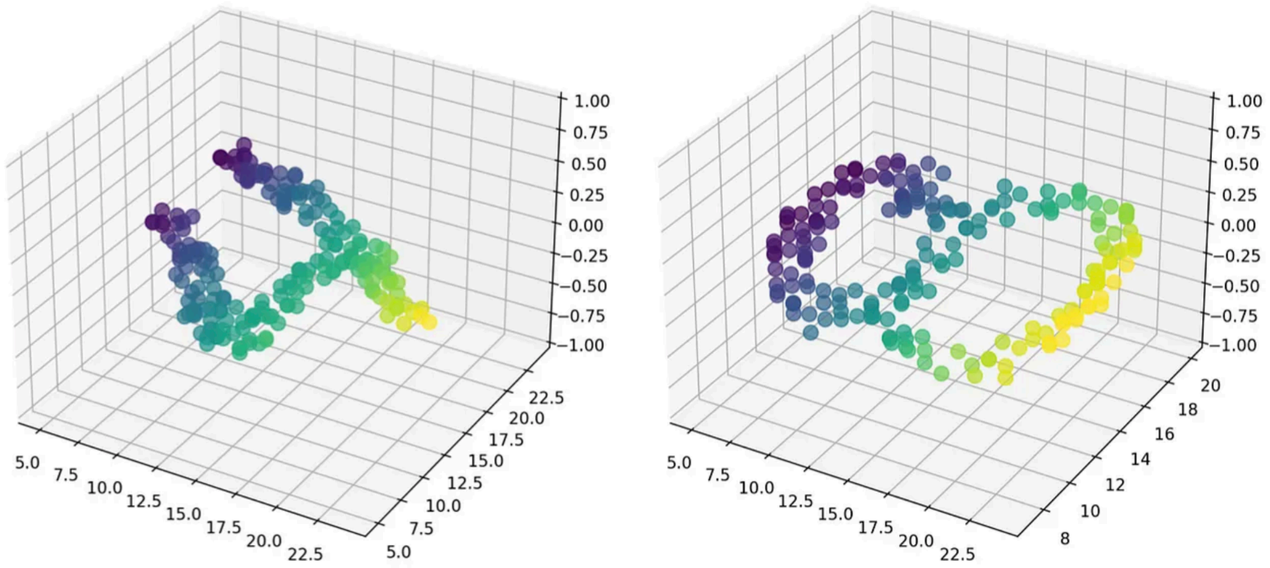
- Can have Irregular geometry
- For most showers only a fraction of cells contain an entry (sparse)
- Point Clouds are a good representation
- Detector independend

Jets

- Jets are point clouds by construction

Point Clouds

- Point Clouds are a a list of coordinates plus features
- Non-fixed size
- permutation equivariant

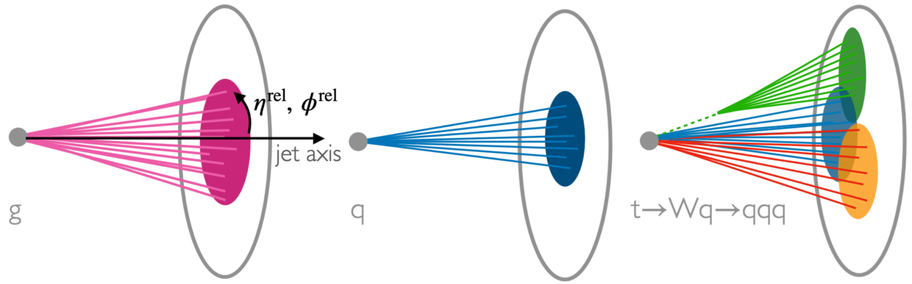


Open Datasets

JetNet

Kansal et al

- Public dataset generated with Pythia
- Anti-kT clustered (R=0.8)
- 3 Classes: gluon, light quark and top

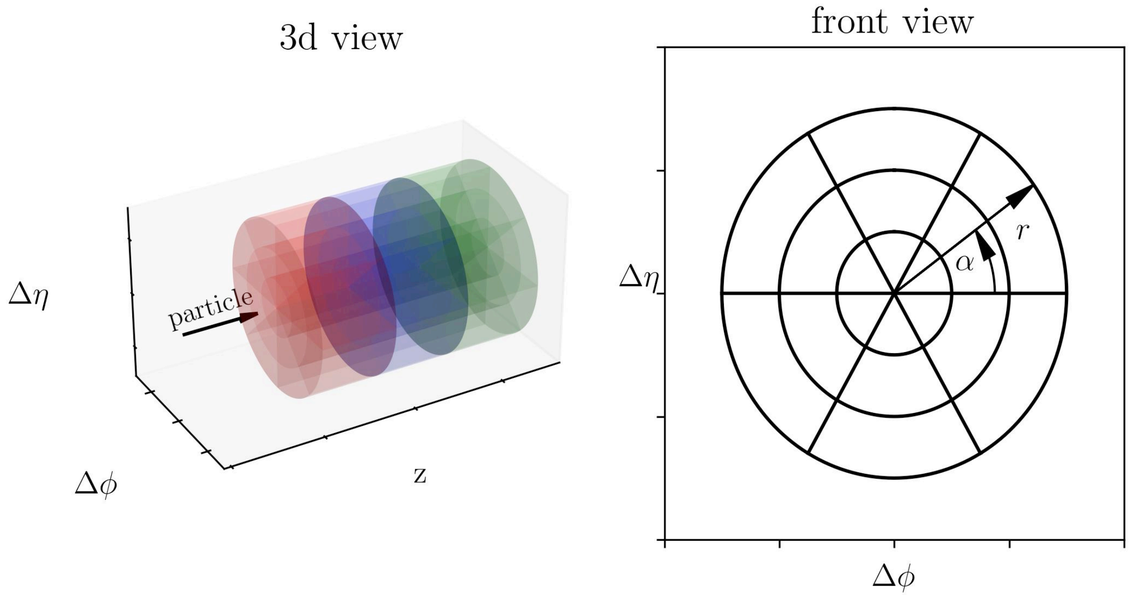


https://jetnet.readthedocs.io/en/latest/

CaloChallenge

Krause et al

- Challenge to build generative models for Calorimeter Showers
- 4 different datasets of increasing complexity

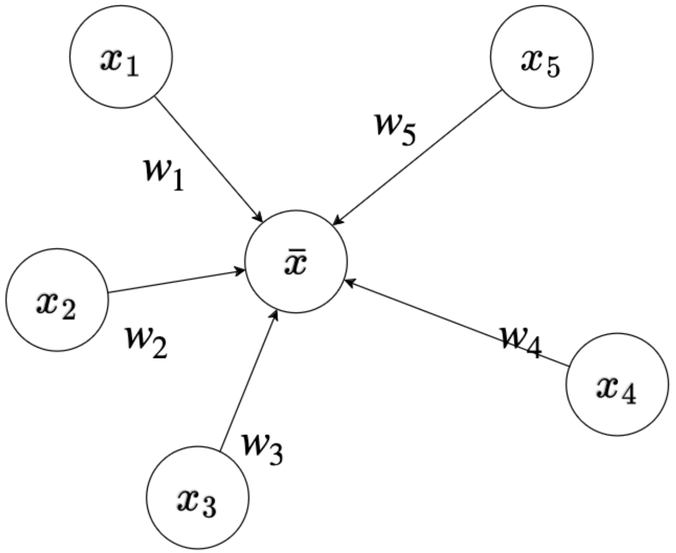


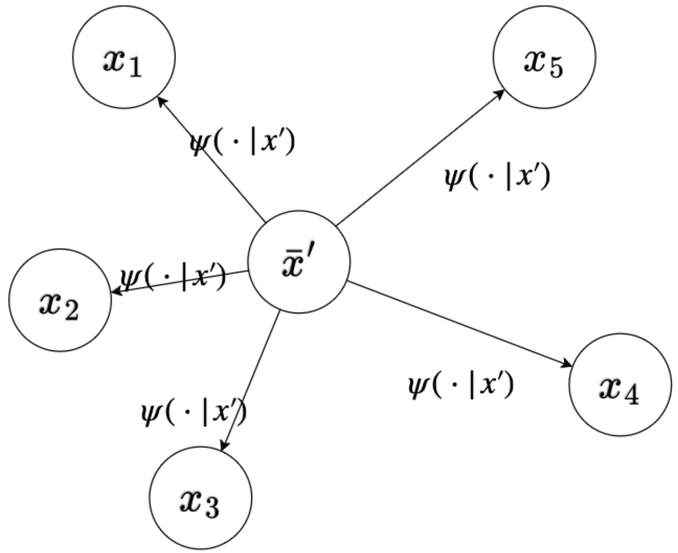
https://calochallenge.github.io/homepage/

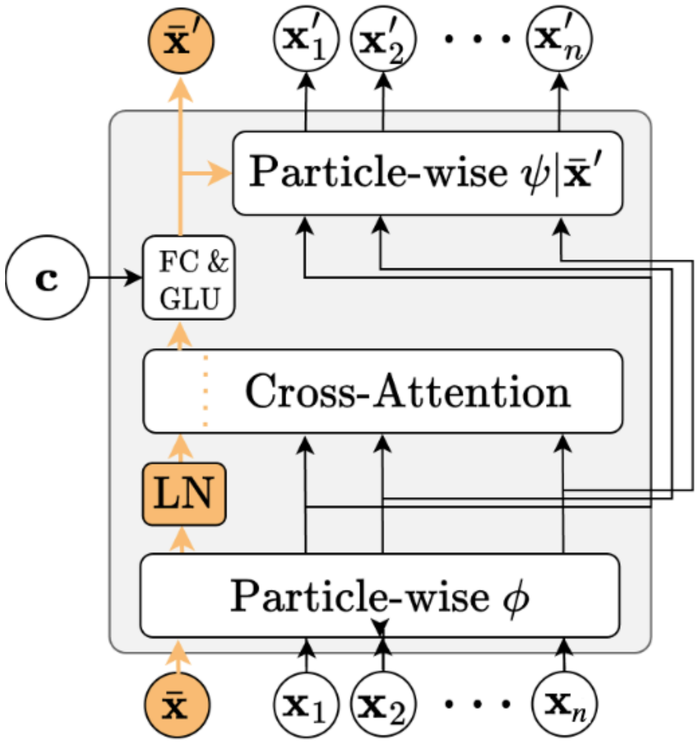
Pay Attention to Mean Fields for Point Cloud Generation

Benno Käch [arXiv: 2305.15254]

- Generative Adversarial Network
- Initialize a point cloud with noise (normal distributed)
- There is one additional mean point introduced (called mean field)
- That mean point aggregates all information by a linear Cross-Attention layer
- The information is added to all points
- The points are individually morphed



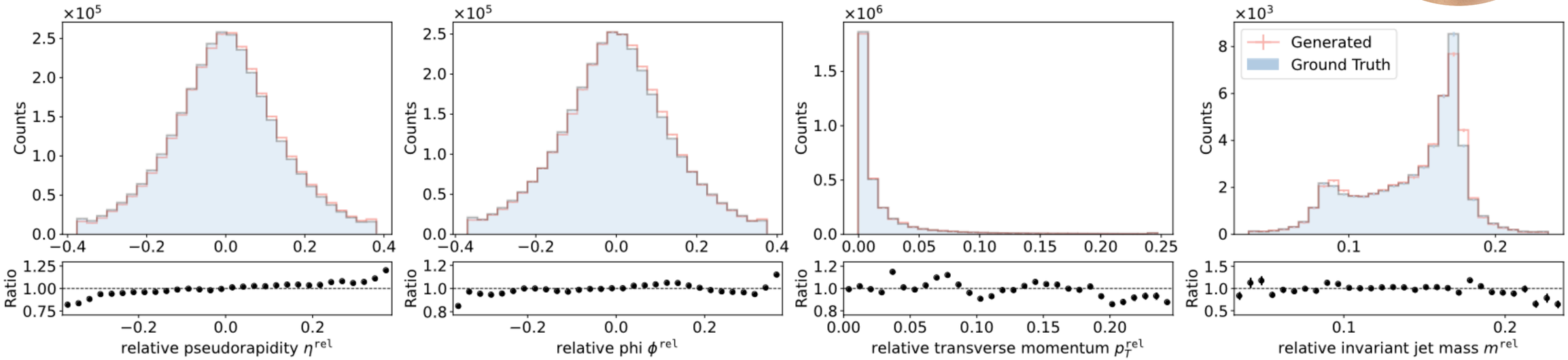




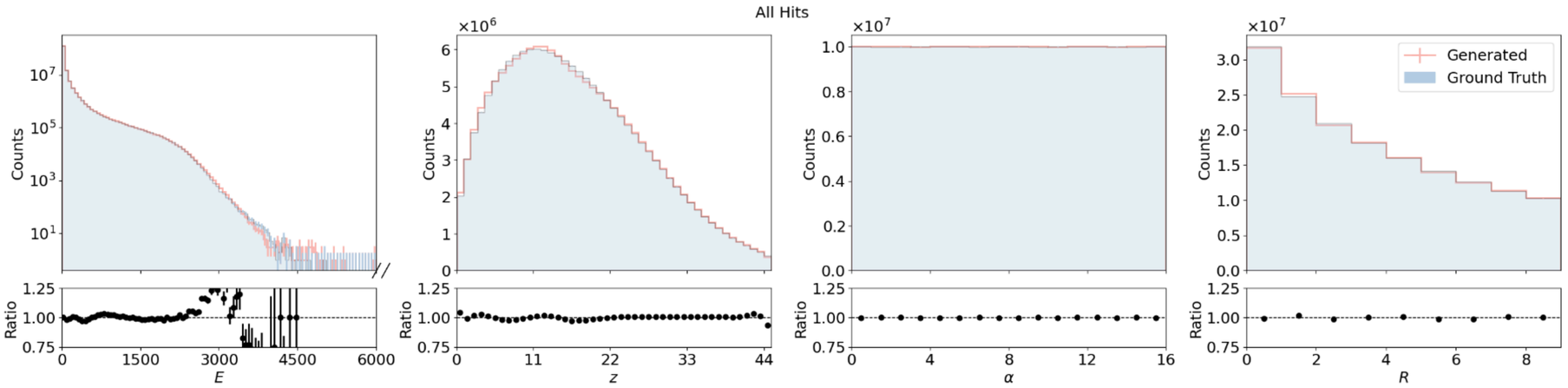
Pay Attention to Mean Fields for Point Cloud Generation

Benno Käch [arXiv: 2305.15254]

Top-quark JetNet150



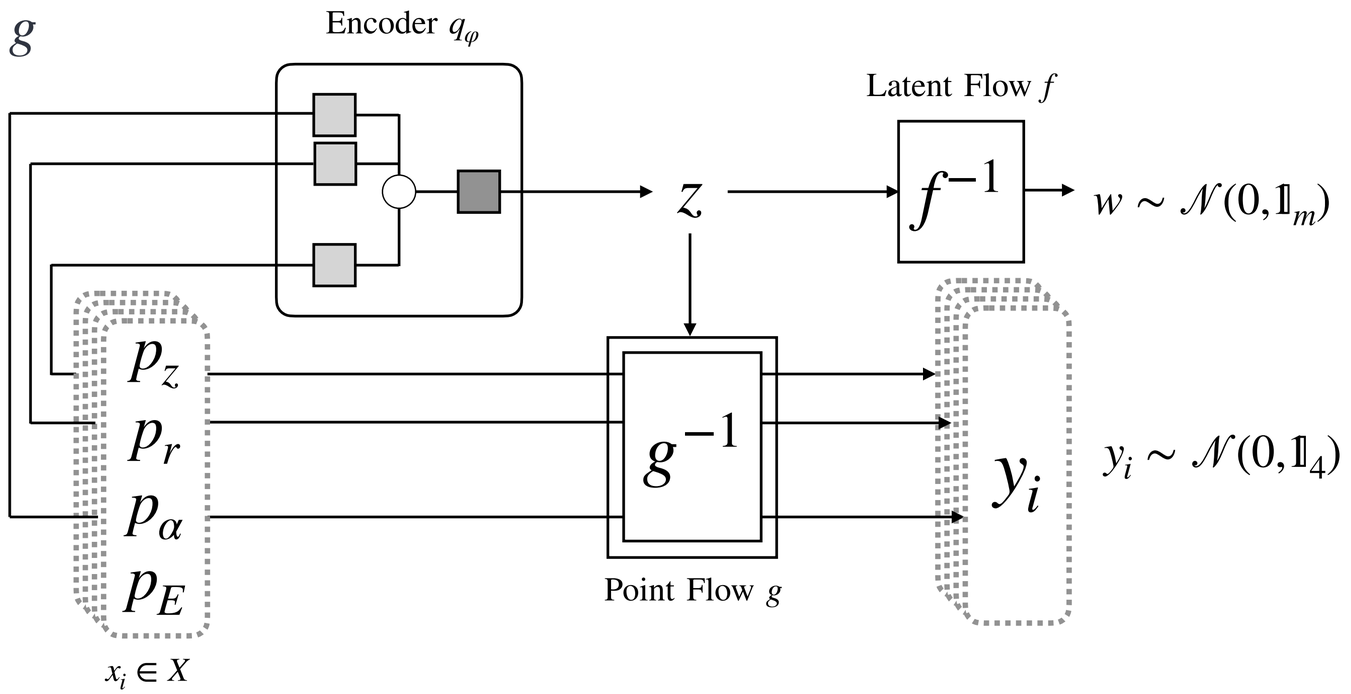
CaloChallenge Dataset 2

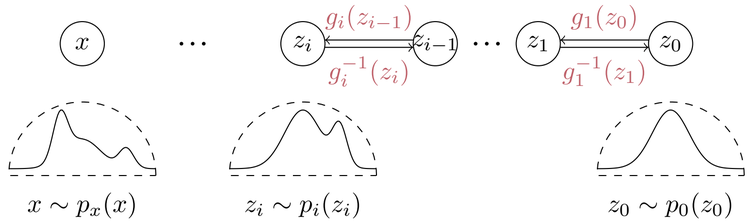


CaloPointFlow

Simon Schnake [arXiv: 2403.15782]

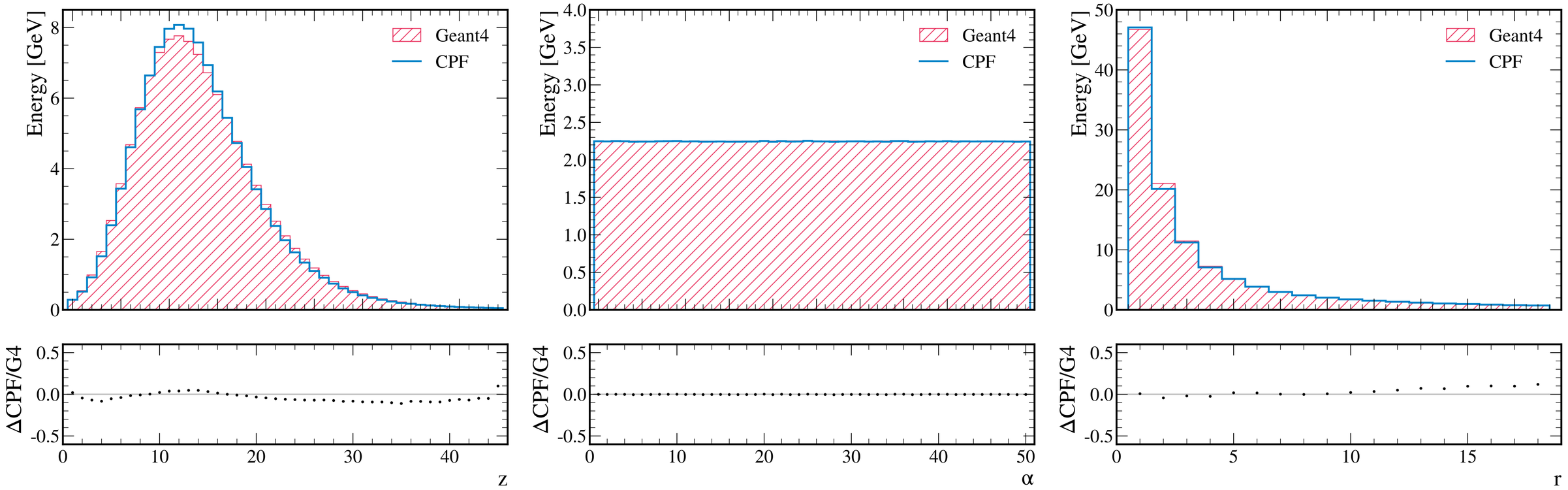
- Point-wise normalizing flow Point Flow g
- Shower shaped encoded in z
- Shower shape is generated by the Latent Flow f





CaloPointFlow

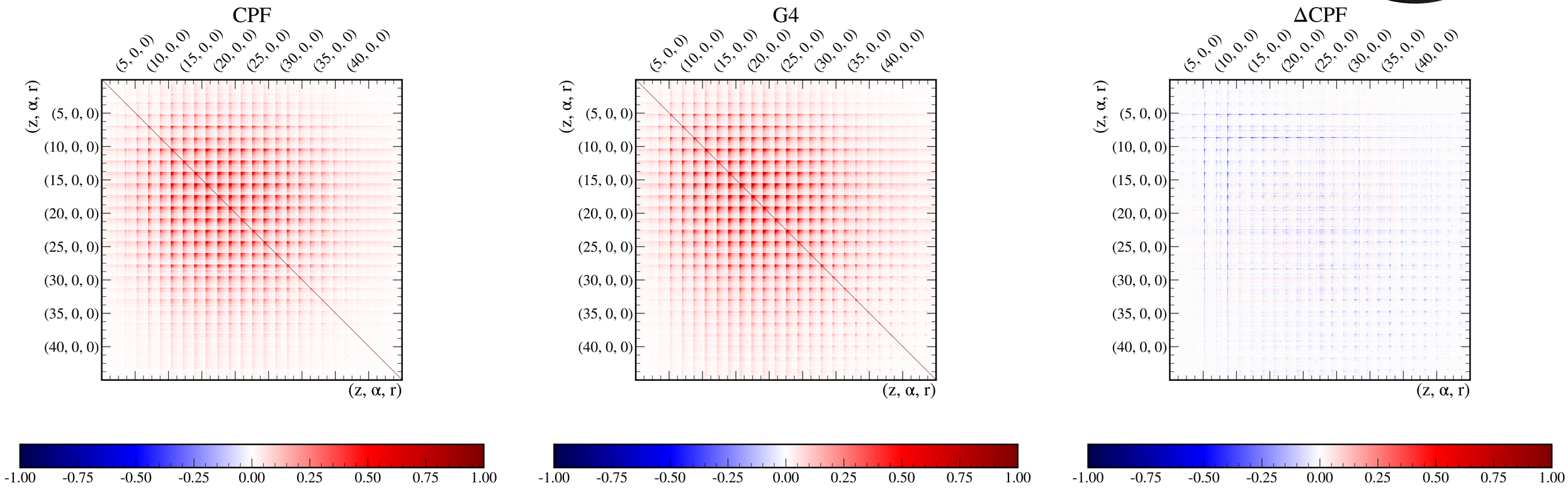
Simon Schnake



CaloChallenge Dataset 3

CaloPointFlow

Simon Schnake



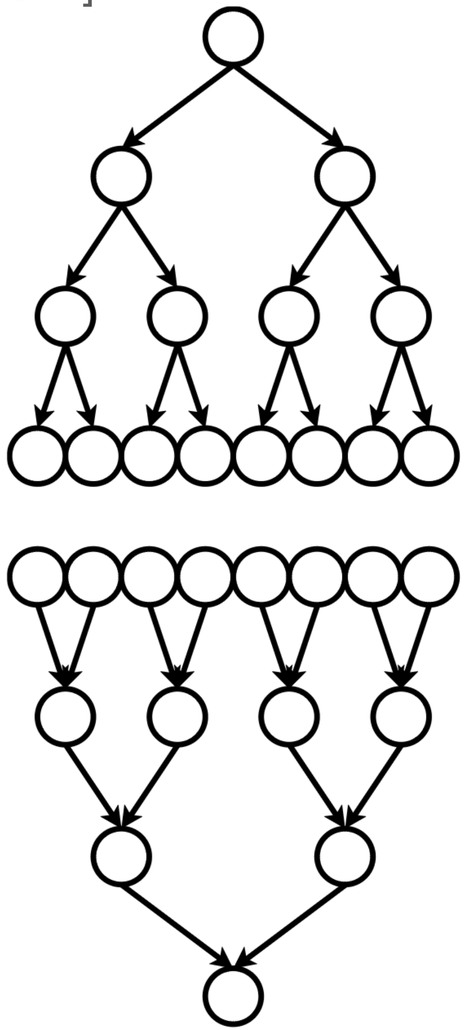
CaloChallenge Dataset 3

DeepTreeGAN

Moritz Scham

[arXiv:2311.12616]
[arXiv:2312.00042]

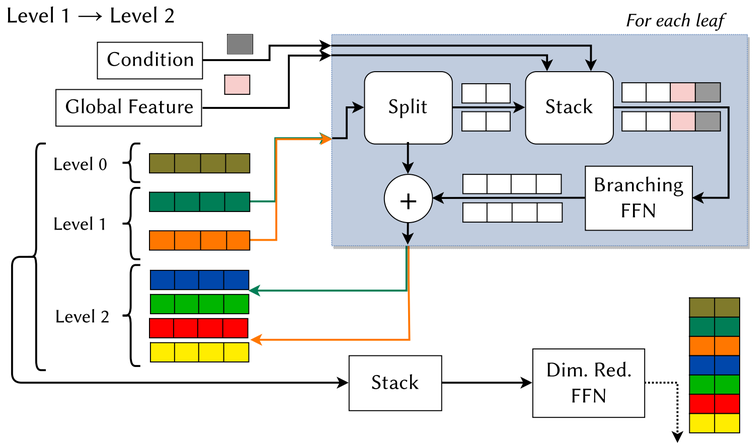
- Generative Adversarial Network
- Generator and Critic have Tree Structure
- Generator:
  - fixed branching + message passing at each level
  - slicing at the end
- Discriminator
  - Bipartite Pooling to fixed point cloud
  - Size reduction of point cloud
  - seperate loss term for each level



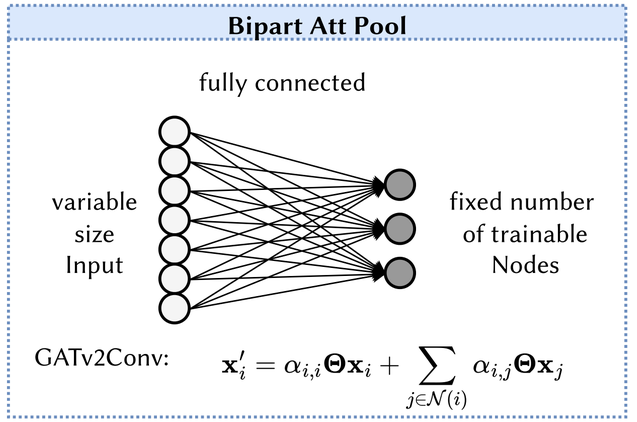
Generator

Critic

Branching



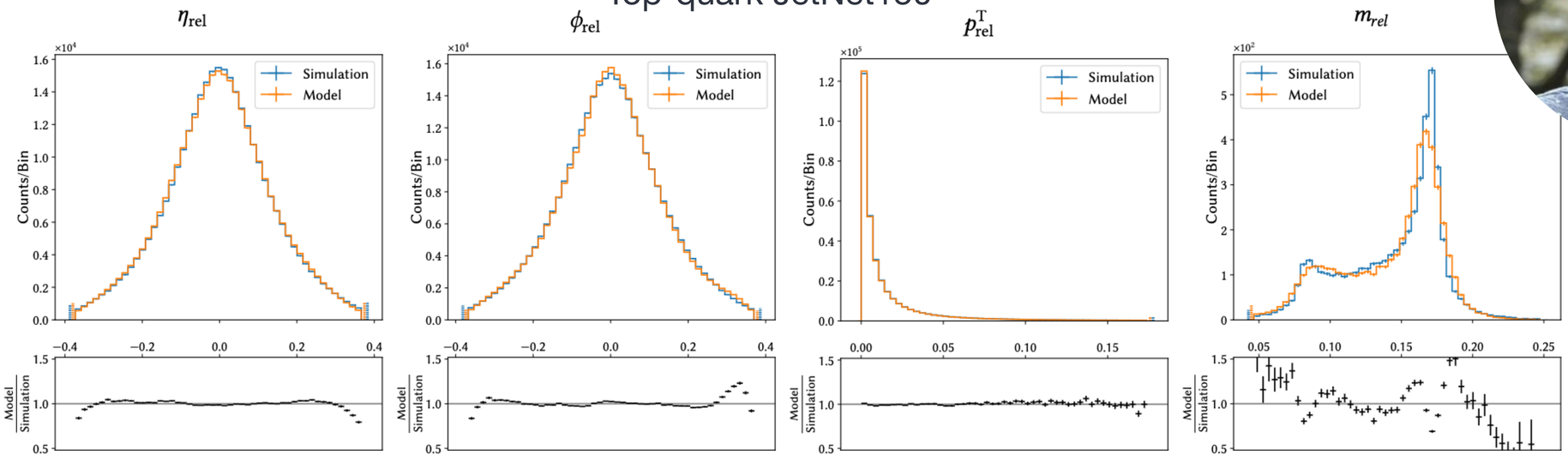
Pooling



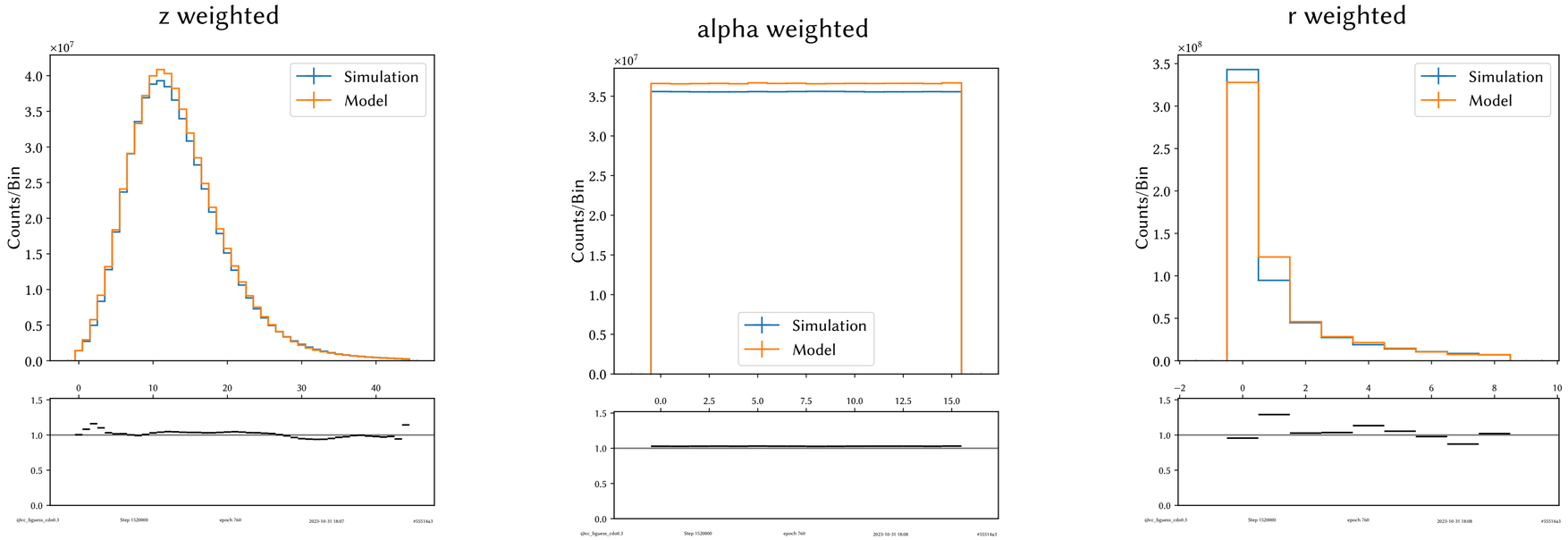
DeepTreeGAN

Moritz

Top-quark JetNet150

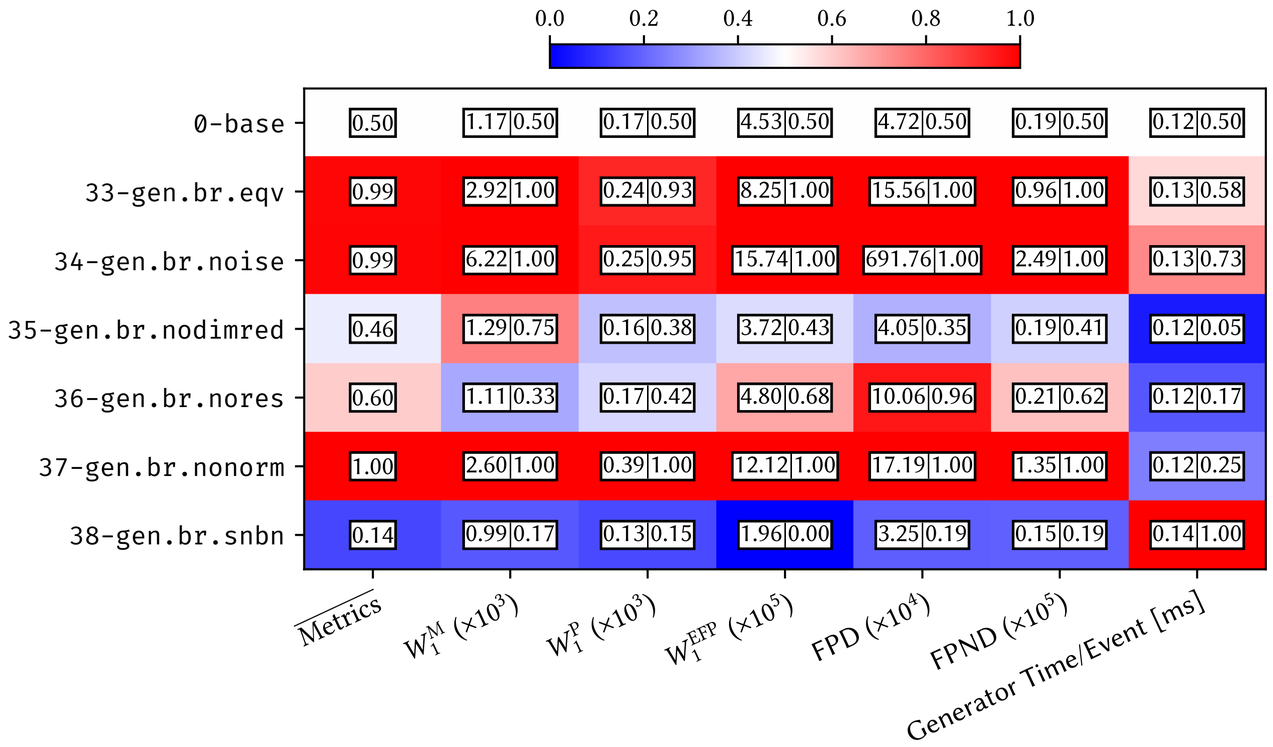


CaloChallenge Dataset 2



Computational Demand

- Development of generative modeling requires often high computing power
- Hyperparameter studies or ablation studies are extremely demanding
- Example: Ablation study by Moritz Scham (93 trainings a 360 h on a v100) to define the final model



Branching Ablation Study by Moritz Scham

Example Plot with 6 of 80 model variations

Conclusion

- Motivated Generative Point Cloud Models For CMS
- Work of the Group
  - Benno: Pay Attention to Mean Fields for Point Cloud Generation [arXiv: 2305.15254]
  - CaloPointFlow [arXiv:2403.15782] (submitted to SciPost)
  - Moritz: DeepTreeGAN [arXiv:2311.12616, arXiv:2312.00042]
- All training is performed on GPUs on the excellent DESY Maxwell cluster
- We are very grateful to have access to these resources

Thank you for your attention!