CaloDiT: Diffusion with transformers for fast shower simulation

Renato Cardoso[1], Vijay Ekambaram[3], Kristina Jaruskova[1], Jayant Kalagnanam[2], Nam Nguyen, Piyush Raikwar[1], Dalila Salamani[1], Mudhakar Srivatsa[2], Sofia Vallecorsa[1], Kyongmin Yeo[2], Anna Zaborowska[1]

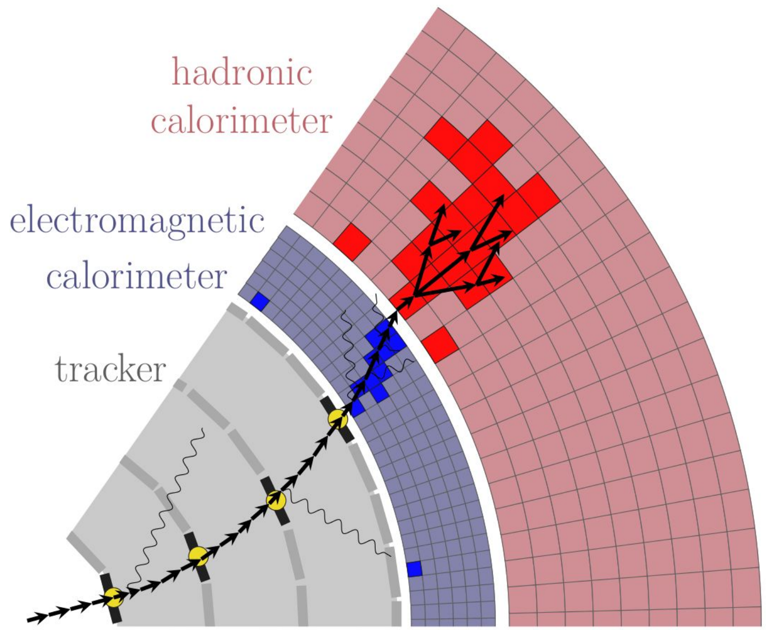
[1]CERN, Geneva, Switzerland
[2]IBM T. J. Watson Research Center, Yorktown Heights, NY USA
[3]IBM Research, India

ACAT 2024
Stony Brook, Long Island NY, USA

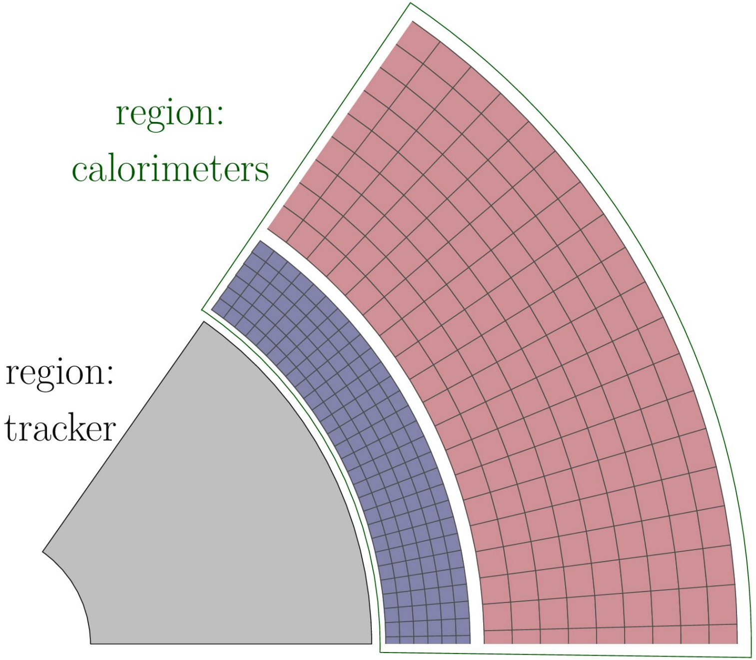
13.03.2024

Fast shower simulation

FullSim



FastSim



Motivation

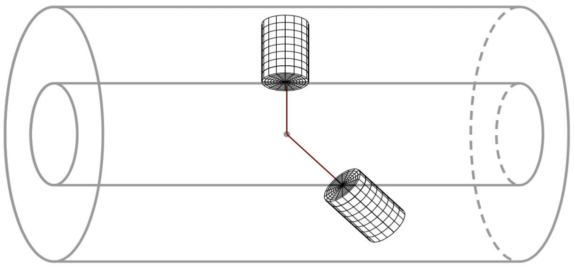
- Development of machine learning models for fast shower simulation is computationally expensive.
- Moreover, designing model for *each experiment* requires dedicated expertise.

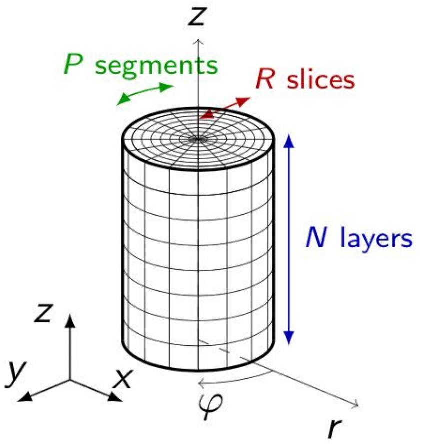
Make FastSim easily available without access to ML expertise.

1. *Generic energy scoring mesh [guide]*
   - Collect energy irrespective of the detector geometry.
   - Ready to use models. (Requires training)
2. *Generalizable ML model*
   - **Train once** on very large & diverse datasets to learn rich representations.
   - **Then adapt** to new detectors, quickly.

Energy scoring

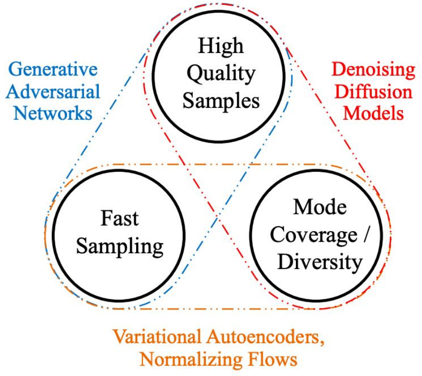
A detector agnostic mesh is constructed to contain the largest shower.





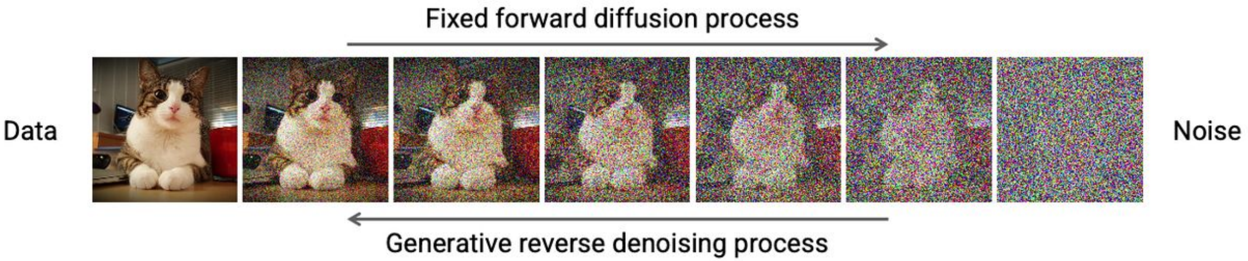
- The mesh aligns with the direction of incident particle.
  - The direction, i.e., the angles are recorded.
- The size of the cells can vary across detectors according to its X_0 & R_M, but the number of cells remains constant[1].
- Explored in LHCb [poster on 14th].

[1] i.e., for a model



Generative model

We use a **diffusion model** for higher accuracy and higher diversity.



As for the architecture, we apply **transformer** blocks.

- A **generalized architecture** that works with any type of data, e.g., text, images, audio, etc.
- Models long-range dependencies (**Attention** mechanism).

https://arxiv.org/abs/2112.07804

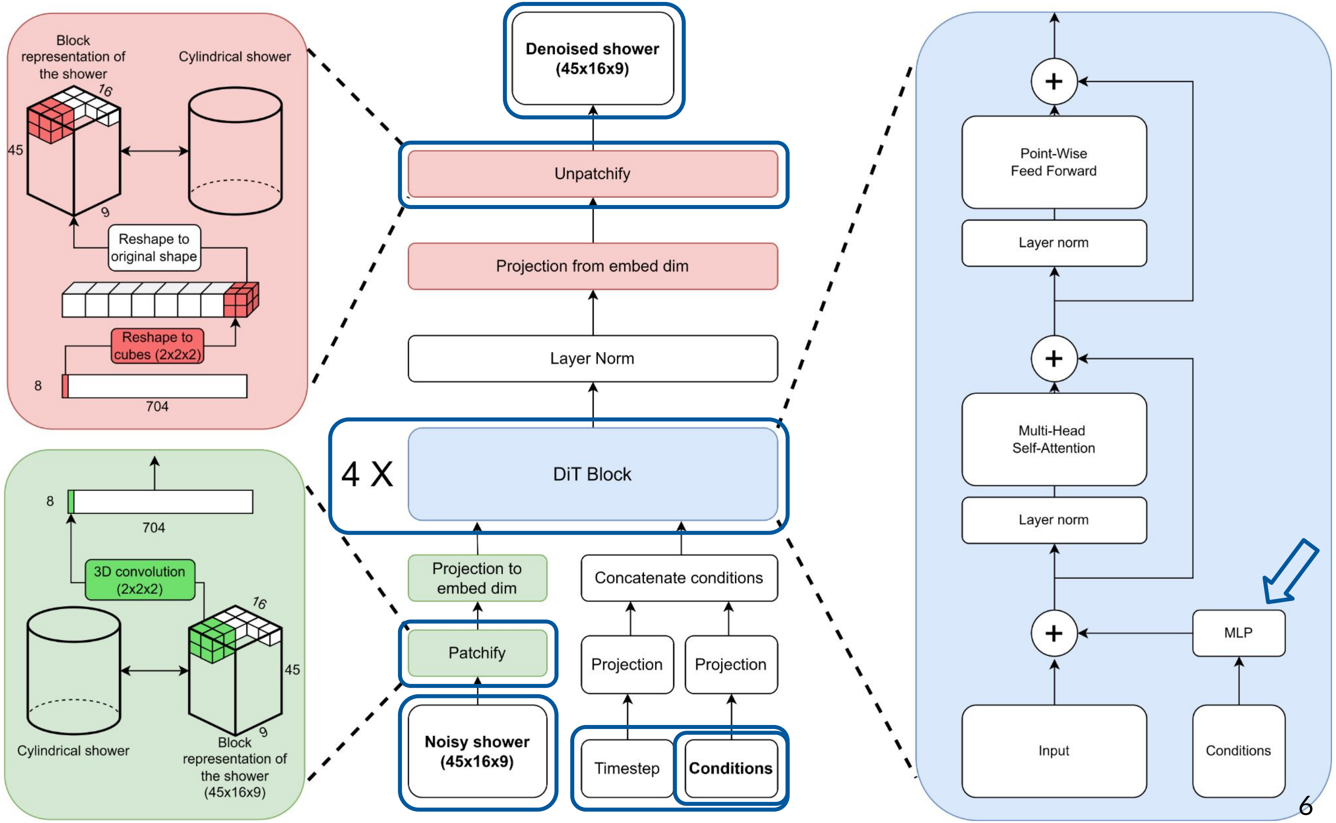
Model architecture

Conditions:

1. Energy (1GeV to 1TeV)
2. Phi (azimuthal angle) - 0 to 2π
3. Theta - 0.87 to 2.27
4. Geometry

The model:

- Diffusion steps = 400
- Embed dim = 144
- Cosine scheduler



https://arxiv.org/abs/2212.09748

Experiments

1. Training on single geometry
   - Par04 (CaloChallenge[1]), simplistic cylindrical geometry
   - 1M samples
2. Joint training on multiple geometries
   - Par04 and [Open Data Detector] (ODD, realistic geometry)
   - 1M samples each
   - Geometry condition - one hot encoding
3. Adaptation on [FCCeeALLEGRO]
   - Checkpoint from 2
   - Baseline - training from scratch
   - 100K, 200K, 400K samples in each case

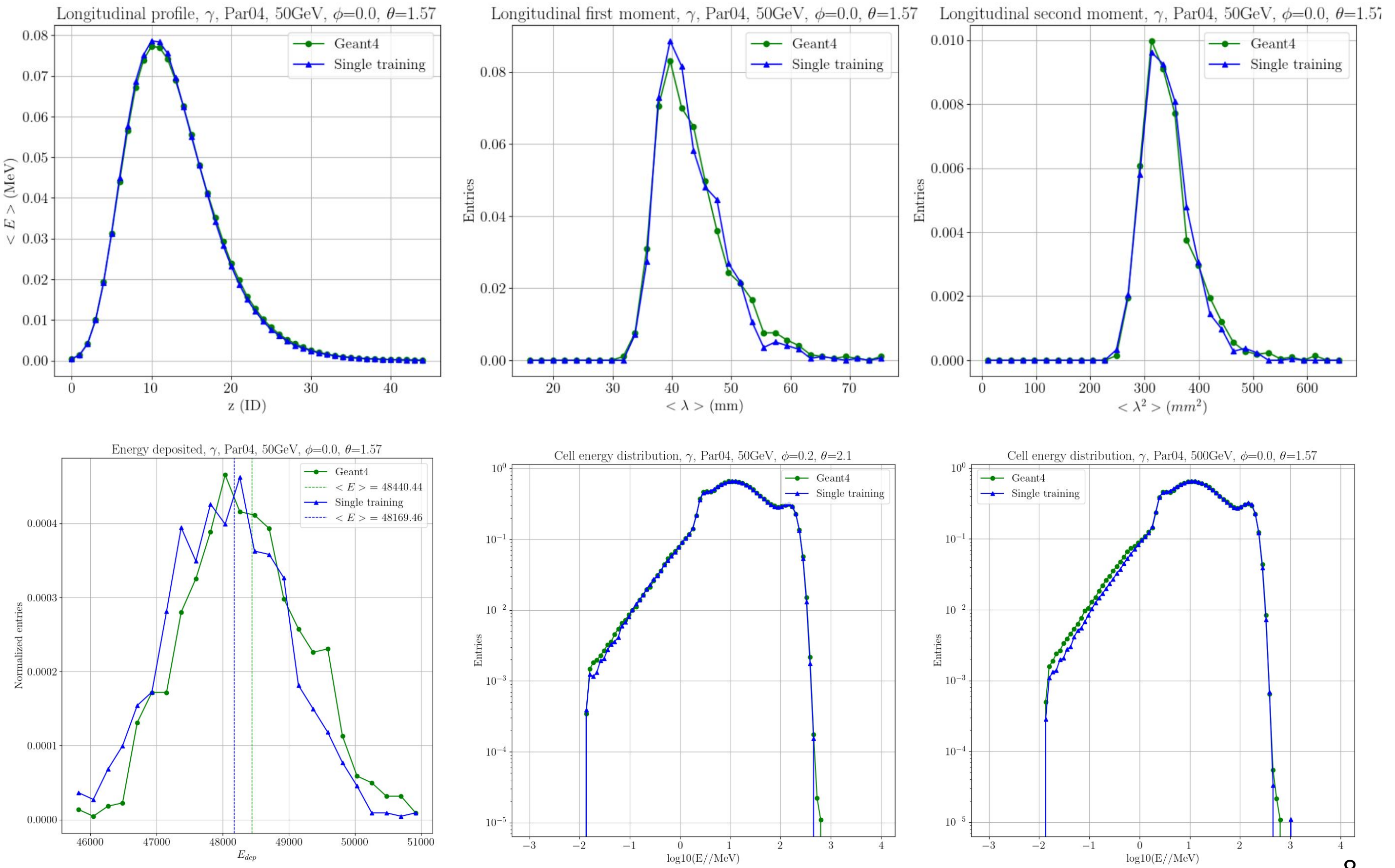
Note: The results are preliminary

[1] More samples and more conditions (phi, theta) compared to CaloChallenge

1. Training on single geometry

Par04

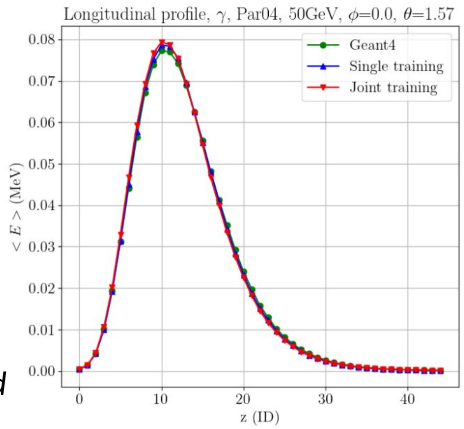
- Good accuracy with diffusion models
- Even cell energy

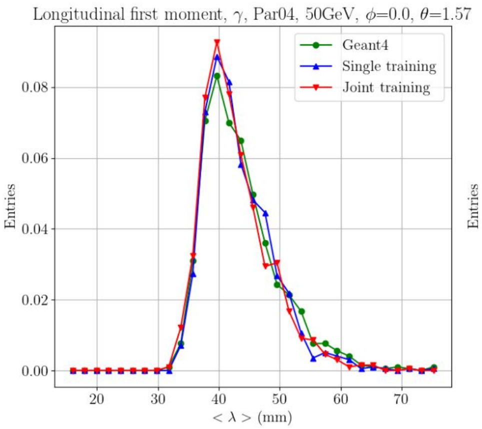


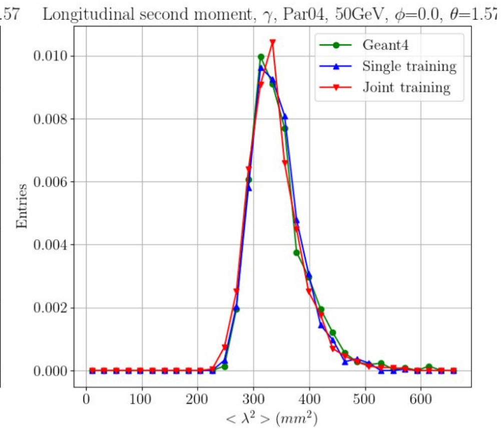
2. Joint training

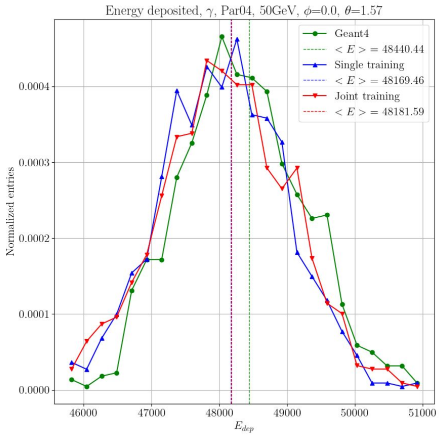
Par04

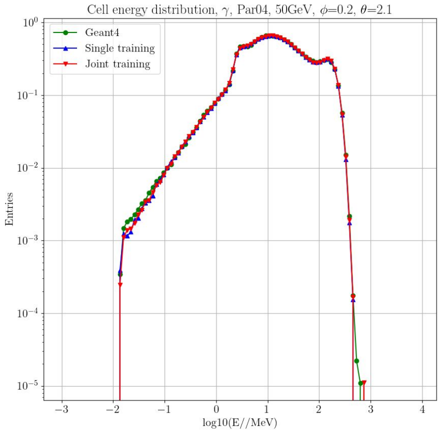
- Almost no degradation after adding geometry condition
- *Even if there was, not intended to use directly*

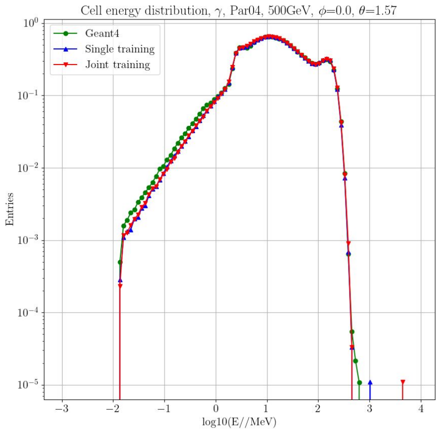












3. Adaptation

FCCeeALLEGRO

250 epochs for training from scratch
20 epochs for adaptation

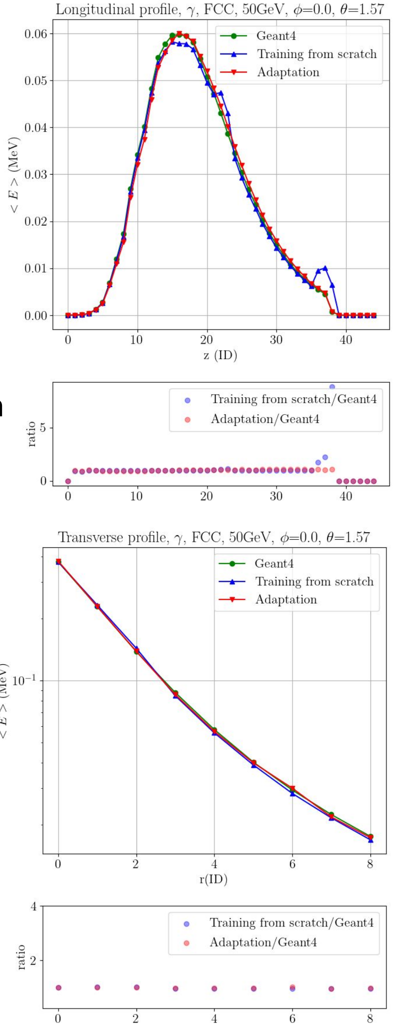
At 200K samples

~25x less training time

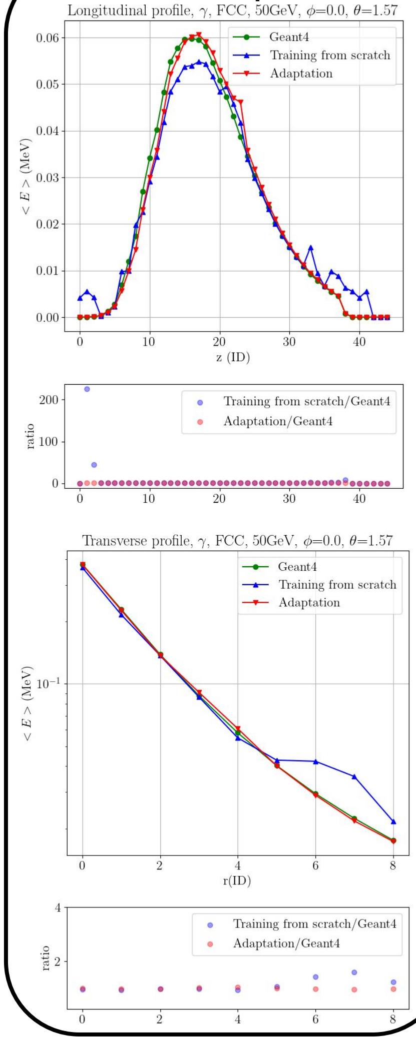
<50% of the data

Preliminary results

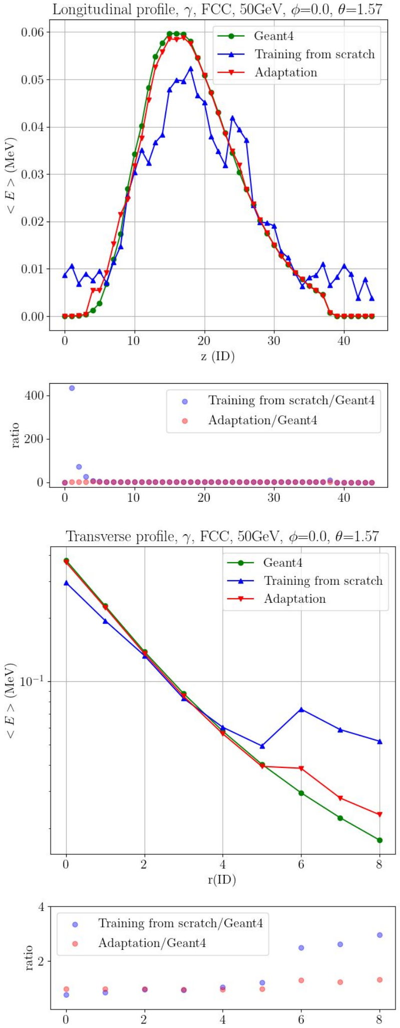
400K samples



200K samples



100K samples



Conclusion

- We present a framework for detector agnostic fastsim model, which can be easily adapted to new detectors.
- Initial results are very promising which significantly reduces the required statistics, and training time from days to just a couple of hours[1].

Future work:

- Pretraining on more geometries.
- Improvements revolving diffusion method and model architecture.
- Faster inference via distillation.
- Testing our framework in experiments.
    - The mesh is already implemented in Gaussino and DD4hep.
    - Work started for ATLAS

[1] Depending on the model size of course

Thank you for listening!

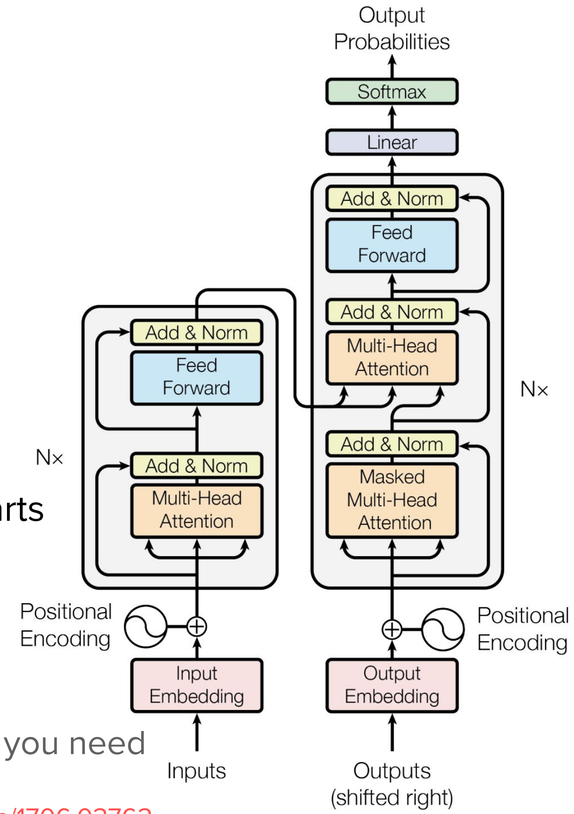
Questions?

piyush.raikwar@cern.ch

Backup

Transformer

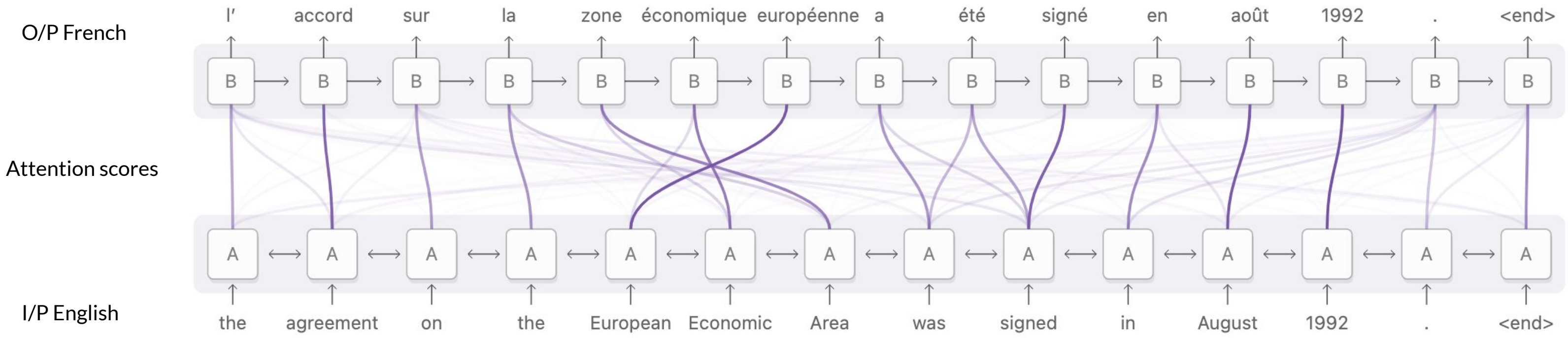
- Proposed for sequence-to-sequence tasks.
- I/O is any type of sequences.
- Encoder-Decoder blocks.
- Positional embeddings.
- **Attention:** Dynamically focus on important parts in the input.
- Multi-headed attention.



Attention is all you need

https://arxiv.org/abs/1706.03762

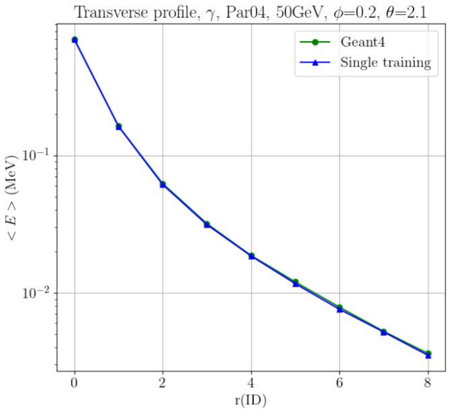
Attention in transformers

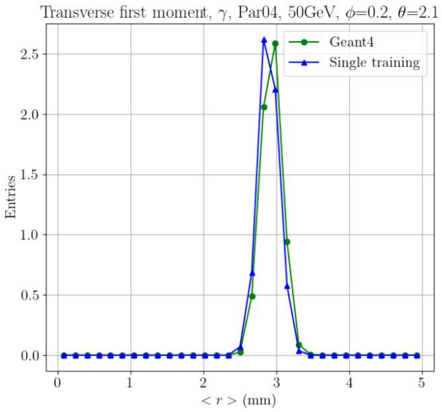


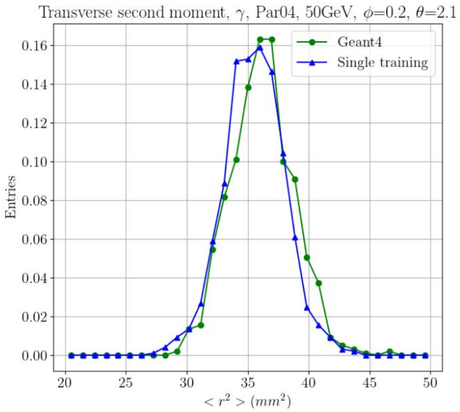
- **Dynamically focuses on important parts** in the input.
- Helps in modelling correlations between energy deposits.

Transverse profiles - Par04

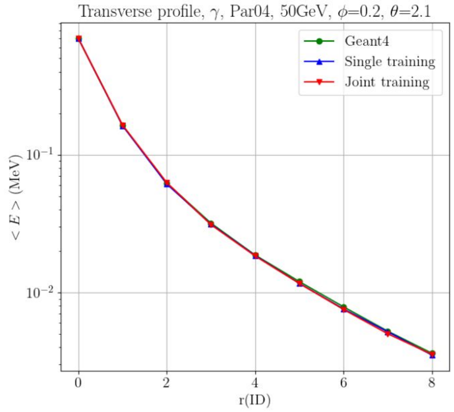
Single geometry training

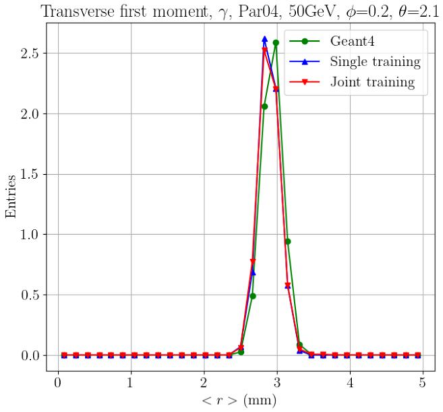


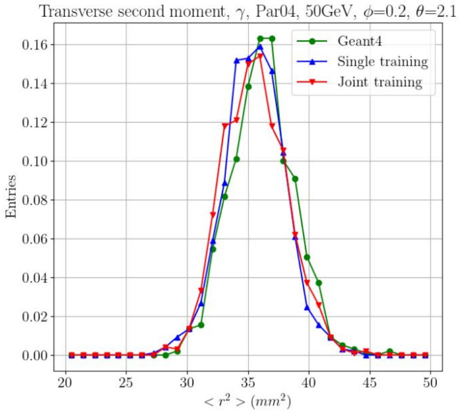




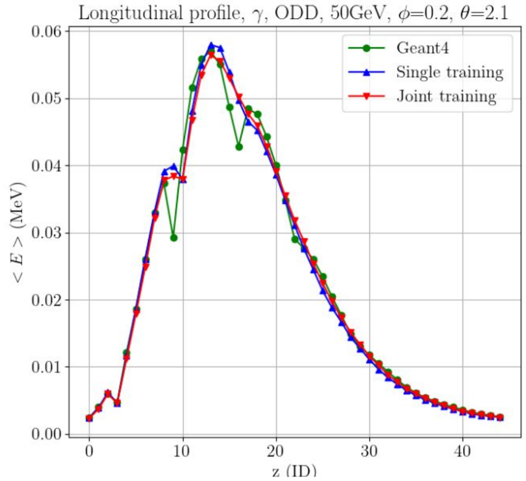
Joint training

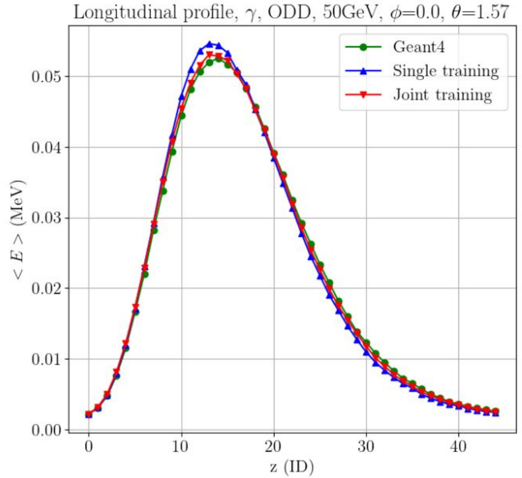


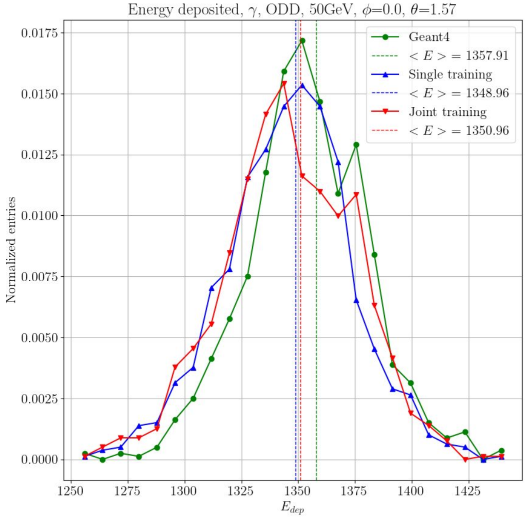


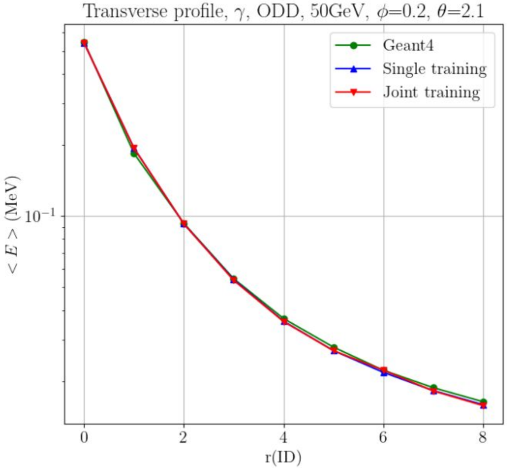


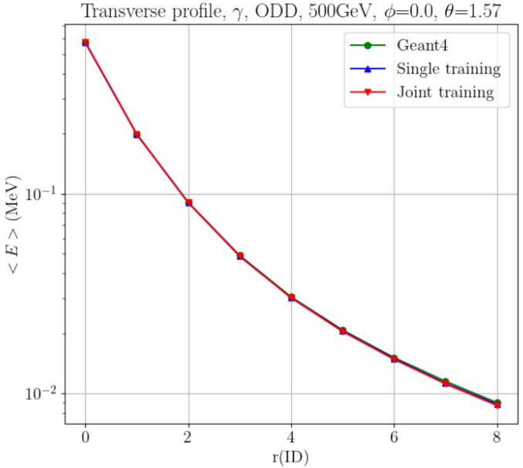
Joint training - ODD

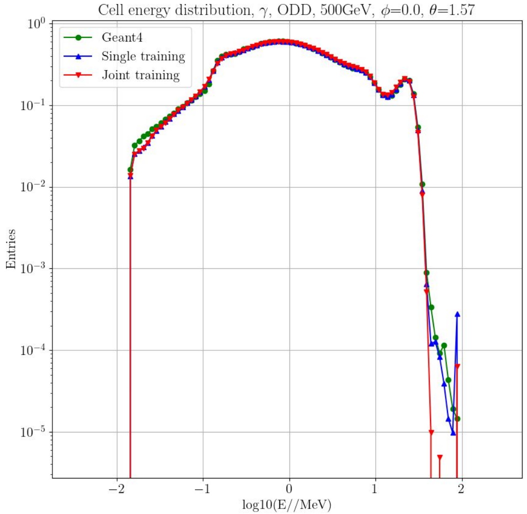


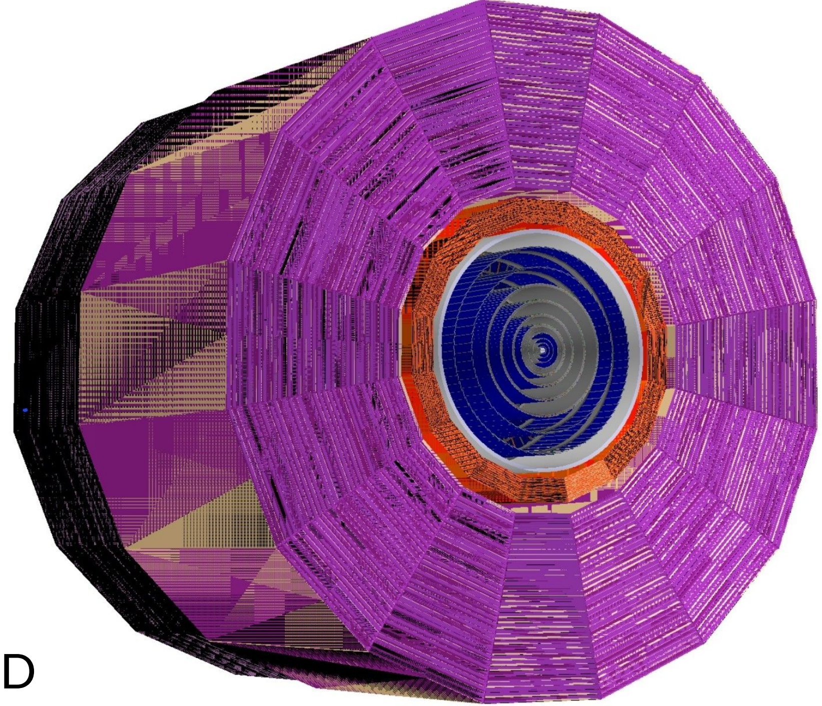




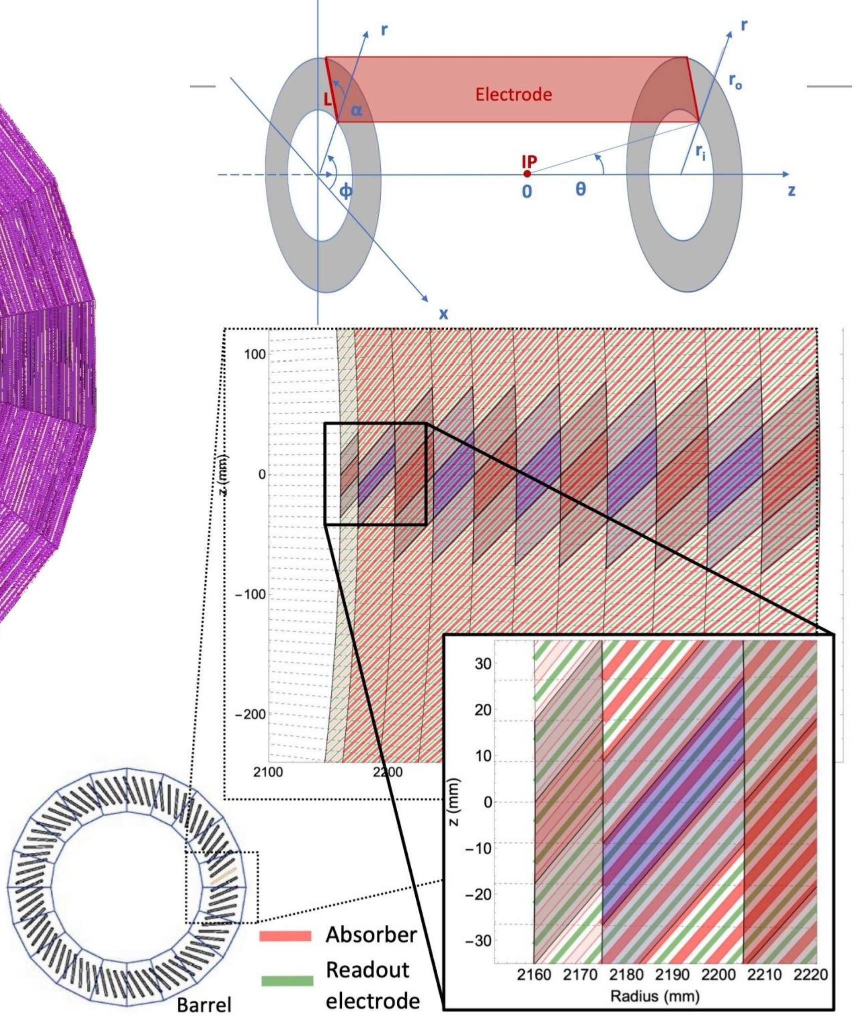








ODD



FCCeeAllegro