We **reconstructed photons** from energy deposits by **transforming** particle reconstruction.

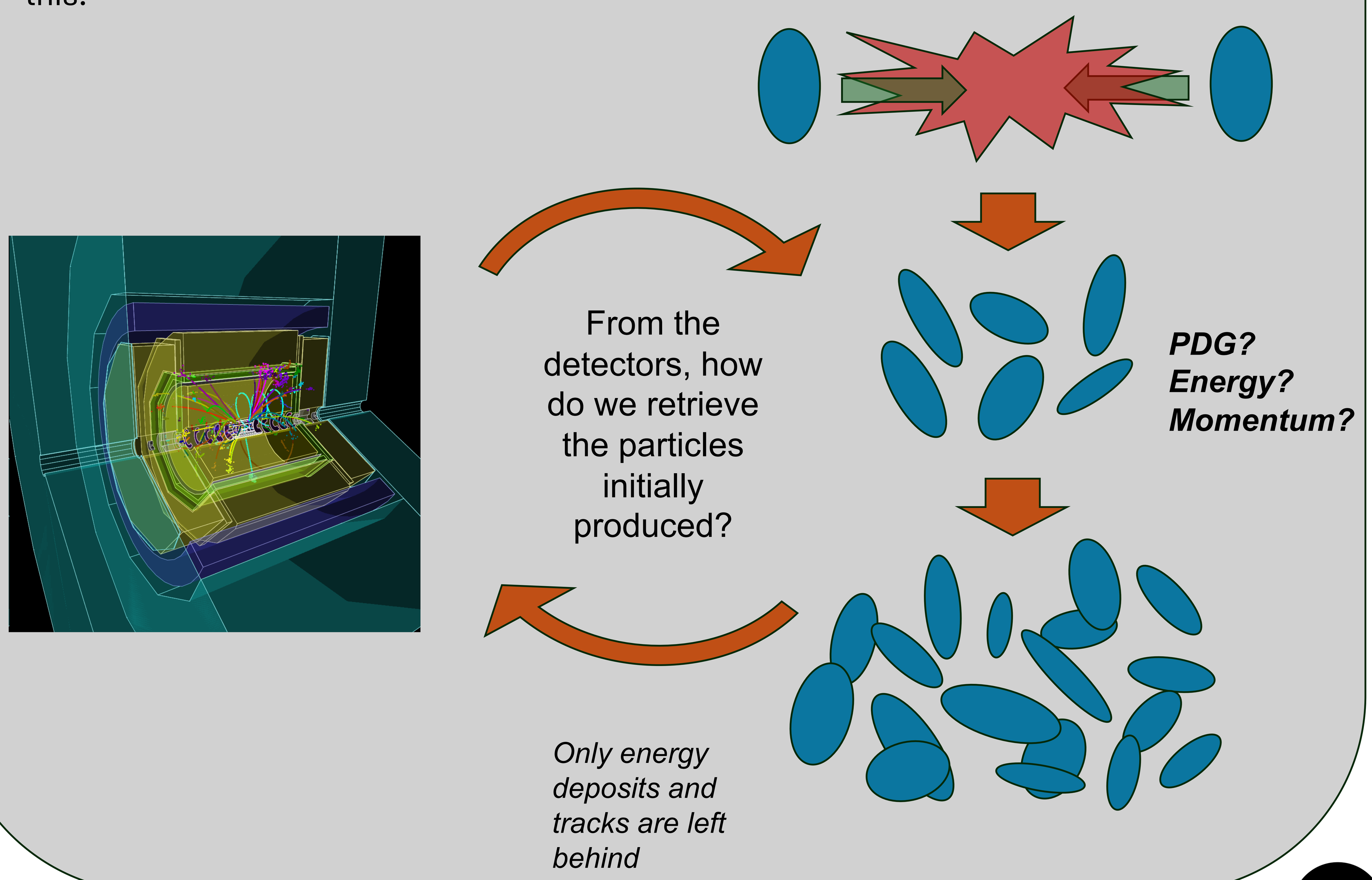
Particle-flow reconstruction with Transformer

Taikan Suehara[1], Paul Wahlen[2] [1]ICEPP, The University of Tokyo. [2] IP Paris, ETHZ, ICEPP Intern

1

Background:

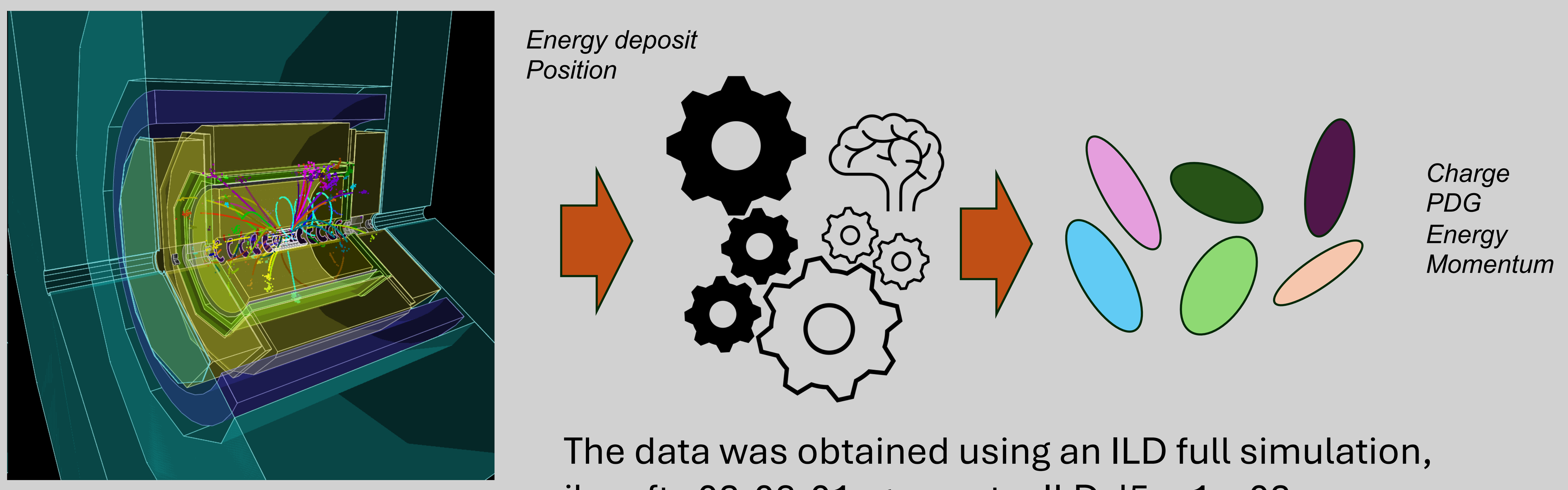
Due to the interaction between particles and matter, the initial energy of an incoming particle into a detector is scattered in hits located at different positions, forming a cluster. This renders the cluster of reconstructing the initial particle notoriously difficult and several algorithm, called Particle Flow Algorithms (PFAs), have been developped to tackle this.



2

Methodology:

Using a network architecture first developped in language models, a Transformer, we are re-constructing particles using only the energy deposits and their positions in the calorimeters.

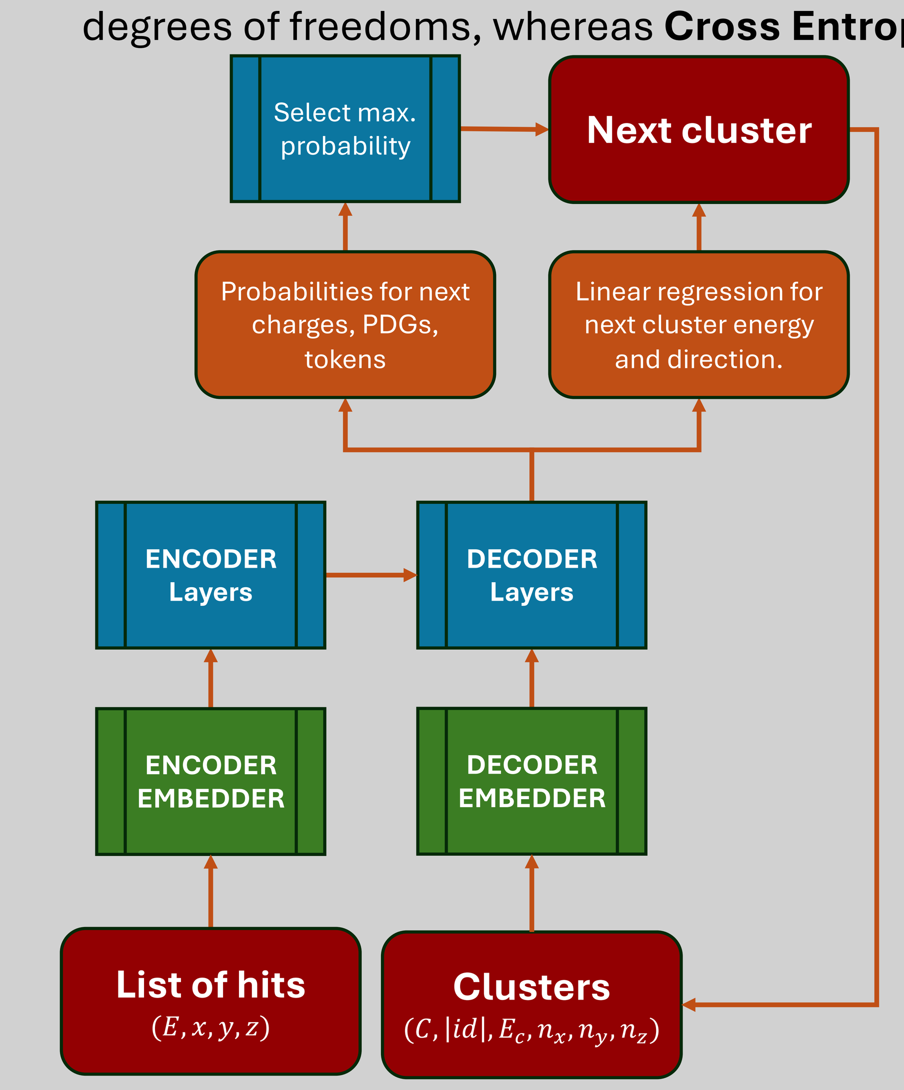


The data was obtained using an ILD full simulation, ilcsoft v02-03-01, geometry ILD_l5_o1_v02

3

Network Architecture:

The architecture is heavily inspired by the original Transformer detailled in *Attention is All you Need* by Vaswani et al. (2017). **Mean Square Error** is used for continuous degrees of freedoms, whereas **Cross Entropy** is used for discrete quantities.



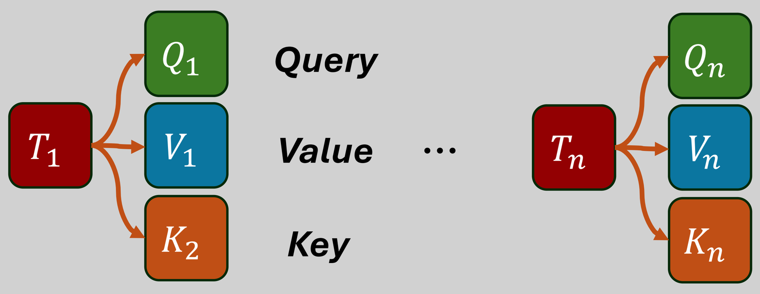
Hits are characterised by their positions in the calorimeters and energy deposits.

Clusters contain information of particles before passing through the detectors. These are:

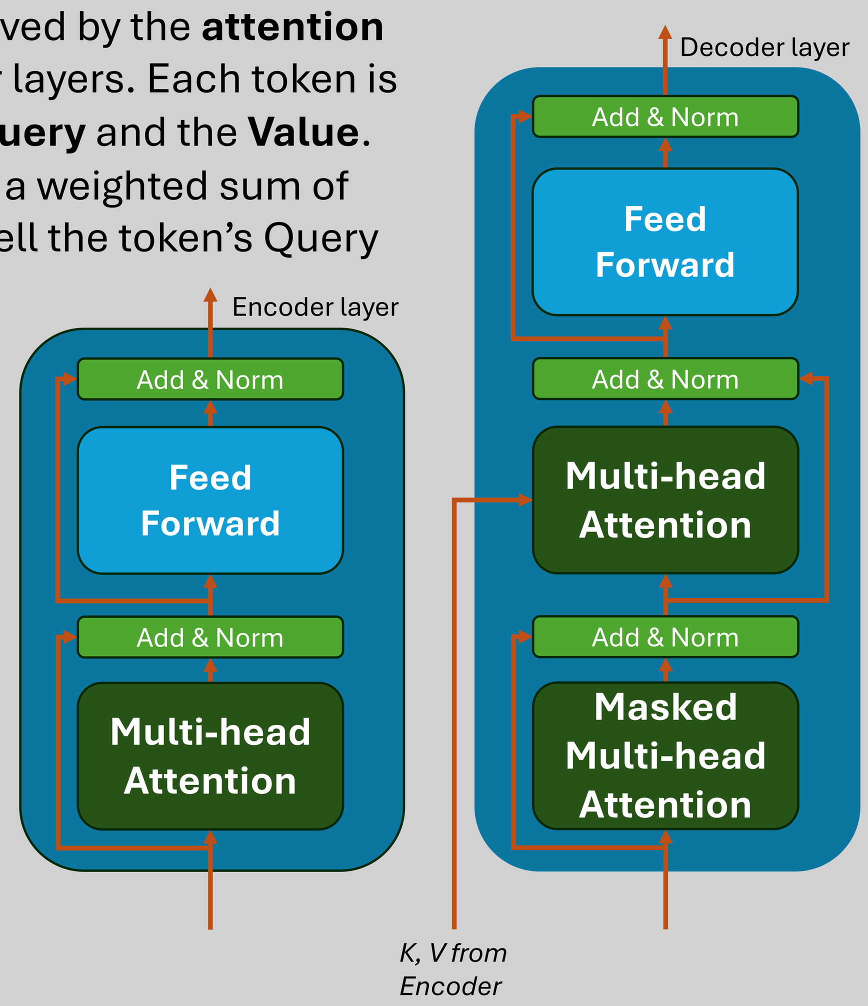
- Its charge, C
- The absolute value of its PDG number, $|id|$
- Its energy, E_c
- Its direction, (n_x, n_y, n_z)

At each **iteration**, the model predicts the **next cluster** by using information contained in the hits and clusters from **previous iterations**. The process is stopped when an end of sequence token is predicted.

Information in hits and clusters is retrieved by the **attention** mechanism in the encoder and decoder layers. Each token is projected onto 3 vectors: the **Key**, the **Query** and the **Value**. During attention, tokens are updated as a weighted sum of the Values, with weights given by how well the token's Query are aligned with the Keys.



$$T_i' = \mathrm{Attention}(Q_i, K_j, V_j) = \sum_{j=1}^{n} \mathrm{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d_k}}\right) V_j$$

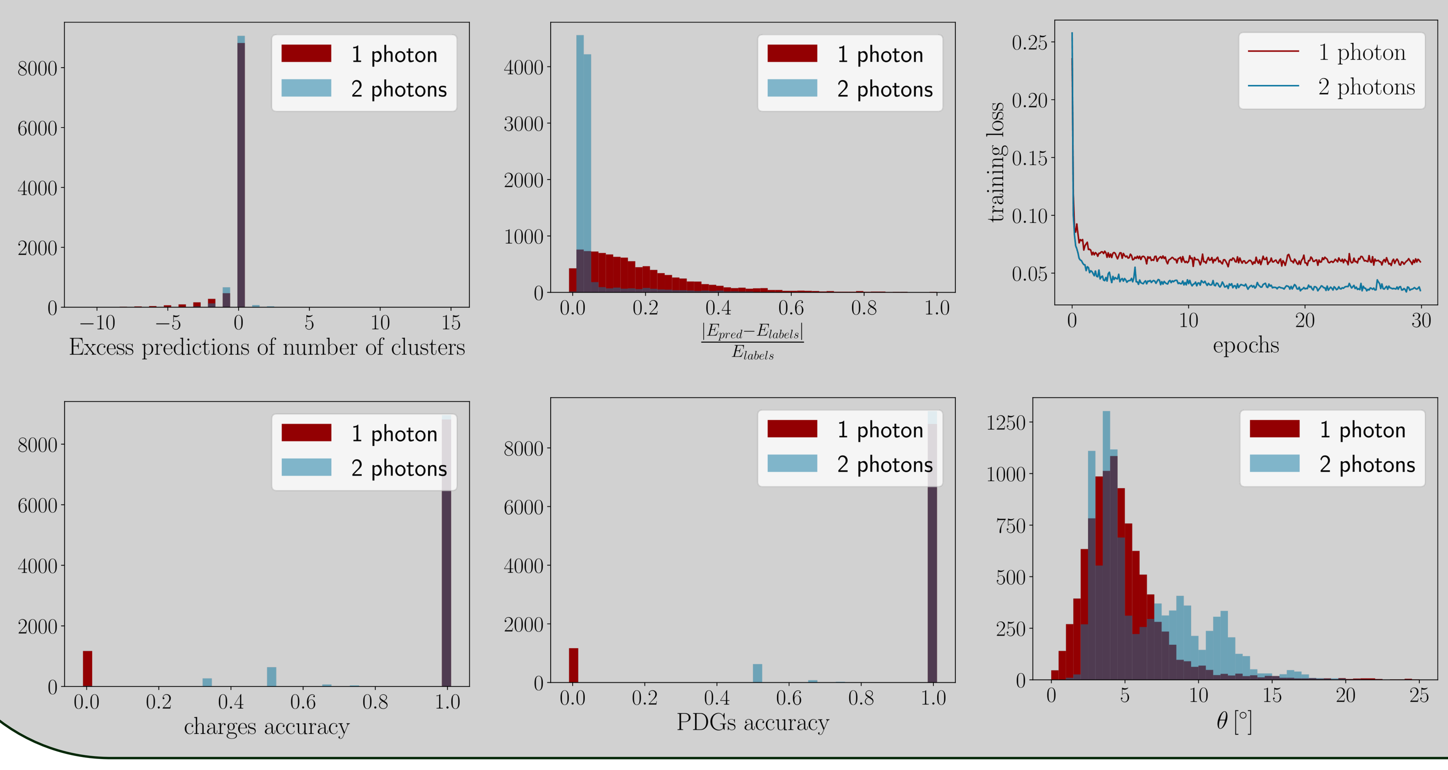


4

Results

- Models are tested against clusters generated by either a **single** or **two photons**
- Maximum accuracy is not achieved since photons can split into particles/antiparticles, etc...

Accuracies are calculated by a one-to-one comparison between overlapping predictions and labels



- **Single photons**: logarithmic distribution of 10 to 100 GeV particles in random directions. 100k events.
- **2 photons:** Fixed energy at 10 GeV, θ fixed at 85 degrees. ϕ random. The 2 photons are separated by an opening angle of 100 mrad.

5

Conclusion

- Current architecture shows reasonable results with simplified datasets of clusters formed by one or two particles

Perspective of future work

- Increasing the **complexity** of the dataset using multiple **taus or jets** to form the clusters
- Focusing on predicting the correct numbers of clusters first, trying different architectures and hyperparameters.

ICEPP
The University of Tokyo

More information and code can be found on the GitHub repository

Scan me!