Advancing Particle Physics with Point Cloud-Based Generative Models

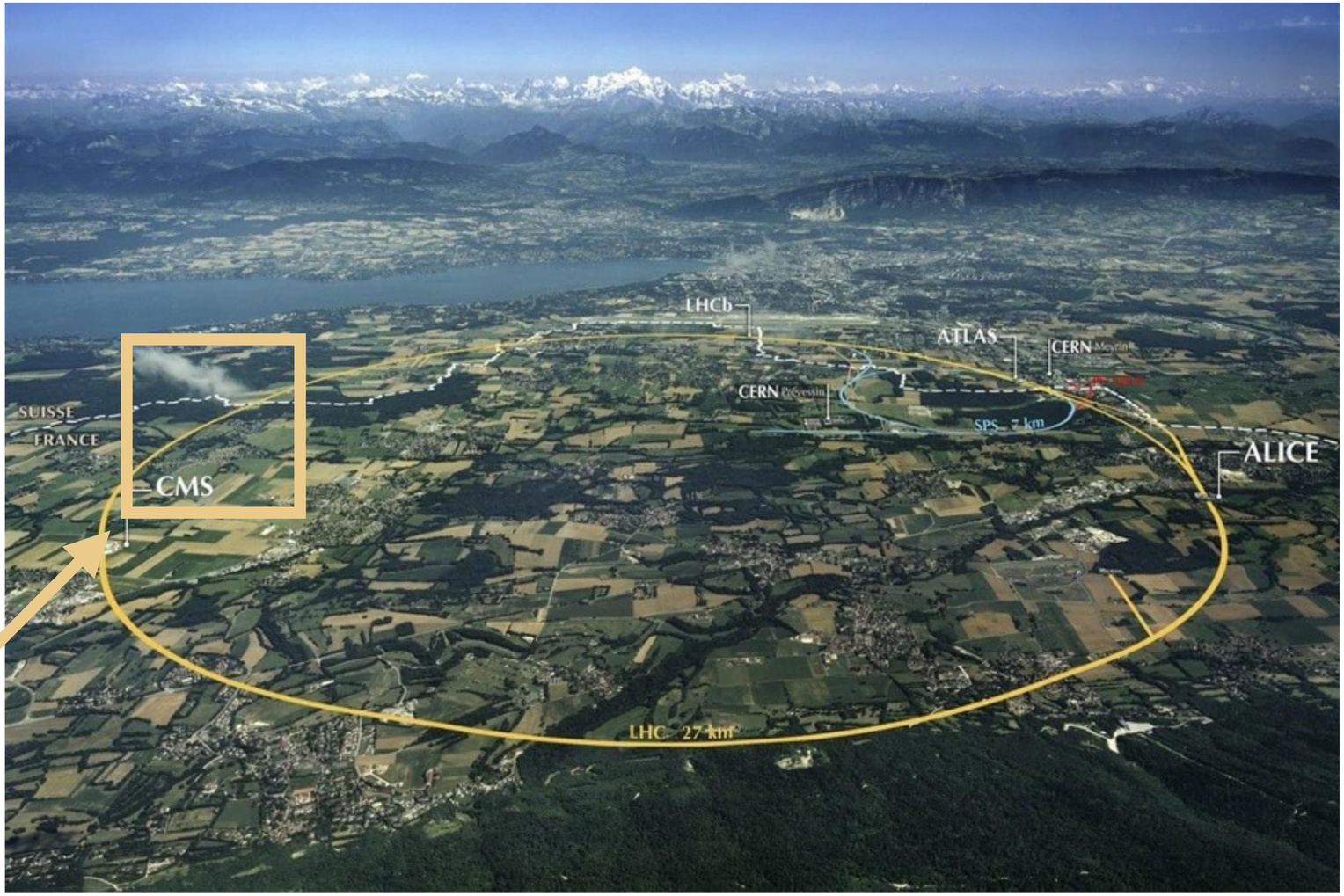
A Generative Machine Learning Group Perspective

DESY CMS: Machine Learning Group
Kerstin Borras, Dirk Krücker, Benno Käch[1], Isabell Melzer-Pellmann, Moritz Scham[2], Simon Schnake
07.11.23
6th Round Table on Deep Learning at DESY 2023

[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

At the Large Hadron Collider

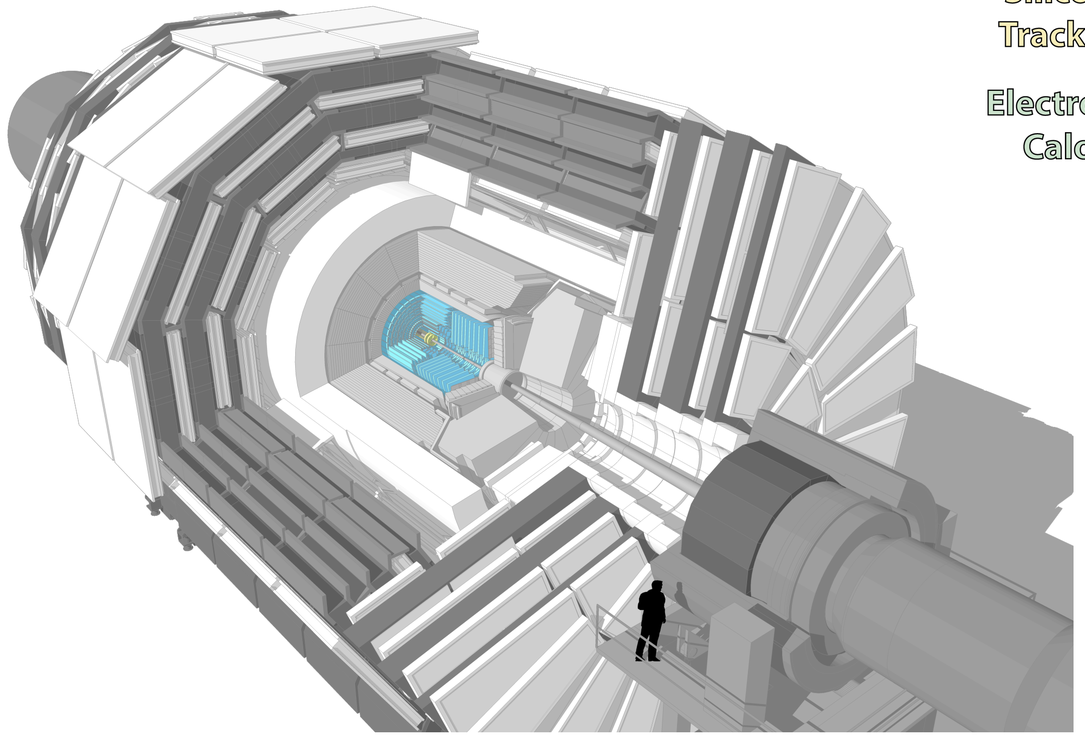


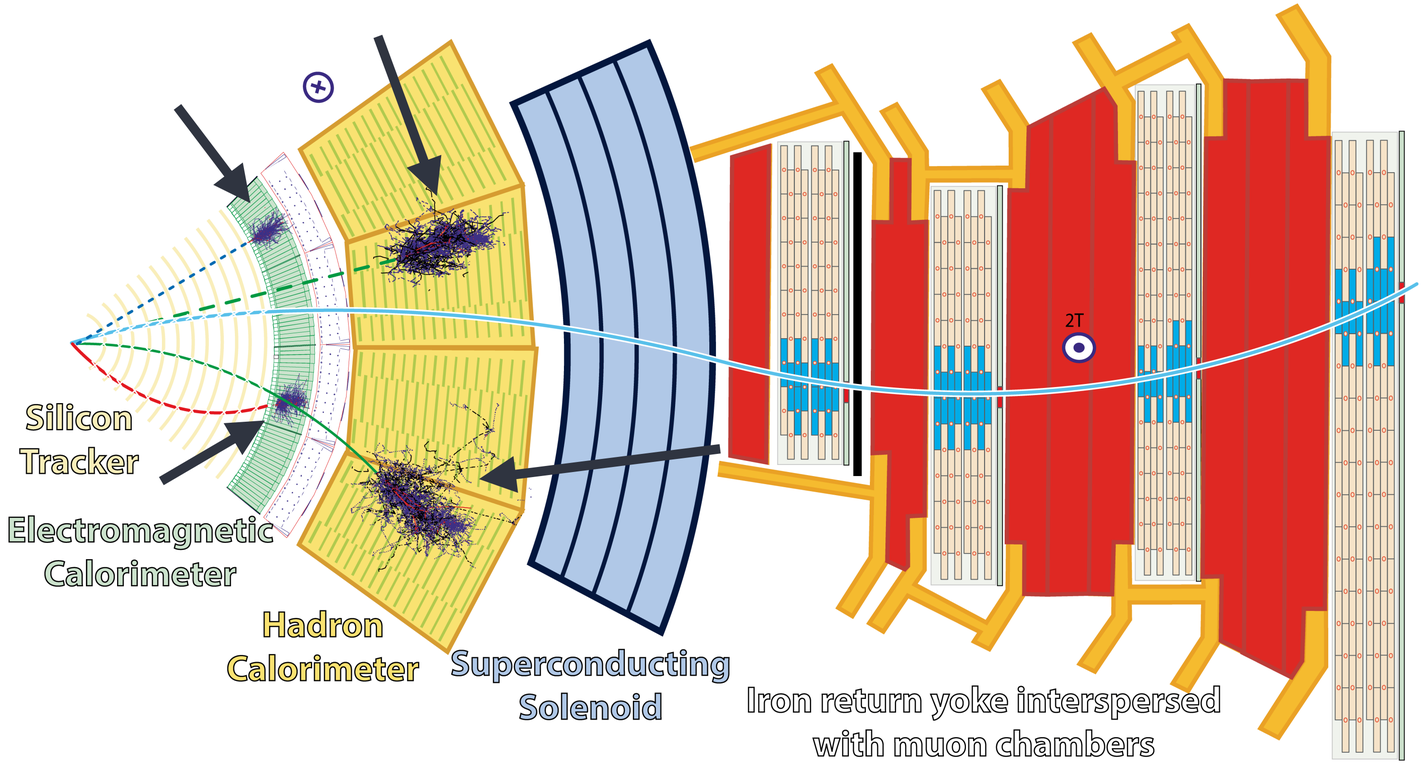
Source: https://www.swissinfo.ch/ger/wissen-technik/teilchenbeschleuniger_large-hadron-collider-feiert-10--geburtstag/44386312

CMS Experiment

Calorimeter Showers

- Particles shower in the calorimeters
- The energy is read out in the active cells
- Each shower is a collection of (E, x, y, z)



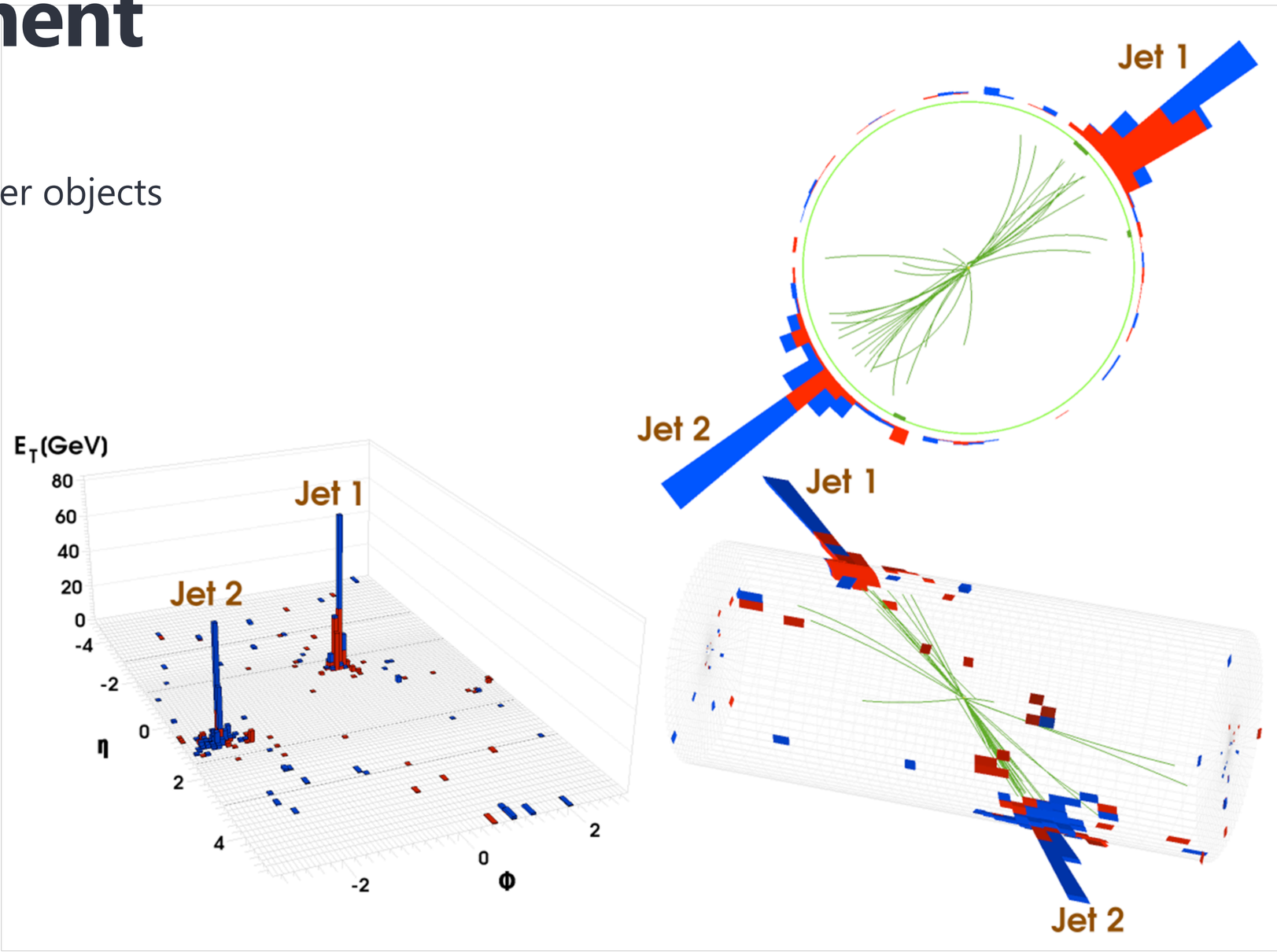


CMS Experiment

Jets

- Reconstructed particle shower objects in our detector
- 3D objects in (η, φ, P_T)

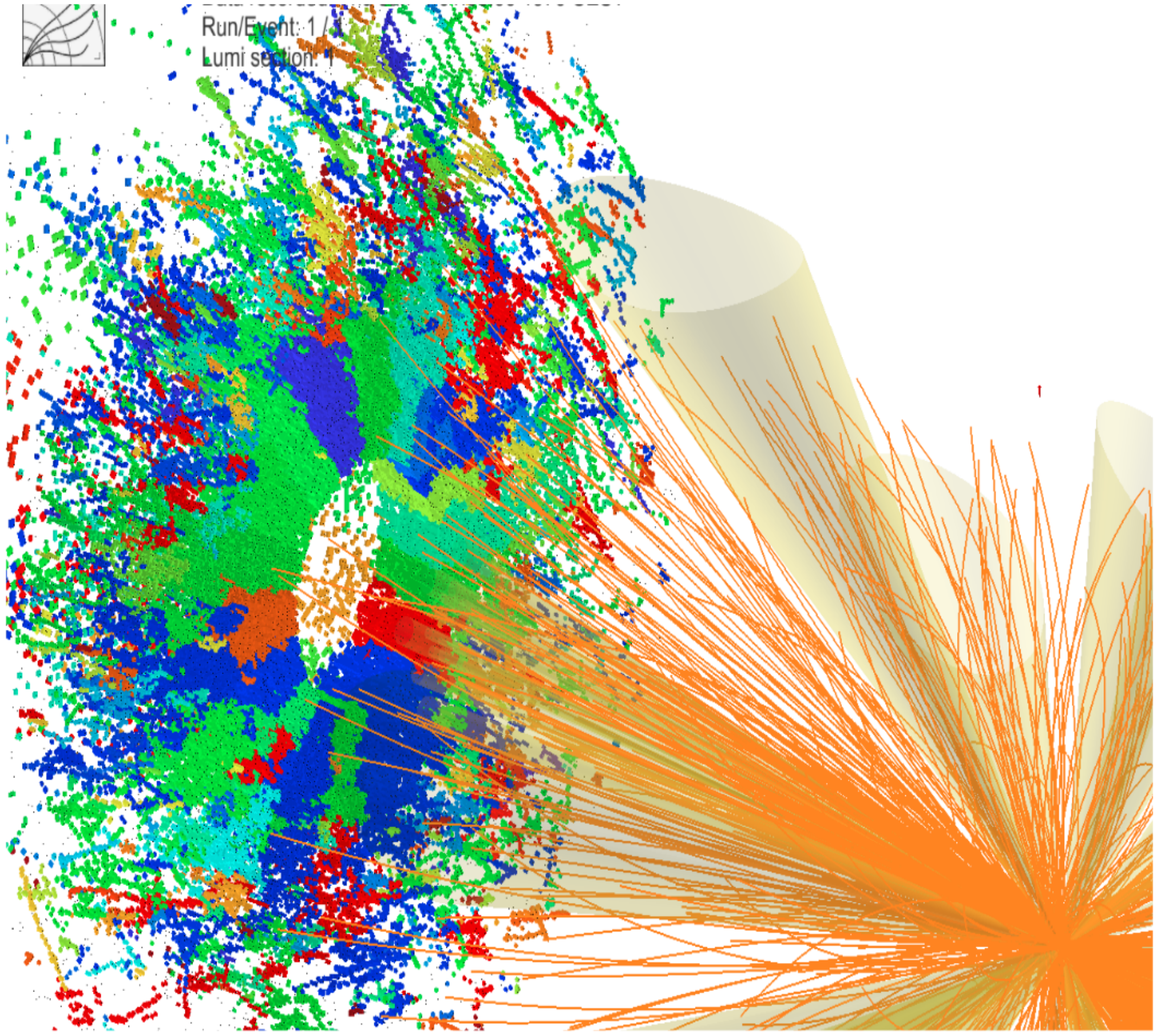
[https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME]



CMS Experiment

High Granularity Calorimeters

- LHC high luminosity phase of the LHC
- upgraded high granular calorimeters
- HGCAL for CMS
- High luminosity ⇒ more tracks
- More tracks ⇒ higher resolution
- higher resolution ⇒ more computation necessary

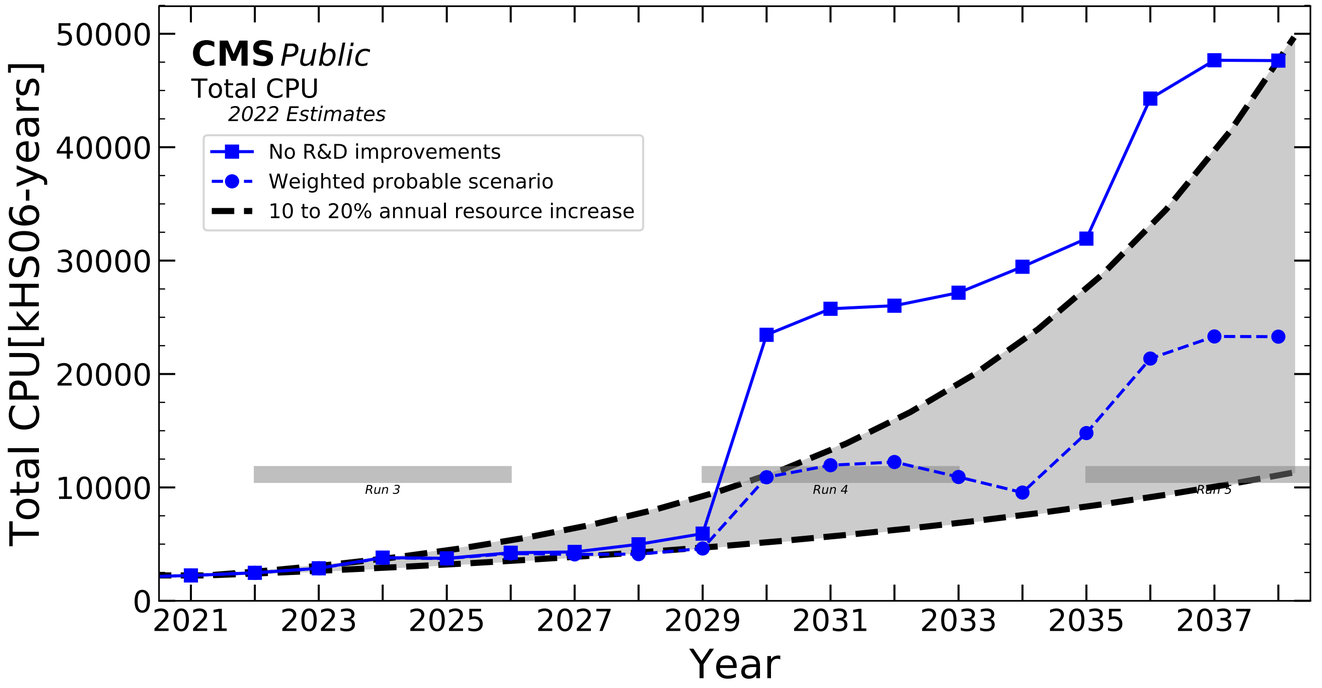


Simulated view of one HGCal endcap, containing particles from the nominal 140 pileup interaction expected at the HL-LHC
[D. Newbold - The High-Luminosity Upgrade of the CMS Detector]

CMS Experiment

Computing Challenge

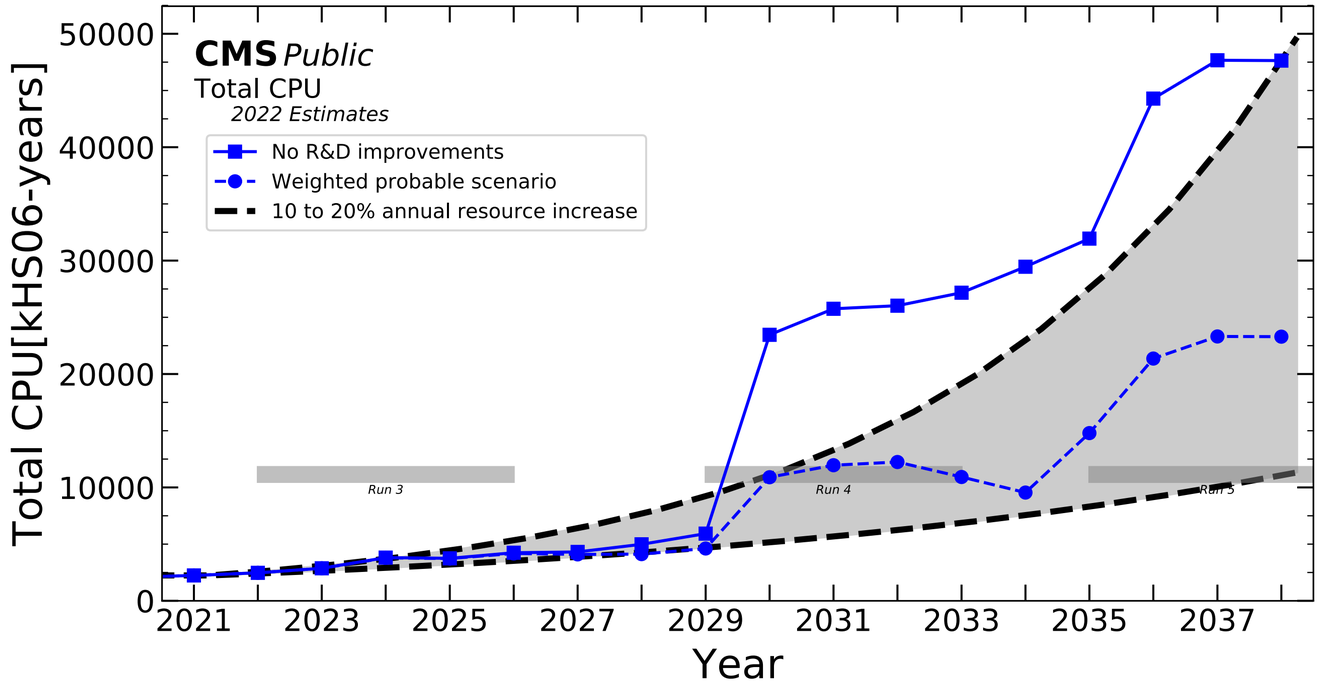
- The projected computation need is larger than the available resources
- Monte Carlo simulation are the most computing intense part
- Already > 50 % of computing budget
- Coming High Luminosity upgrades makes MC approach challenging



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

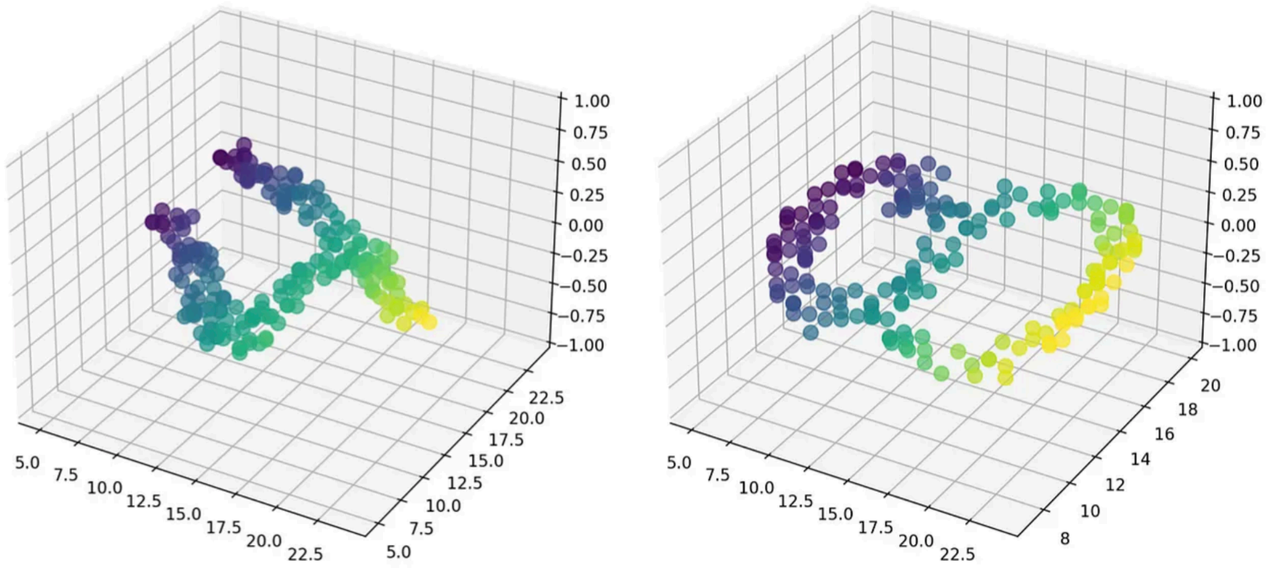
- Can have Irregular geometry
- For most showers only a fraction of cells contain an entry (sparse)
- Point Clouds are a good representation
- Detector independend

Jets

- Jets are point clouds by construction

Point Clouds

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

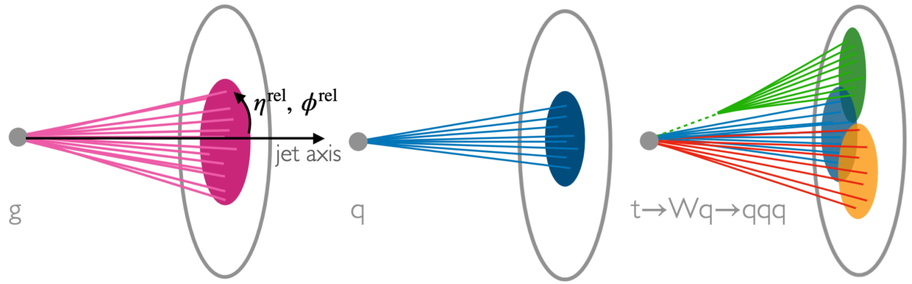


Open Datasets

JetNet

Kansal et al

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

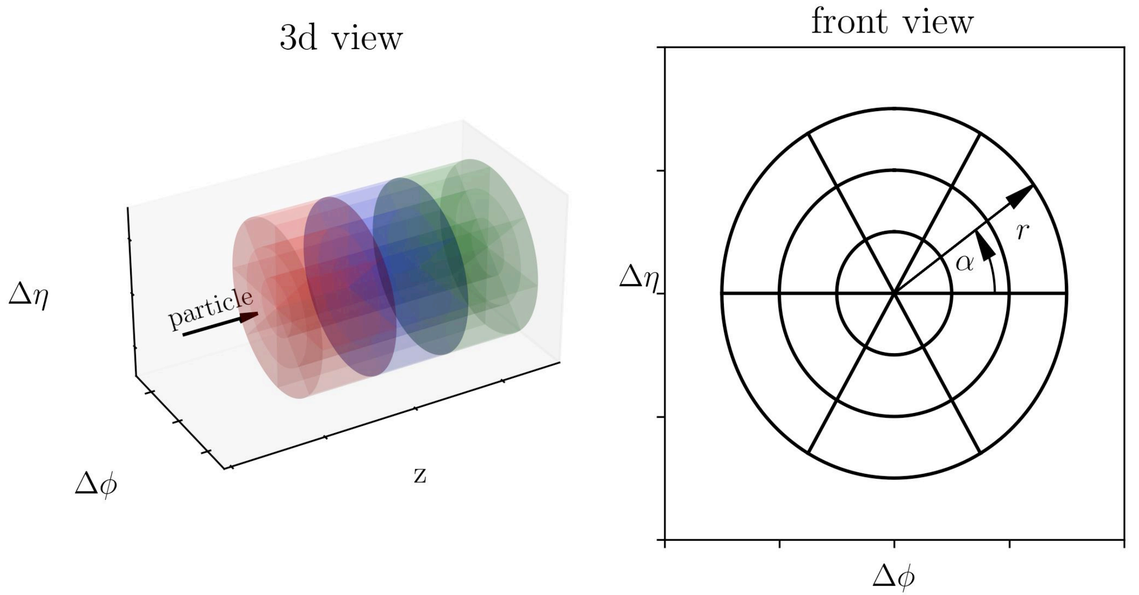


https://jetnet.readthedocs.io/en/latest/

CaloChallenge

Krause et al

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

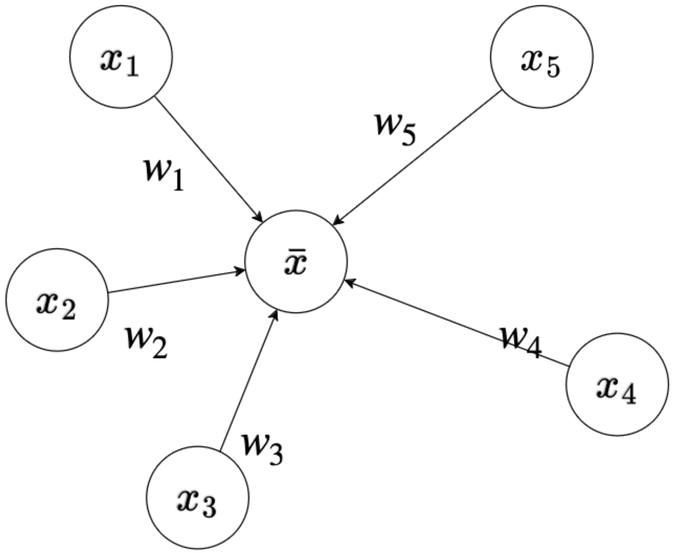


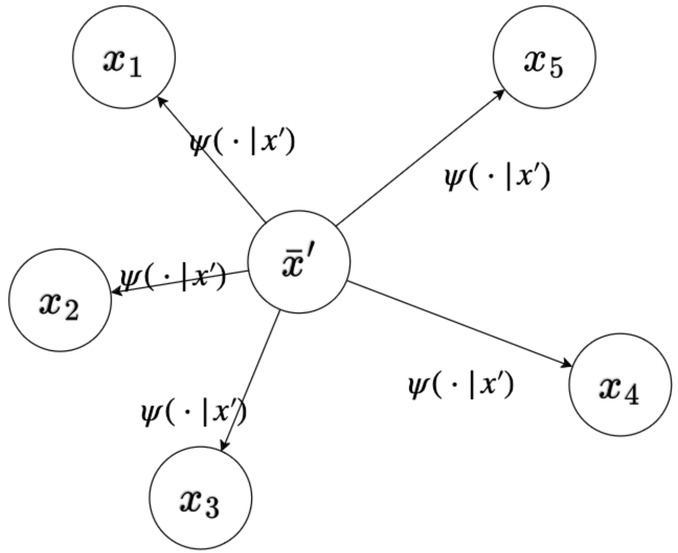
https://calochallenge.github.io/homepage/

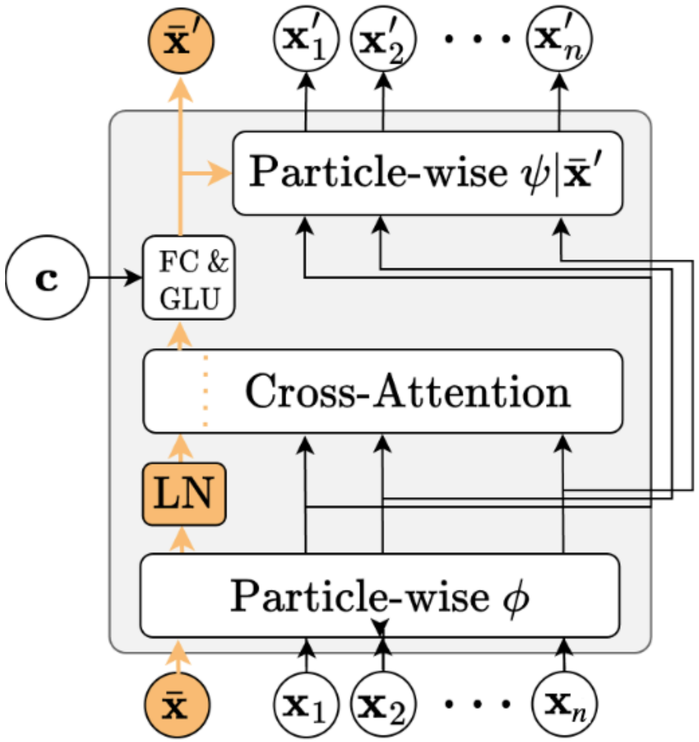
Pay Attention to Mean Fields for Point Cloud Generation

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



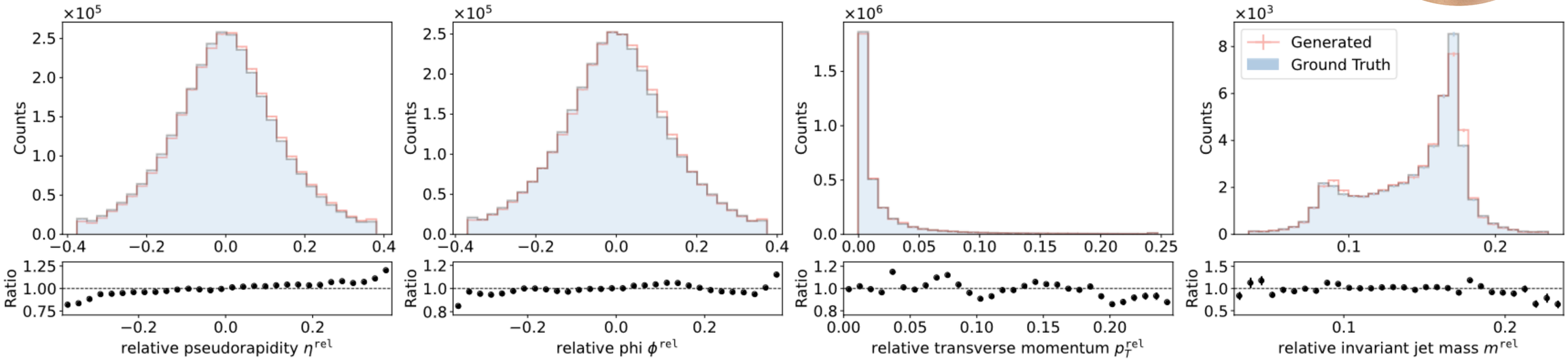




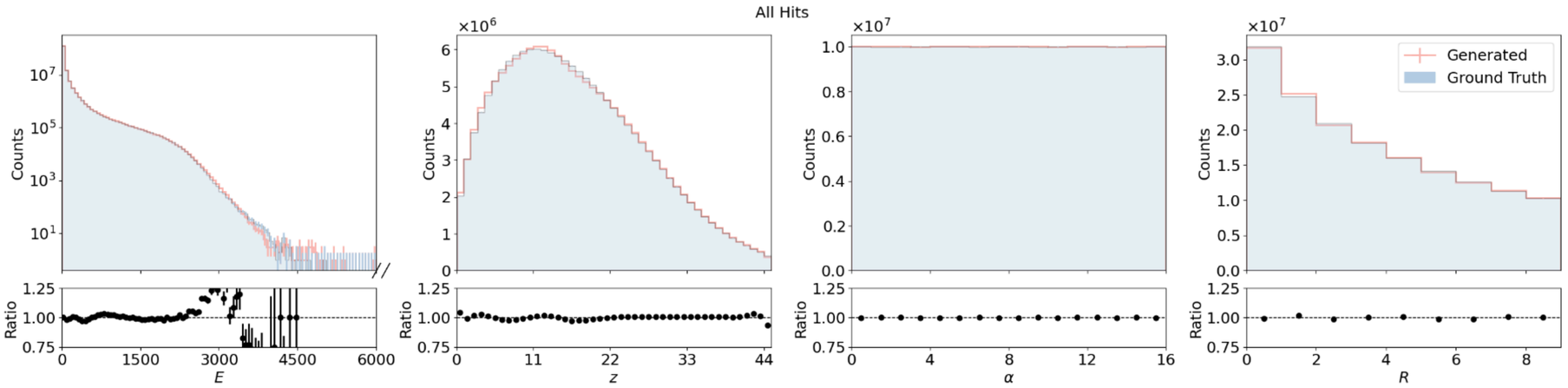
Pay Attention to Mean Fields for Point Cloud Generation

Benno Käch [arXiv: 2305.15254]

Top-quark JetNet150



CaloChallenge Dataset 2

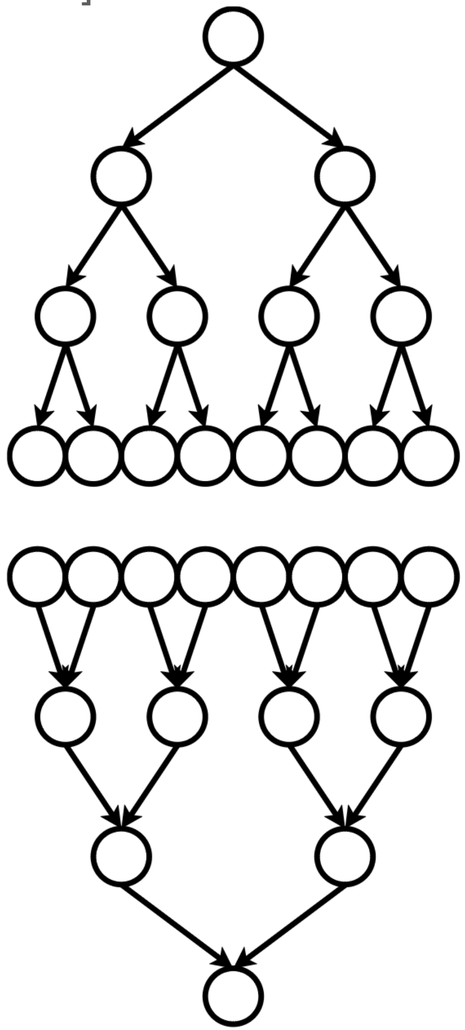


DeepTreeGAN

Moritz Scham

[arXiv:2311.12616]
[arXiv:2312.00042]

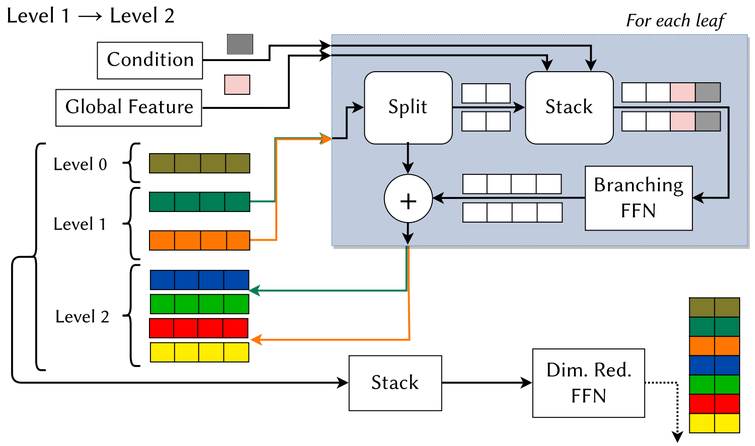
- Generative Adversarial Network
- Generator and Critic have Tree Structure
- Generator:
  - fixed branching + message passing at each level
  - slicing at the end
- Discriminator
  - Bipartite Pooling to fixed point cloud
  - Size reduction of point cloud
  - seperate loss term for each level



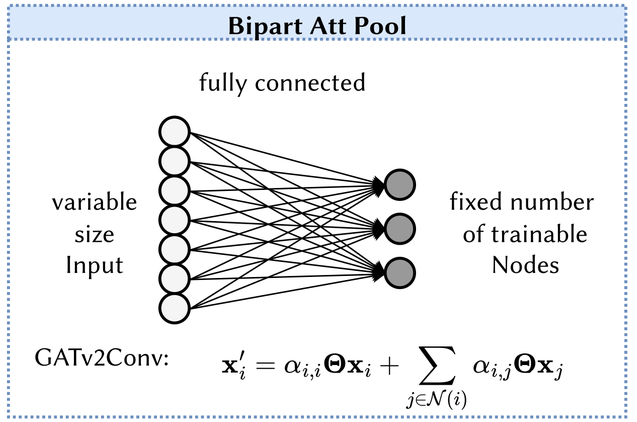
Generator

Critic

Branching



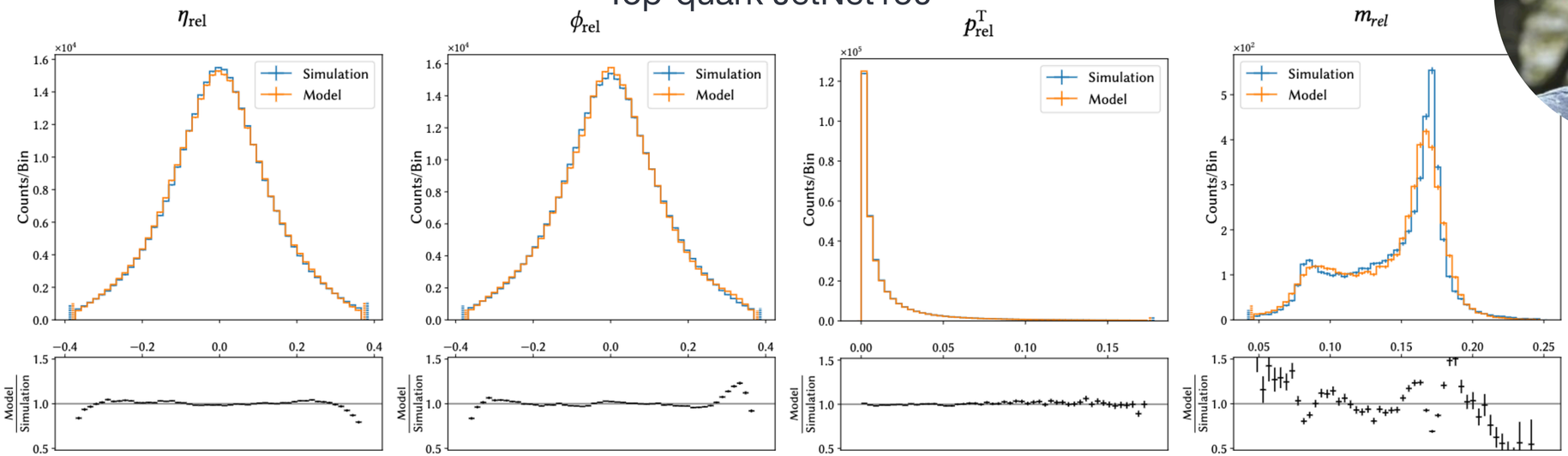
Pooling



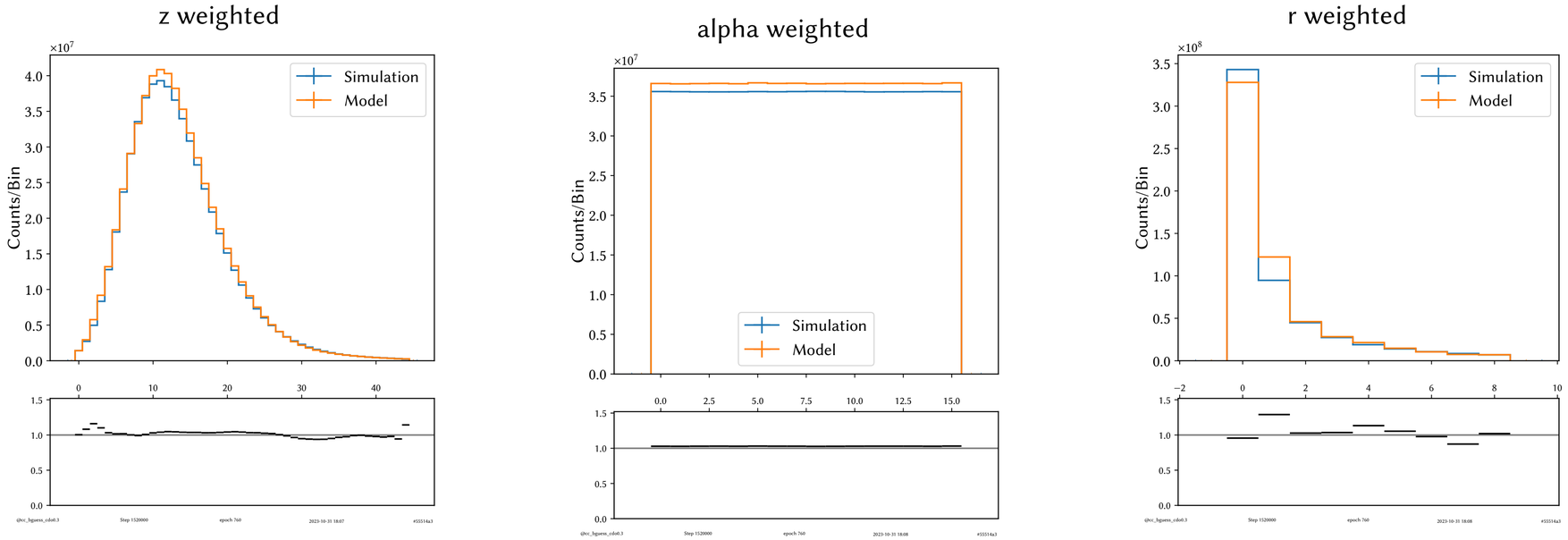
DeepTreeGAN

Moritz

Top-quark JetNet150



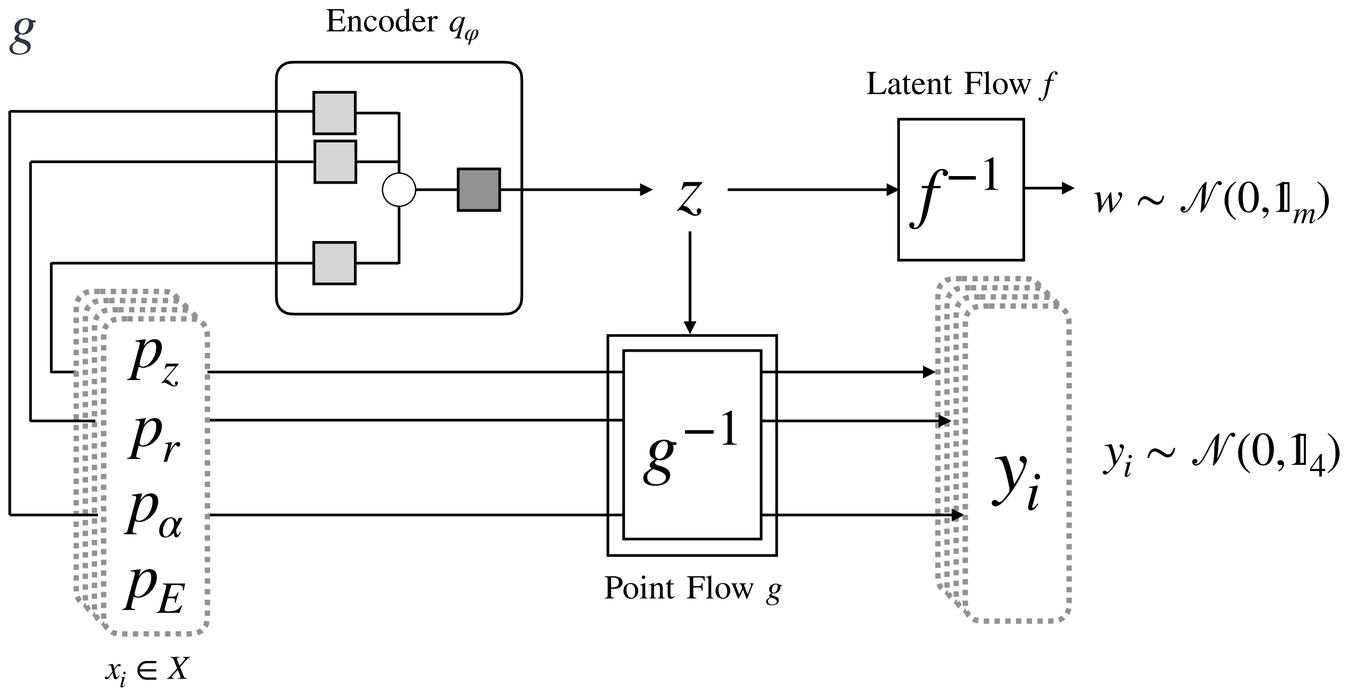
CaloChallenge Dataset 2

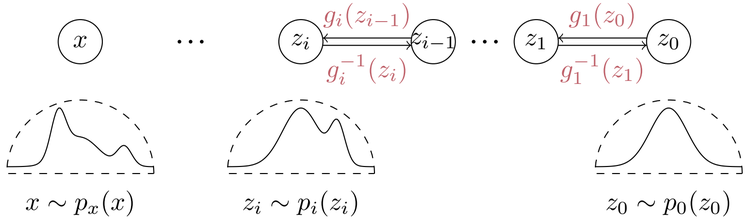


CaloPointFlow

Simon Schnake

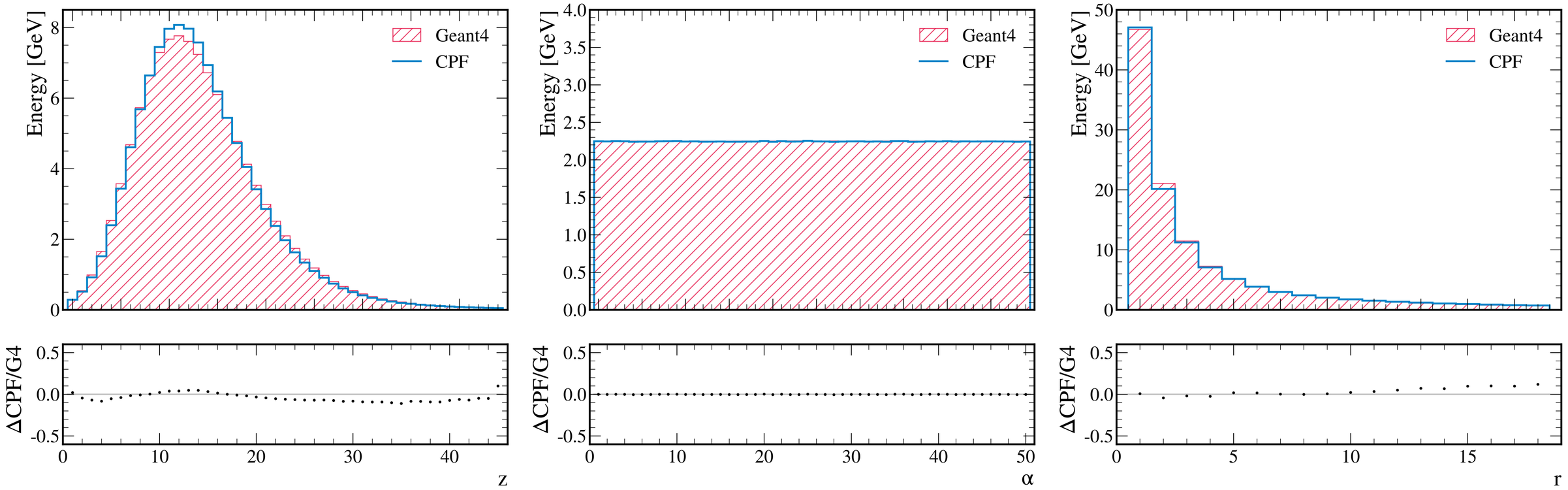
- Point-wise normalizing flow Point Flow g
- Shower shaped encoded in z
- Shower shape is generated by the Latent Flow f





CaloPointFlow

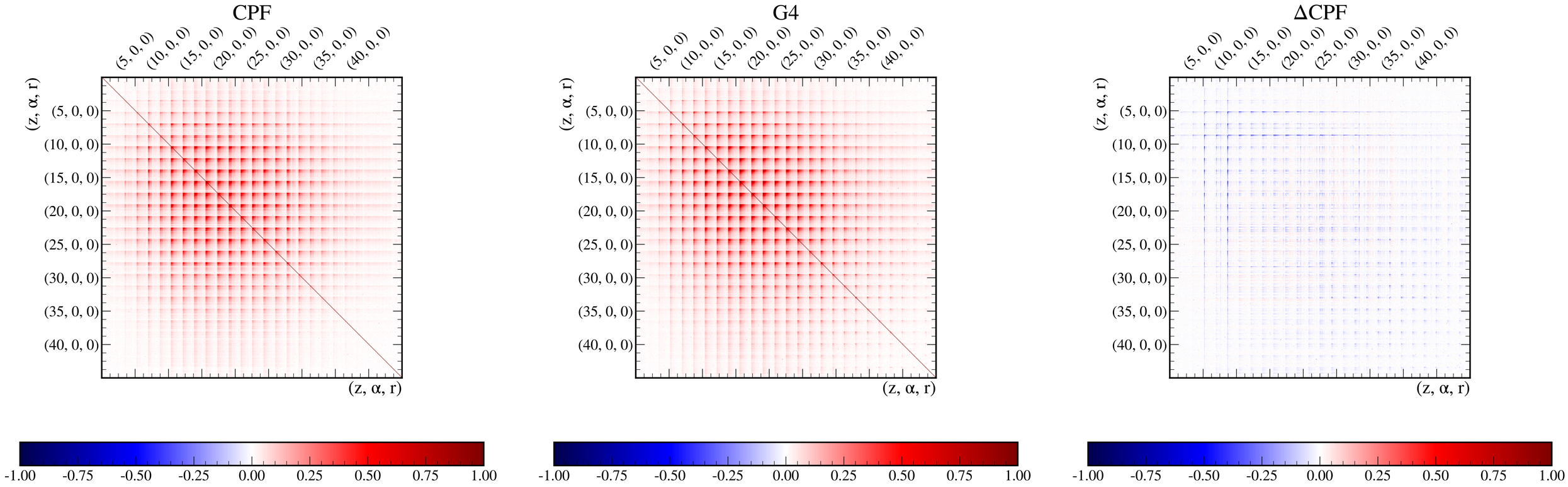
Simon Schnake



CaloChallenge Dataset 3

CaloPointFlow

Simon Schnake



CaloChallenge Dataset 3

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]
  - Moritz: DeepTreeGAN [arXiv:2311.12616, arXiv:2312.00042]
  - CaloPointFlow
- Benno and Moritz will present there work next week at the NeurIPS ML4PS Workshop
- Results of all models on the CaloChallenge will be in the summary paper to be published soon

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]
  - Moritz: DeepTreeGAN [arXiv:2311.12616, arXiv:2312.00042]
  - CaloPointFlow
- Benno and Moritz will present there work next week at the NeurIPS ML4PS Workshop
- Results of all models on the CaloChallenge will be in the summary paper to be published soon

Thank you for your attention!