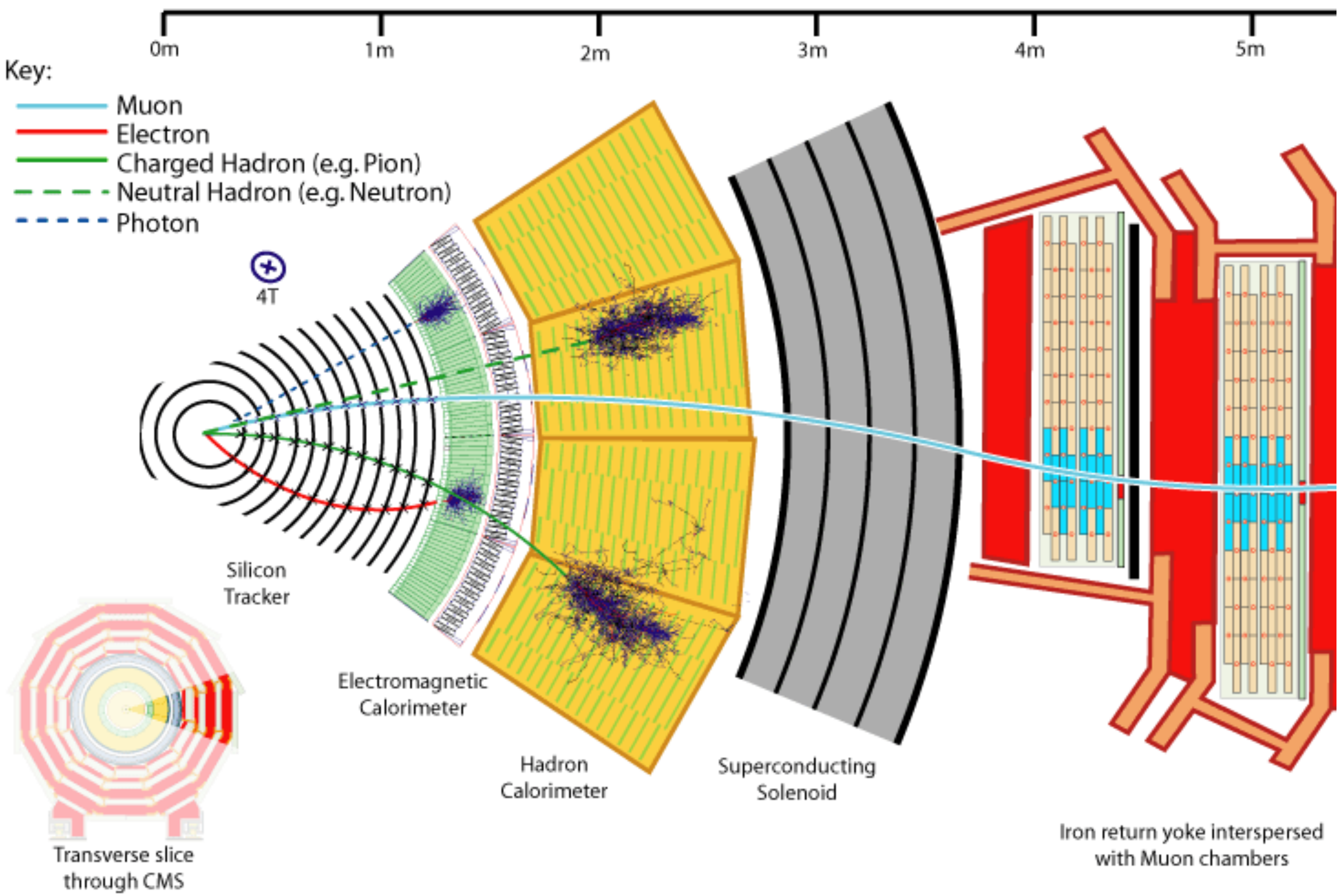
Scalable neural networks for event reconstruction at current and future colliders

Eric Wulff[1], Farouk Mokhtar[2], David Southwick[1], Mengke Zhang[2], Maria Girone[1], Javier Duarte[2], Joosep Pata[3]

[1]CERN, [2]UCSD, [3]NICPB

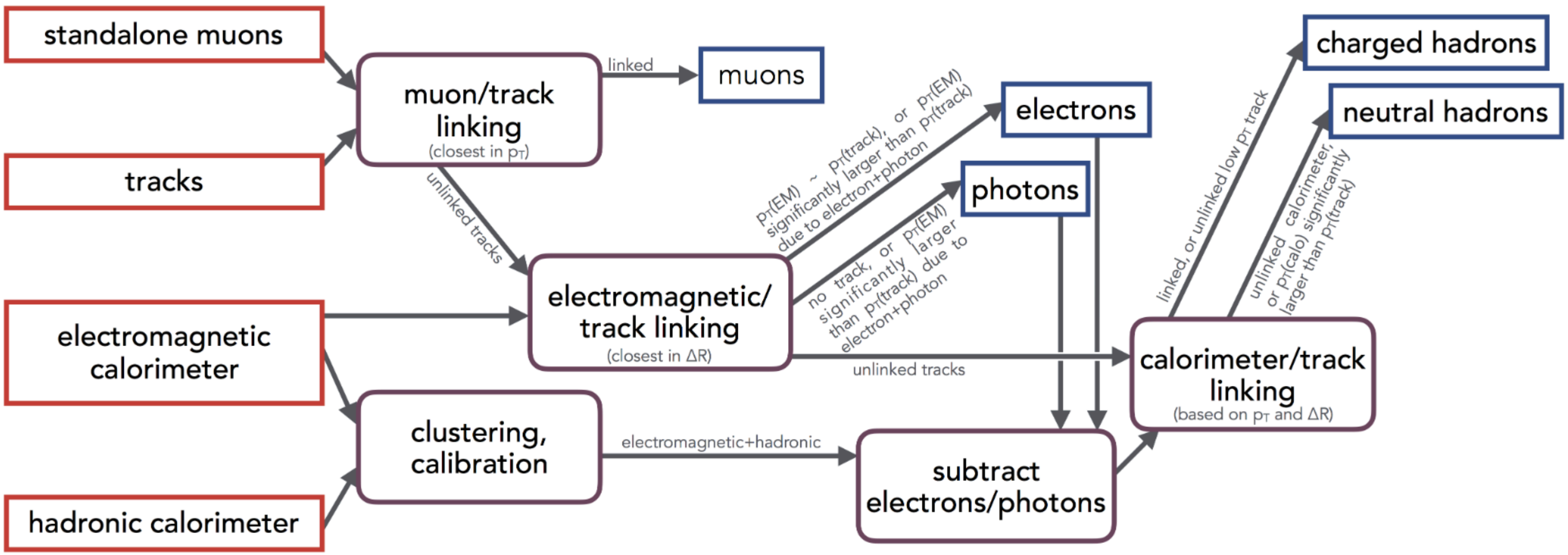
ACAT 2024 @ Stony Brook, March 14, 2023

Current and future multilayered detectors…



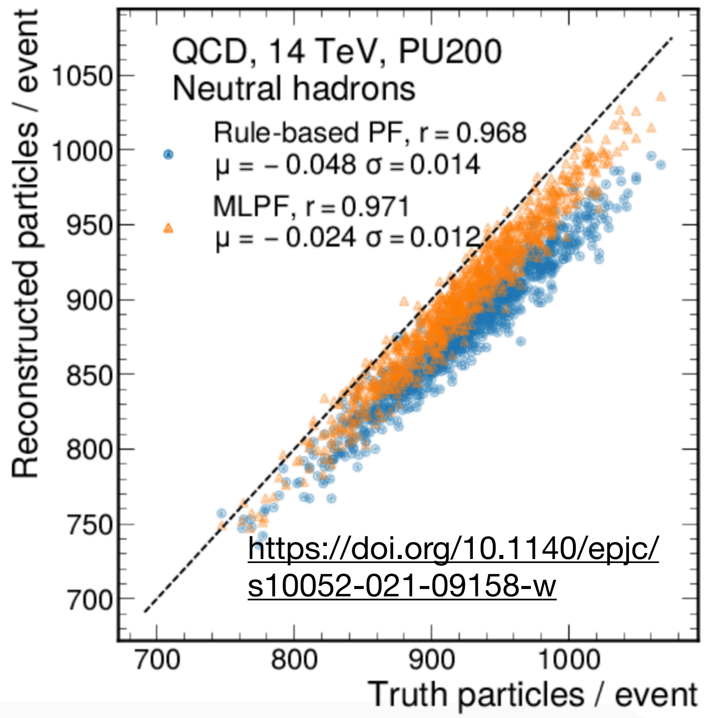
…need complex data reconstruction → particle flow algorithm

Existing full event reconstruction works very well…

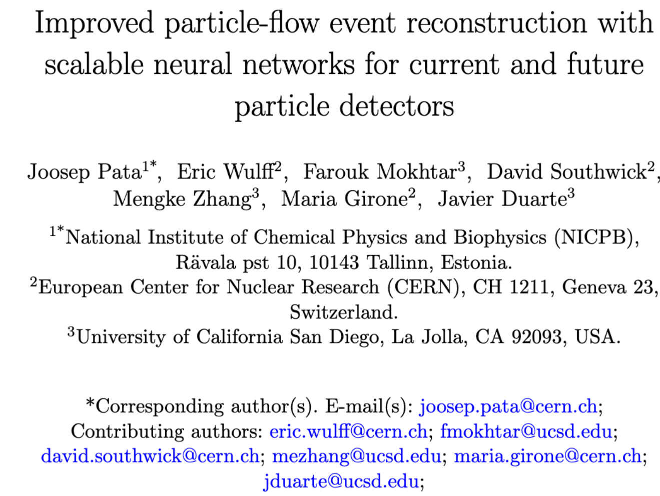


… but largely handcoded heuristics, can be tricky to extend, tune, port to new computational hardware.

proof of concept



event-level SOTA on full simulation


Improved particle-flow event reconstruction with scalable neural networks for current and future particle detectors

Joosep Pata[1*], Eric Wulff[2], Farouk Mokhtar[3], David Southwick[2], Mengke Zhang[3], Maria Girone[2], Javier Duarte[3]

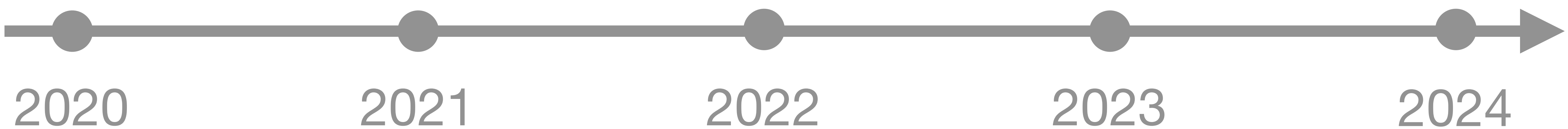
[1*]National Institute of Chemical Physics and Biophysics (NICPB), Rävala pst 10, 10143 Tallinn, Estonia.
[2]European Center for Nuclear Research (CERN), CH 1211, Geneva 23, Switzerland.
[3]University of California San Diego, La Jolla, CA 92093, USA.

*Corresponding author(s). E-mail(s): joosep.pata@cern.ch; Contributing authors: eric.wulff@cern.ch; fmokhtar@ucsd.edu; david.southwick@cern.ch; mezhang@ucsd.edu; maria.girone@cern.ch; jduarte@ucsd.edu;

https://doi.org/10.48550/arXiv.2309.06782

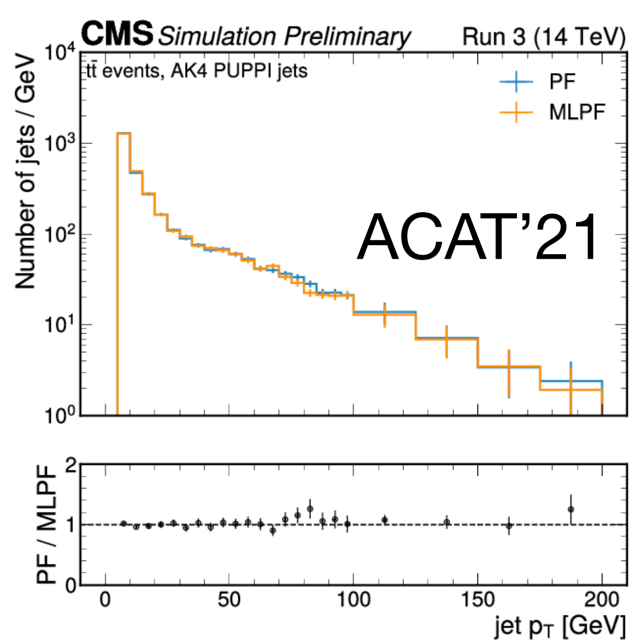
FastML'23, ML4Jets'23, ACAT'24

open data



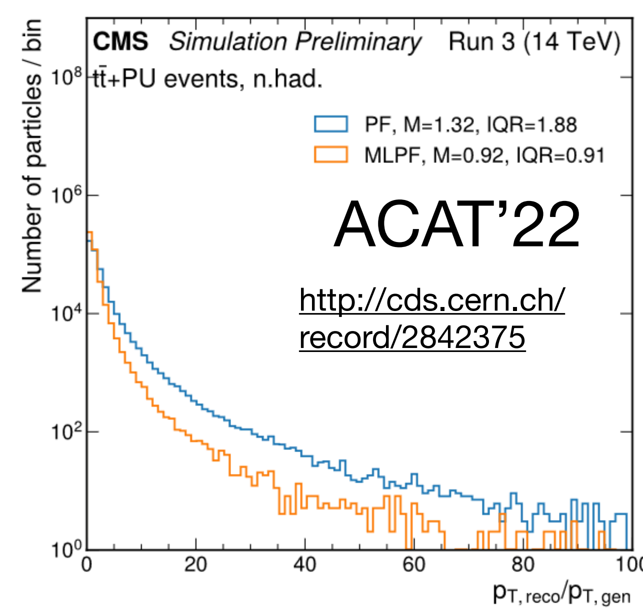
real detector

proof of concept

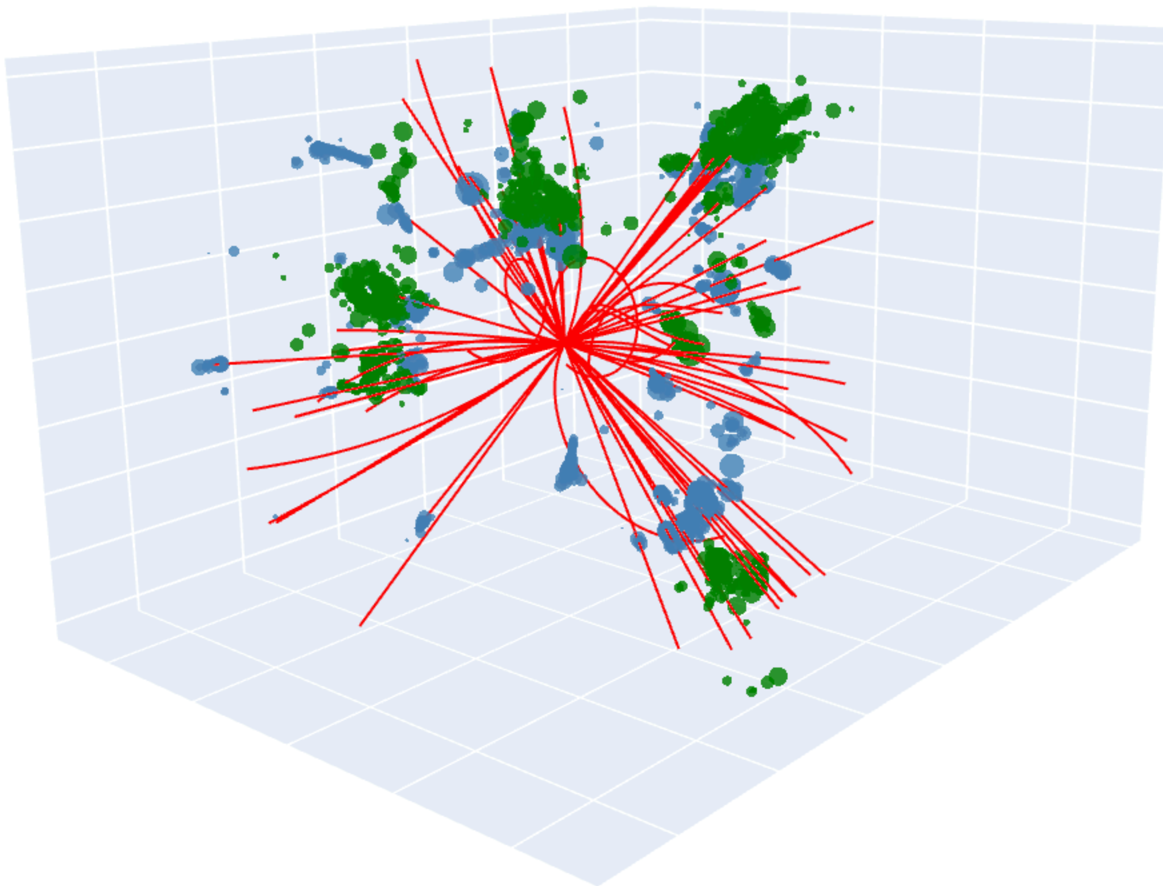


10.1088/1742-6596/2438/1/012100

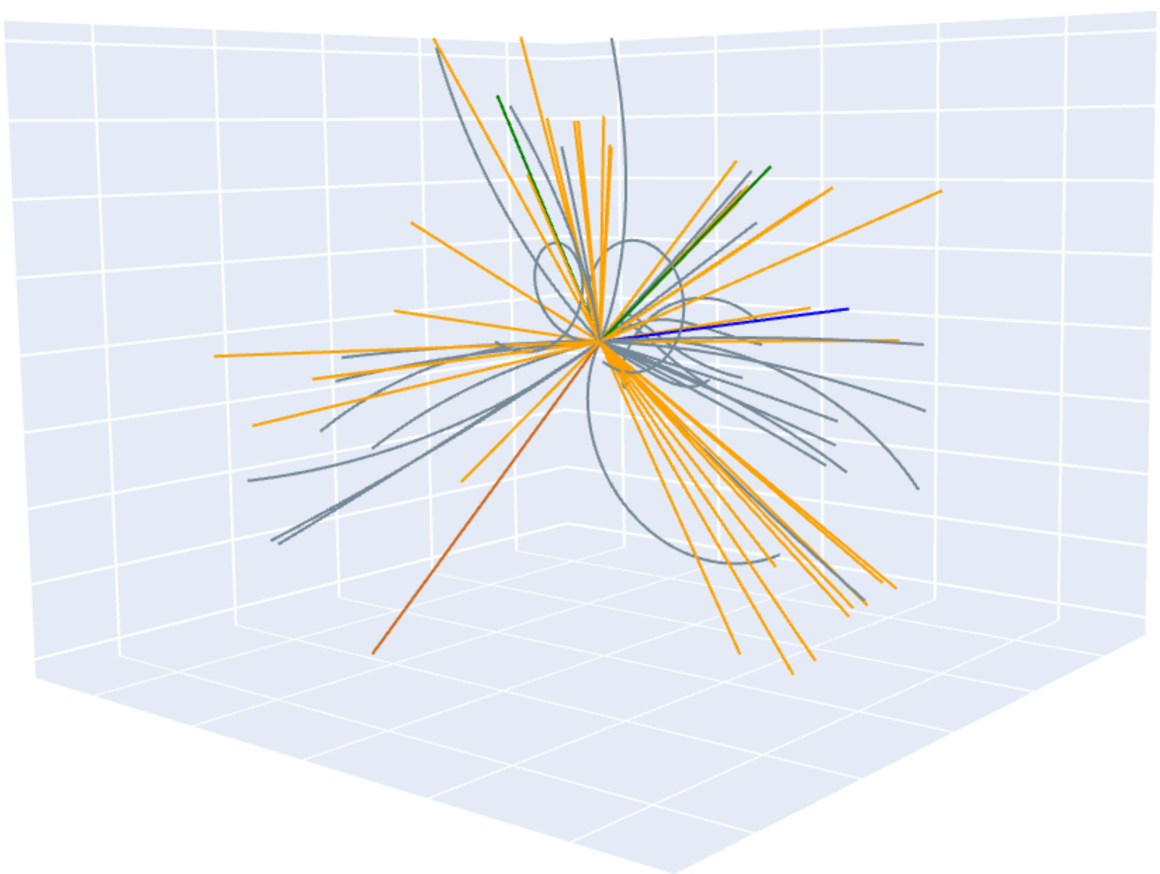
particle-level SOTA



One simulated event with full simulation



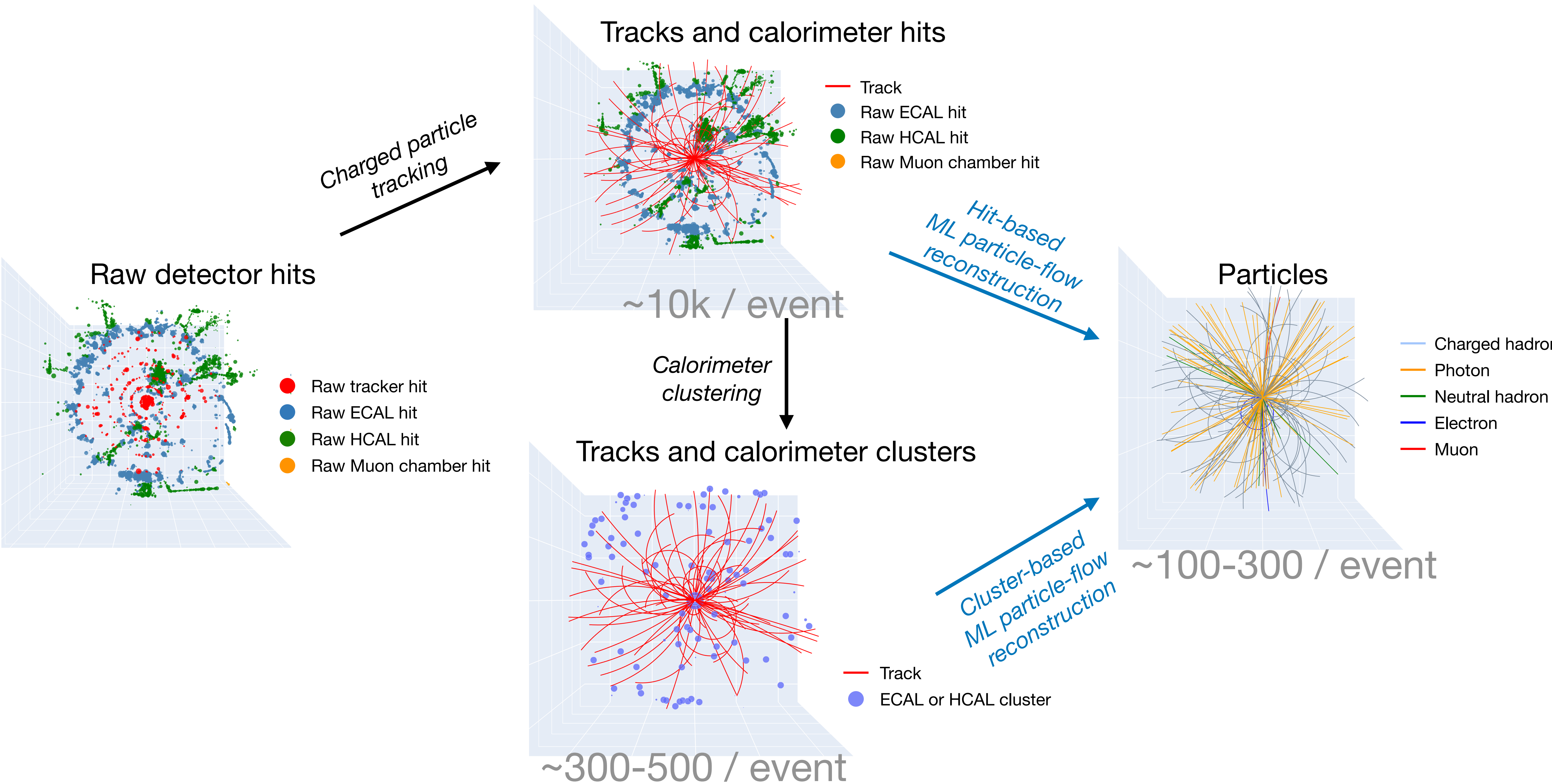
<u>tracks and hits</u>



<u>particles</u>

CLIC detector model, ddSim, Marlin reconstruction via Key4HEP

Highly diverse and granular dataset



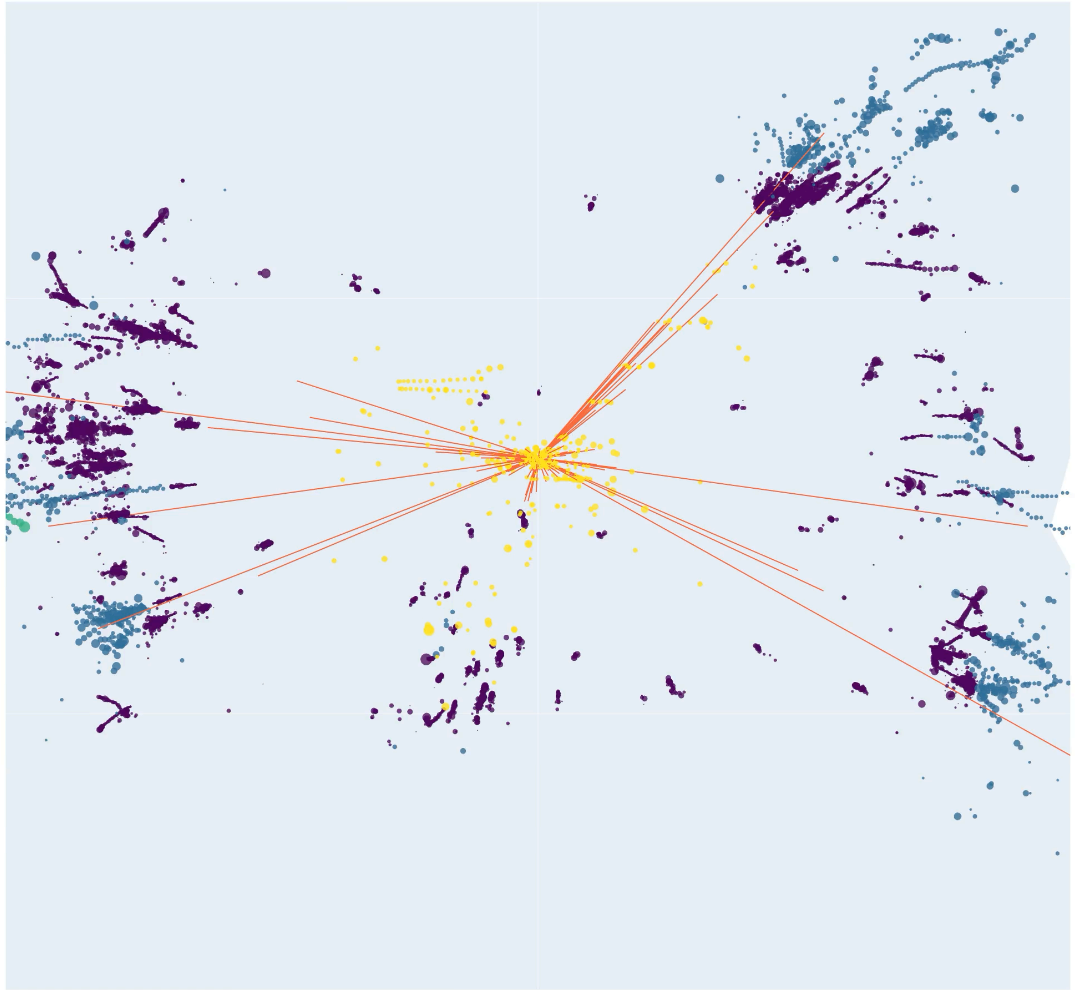
…suitable for tracking, clustering and particle flow

We’ve made the dataset accessible for ML reconstruction studies

Particle Flow Reconstruction

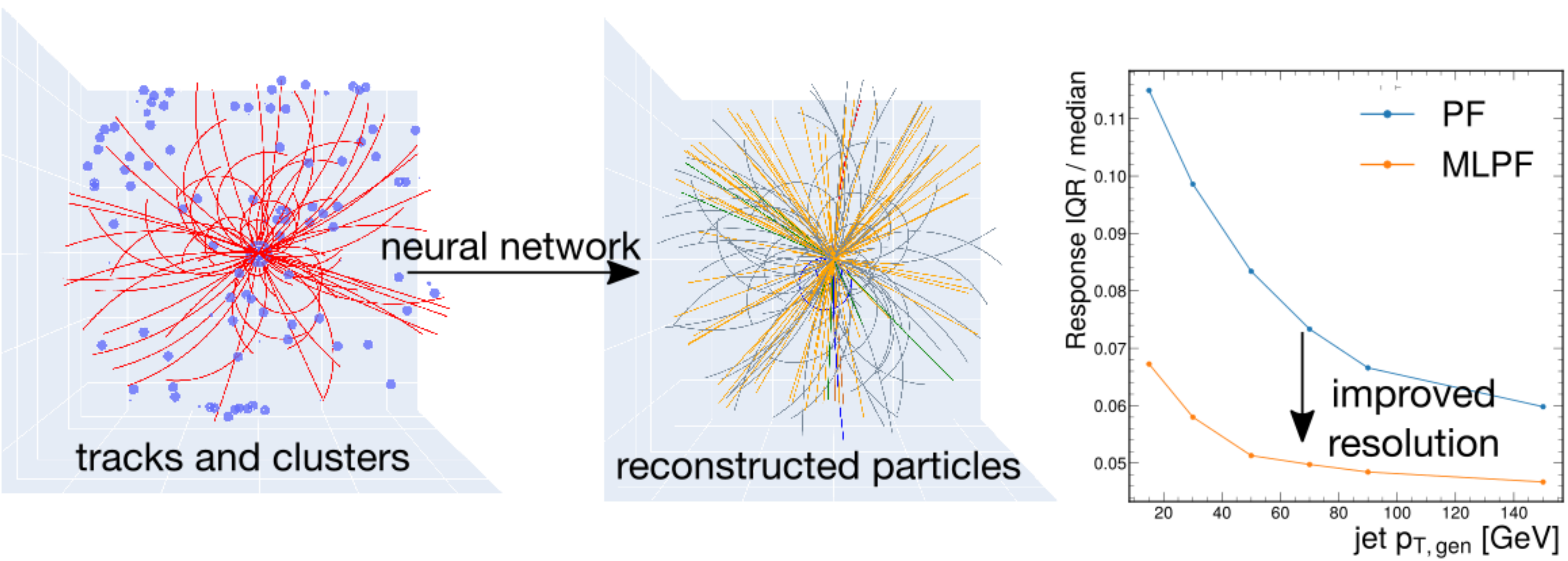
Scalable Neural Network Models and Terascale Datasets

One of the main approaches for event reconstruction at the Large Hadron Collider (LHC) currently relies on particle flow (PF), which combines hits across subdetectors, considering the full event to reconstruct all stable particles in the event. Given the planned High-Luminosity (HL) LHC program, as well as possible future experimental programs of e.g., the Future Circular Collider (FCC), computationally efficient and physically optimal evolutions of the PF-based event reconstruction need to be developed and tested.



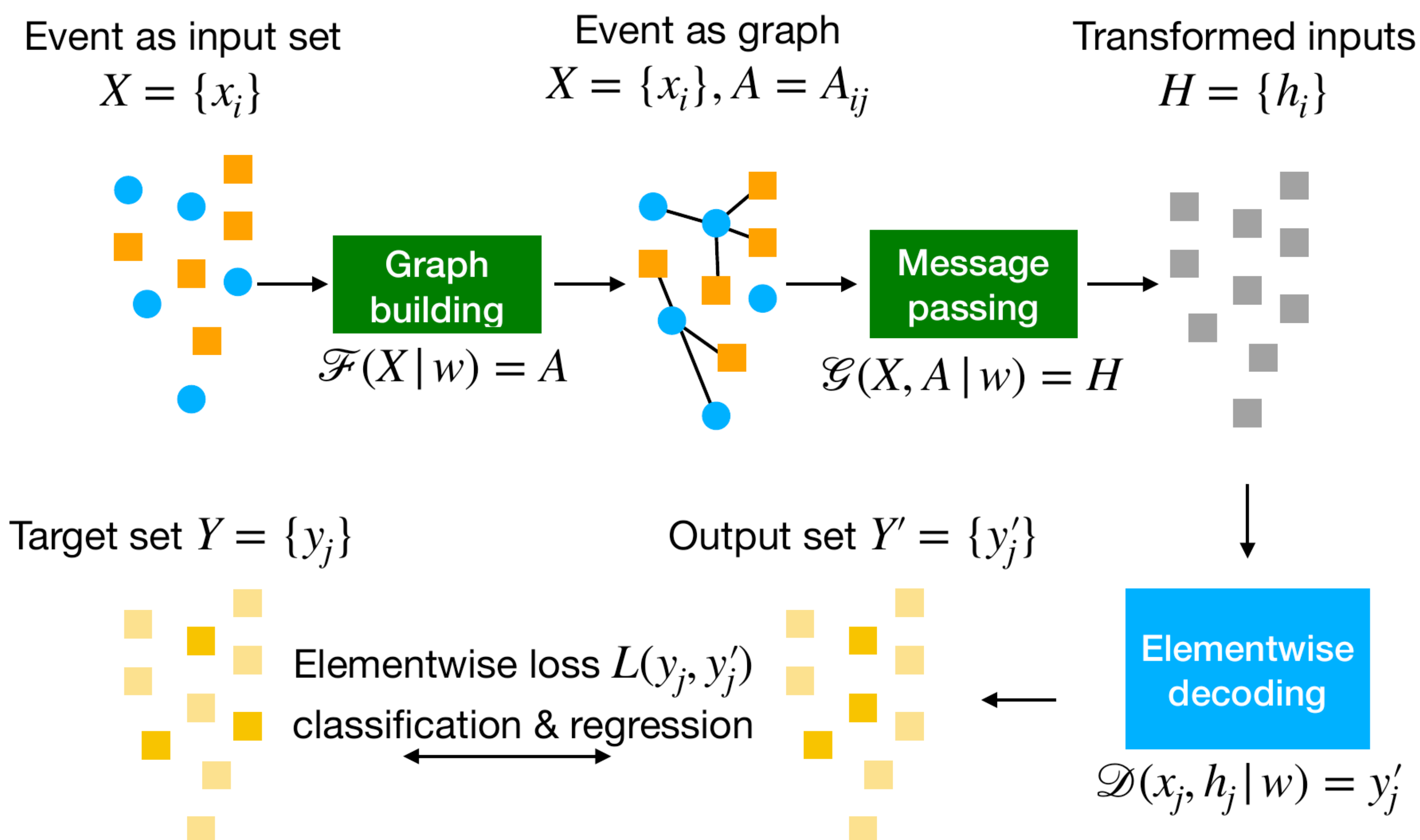
https://www.coe-raise.eu/od-pfr
https://doi.org/10.5281/zenodo.8260741

We can improve event reconstruction fidelity using ML



… while also keeping the models portable, aimed at possible future deployment scenarios.

graph structure learning + message passing
supervised multi-task output



$$x_i = [\mathrm{elem.type}, p_\mathrm{T}, E_\mathrm{ECAL}, E_\mathrm{HCAL}, \eta, \phi, \eta_\mathrm{outer}, \phi_\mathrm{outer}, q, \ldots]$$

$$y_j = [\mathrm{PID}, p_\mathrm{T}, E, \eta, \phi, q], \quad \mathrm{PID} \in \{\text{none}, \text{charged hadron}, \text{neutral hadron}, \gamma, e^\pm, \mu^\pm, \ldots\}$$

$$h_i \in \mathbb{R}^{N_\mathrm{hidden}}$$

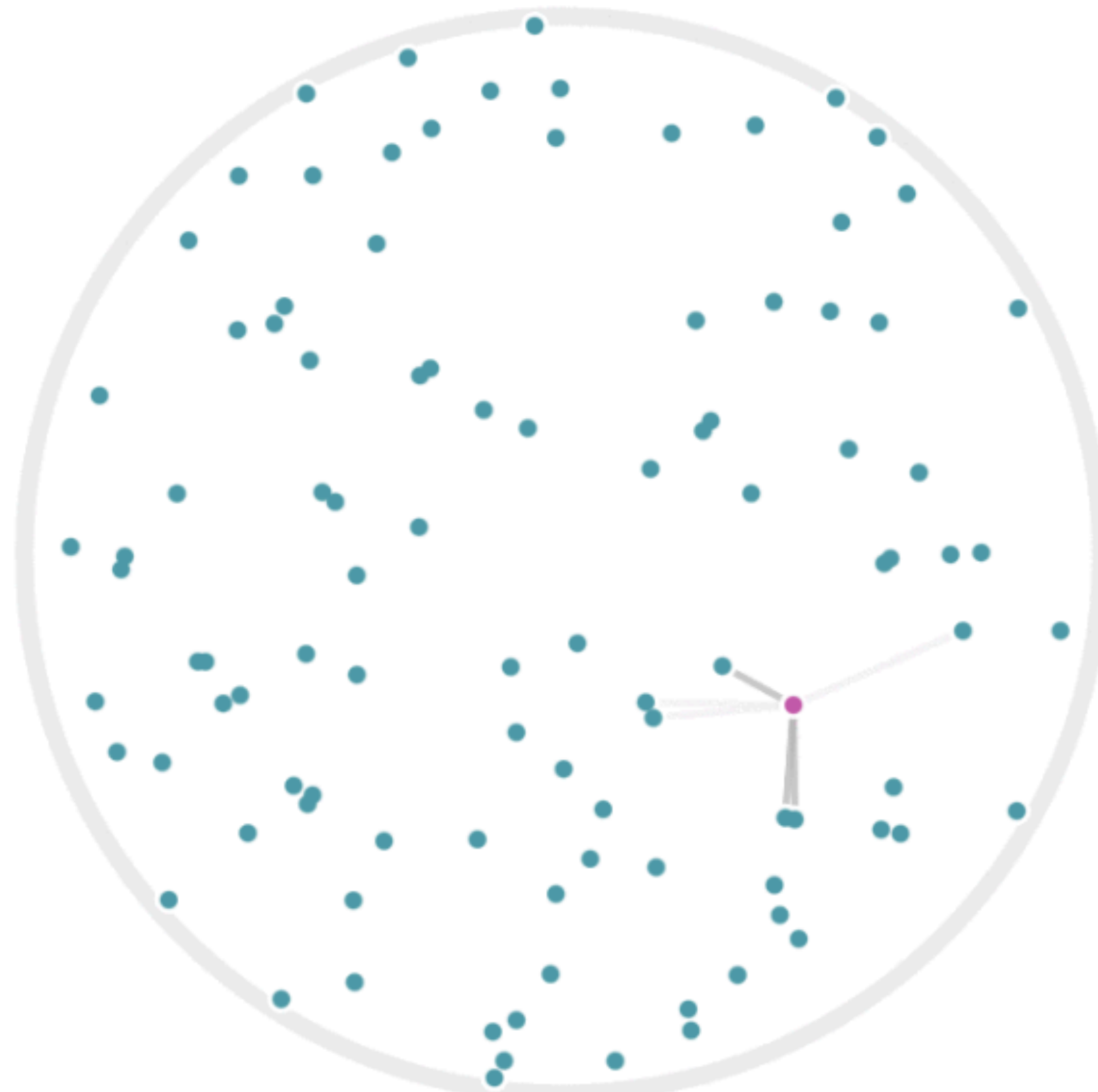
Trainable neural networks: $\mathcal{F}, \mathcal{G}, \mathcal{D}$

● - track, ■ - calorimeter cluster, ■ - encoded element

■ - target (predicted) particle, ■ - no target (predicted) particle

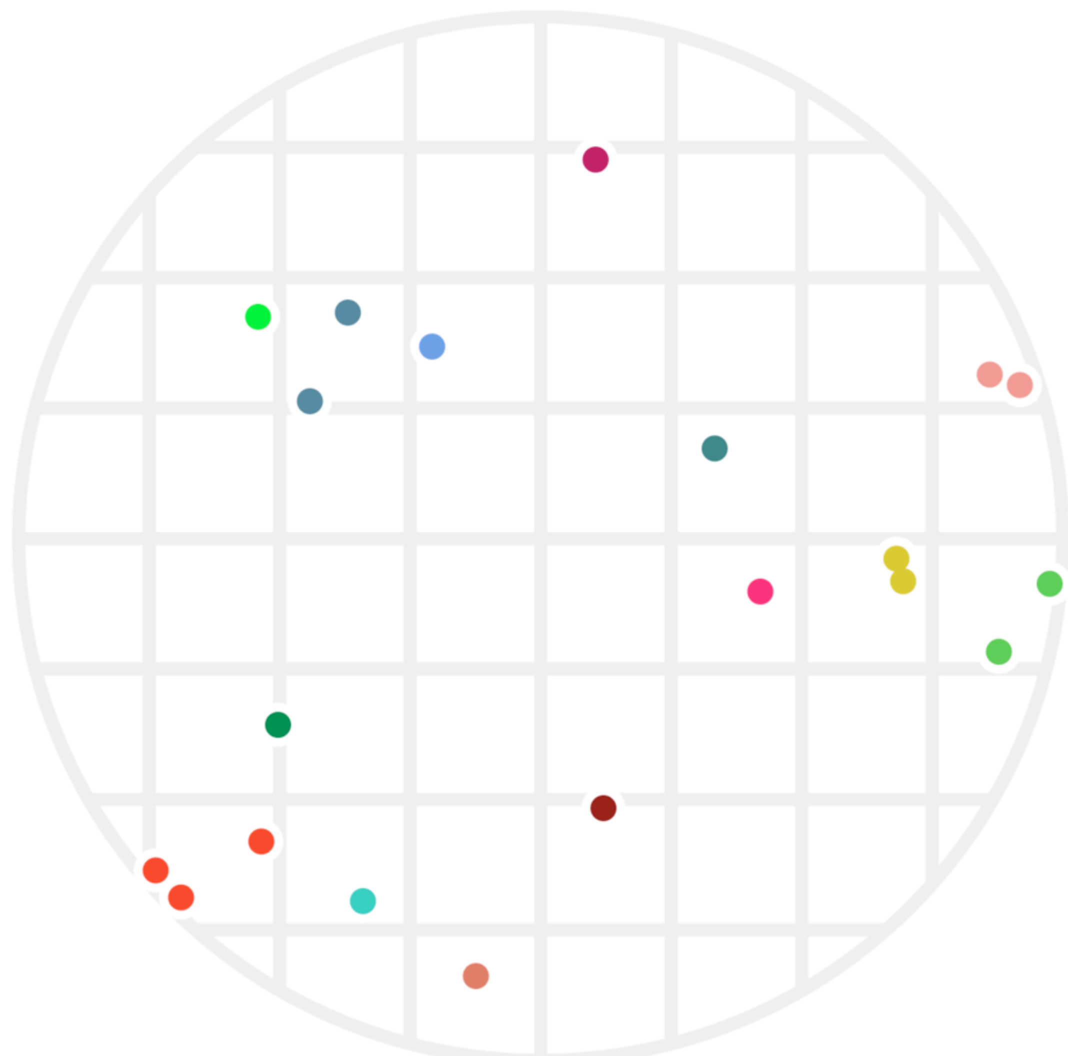
Pata, J., Duarte, J., Vlimant, JR. *et al.* MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks. *Eur. Phys. J. C* **81**, 381 (2021). https://doi.org/10.1140/epjc/s10052-021-09158-w

Naive nearest neighbours graph building: need to compare each pair of particles, $\mathcal{O}(N^2)$ complexity



Credit: https://unboxresearch.com/articles/lsh_post1.html

Divide space to bins, particles are nearby if they are in the same bin.

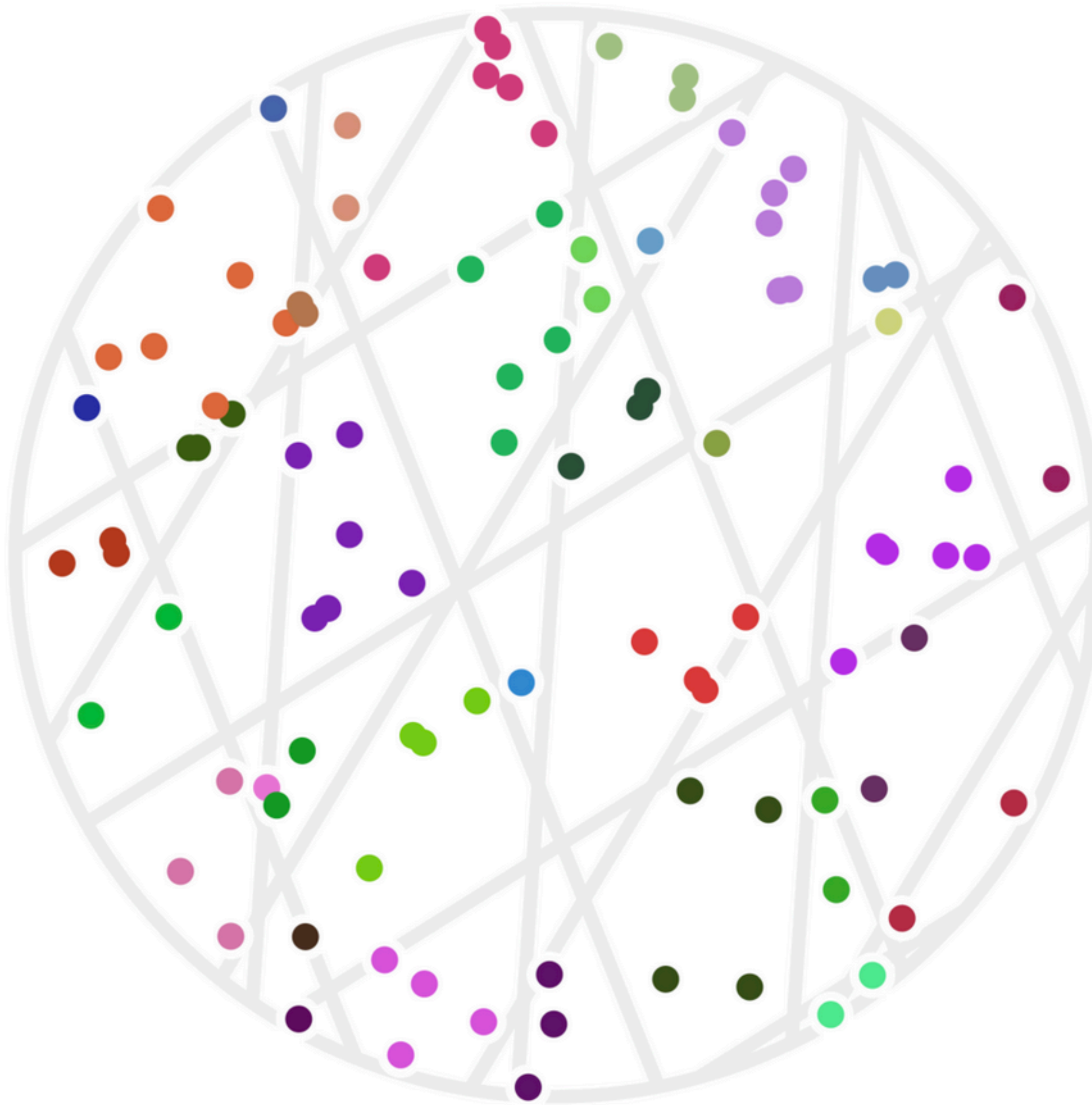


particle features to bin index: hash function

Locality sensitive hashing

Credit: https://unboxresearch.com/articles/lsh_post1.html

Randomized bins (hash functions) work even better!

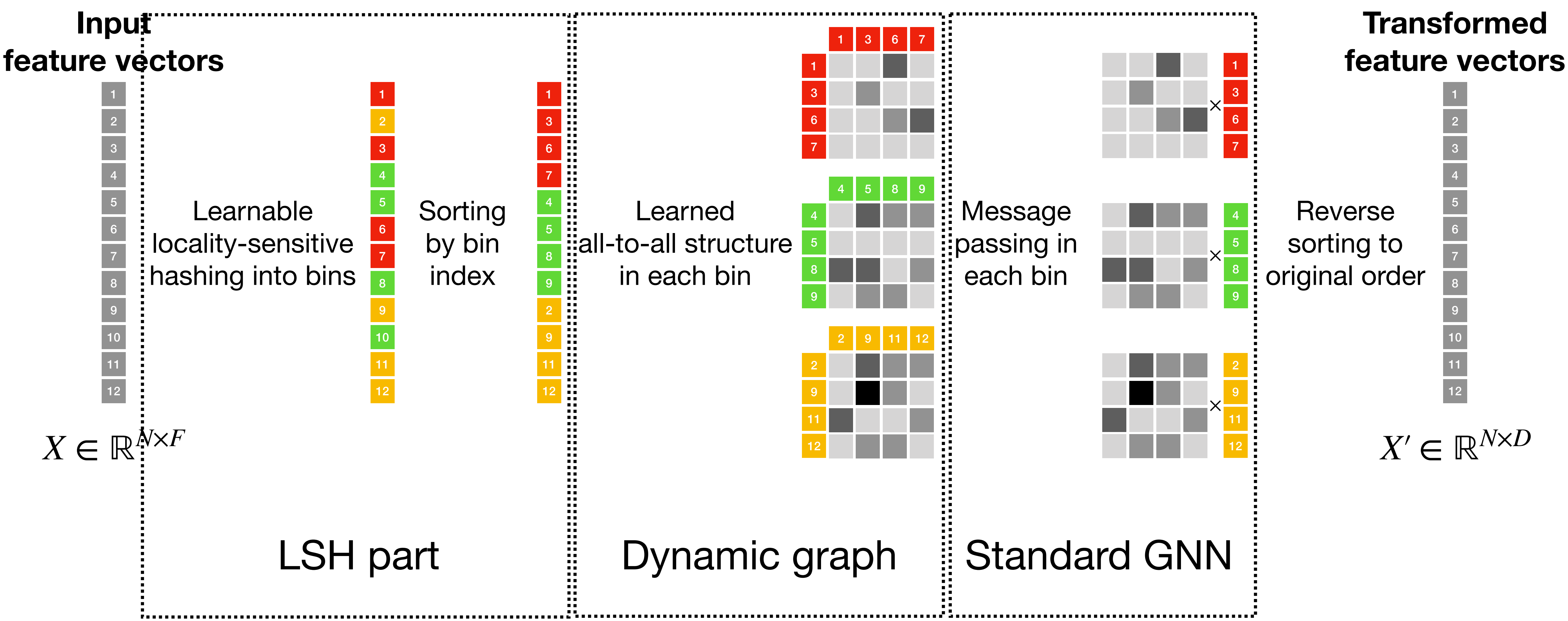


Simple to implement in TensorFlow, PyTorch, JAX using native operations: high portability to Nvidia, AMD, Intel Gaudi etc. **today**

Credit: https://unboxresearch.com/articles/lsh_post1.html

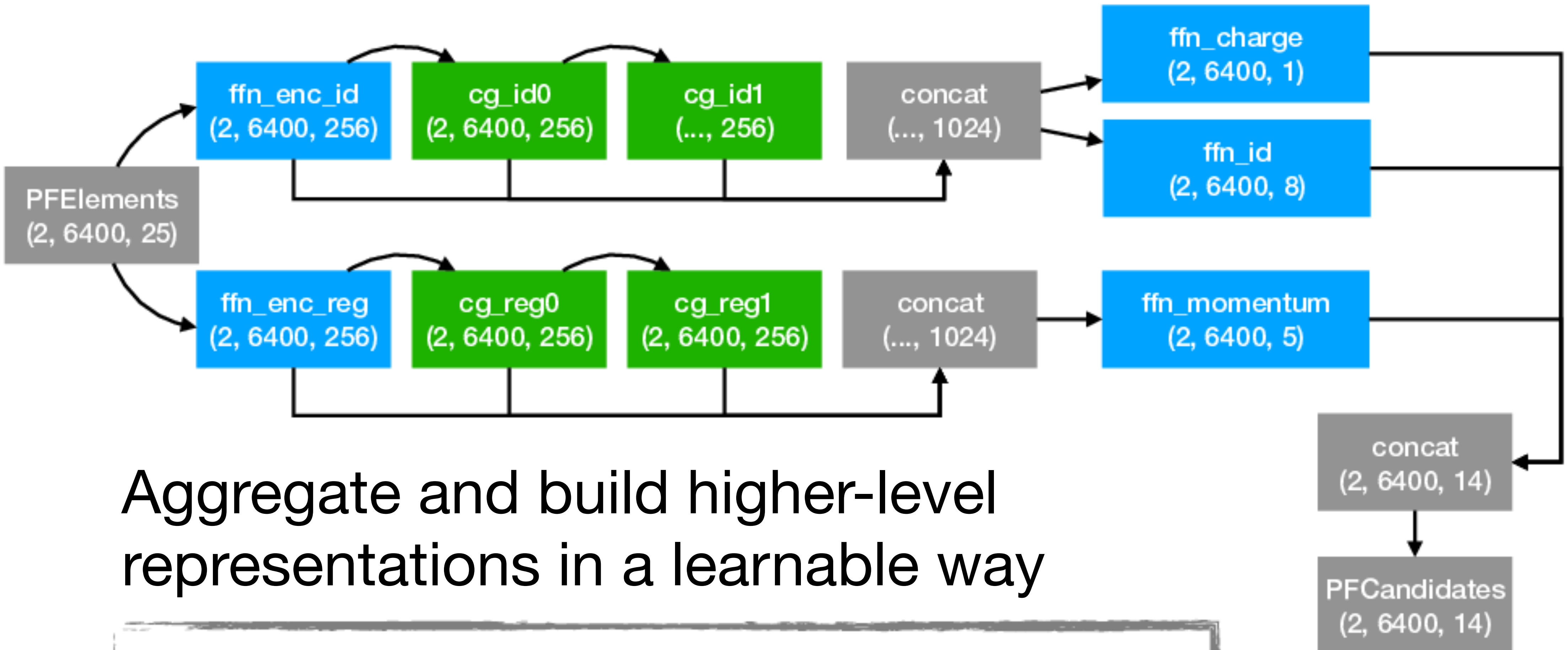
Scalable GNN based on the Reformer architecture

One layer of learnable graph building with locality sensitive hashing and message passing



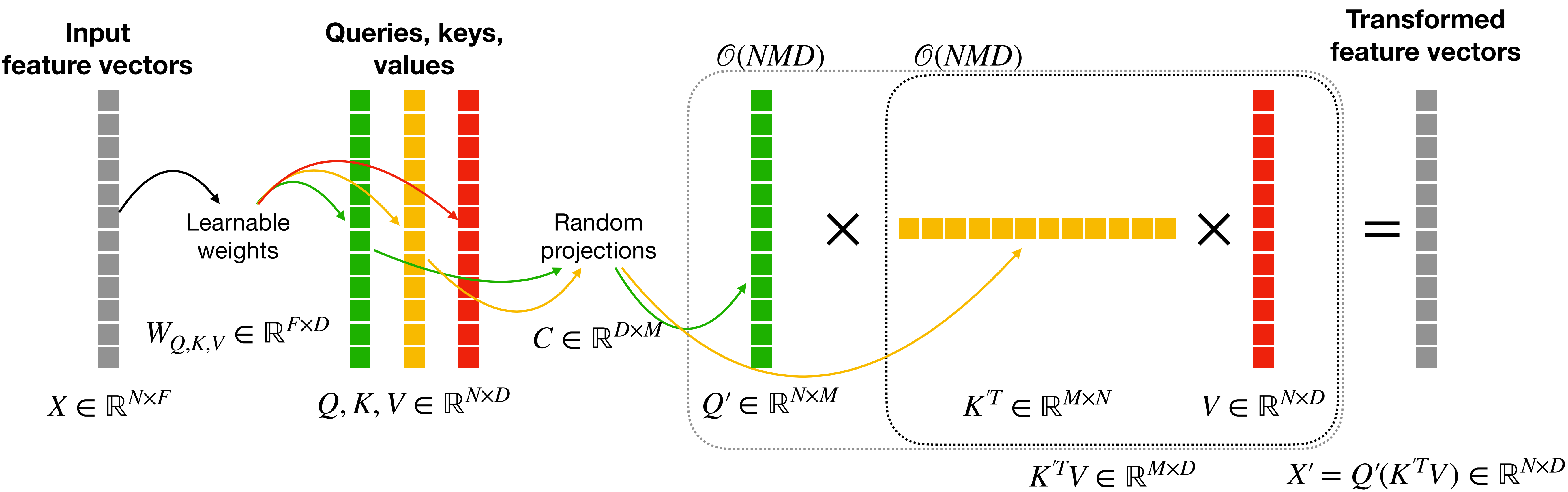
https://blog.research.google/2020/01/reformer-efficient-transformer.html

Can construct multilayered networks from the scalable GNN-LSH building block



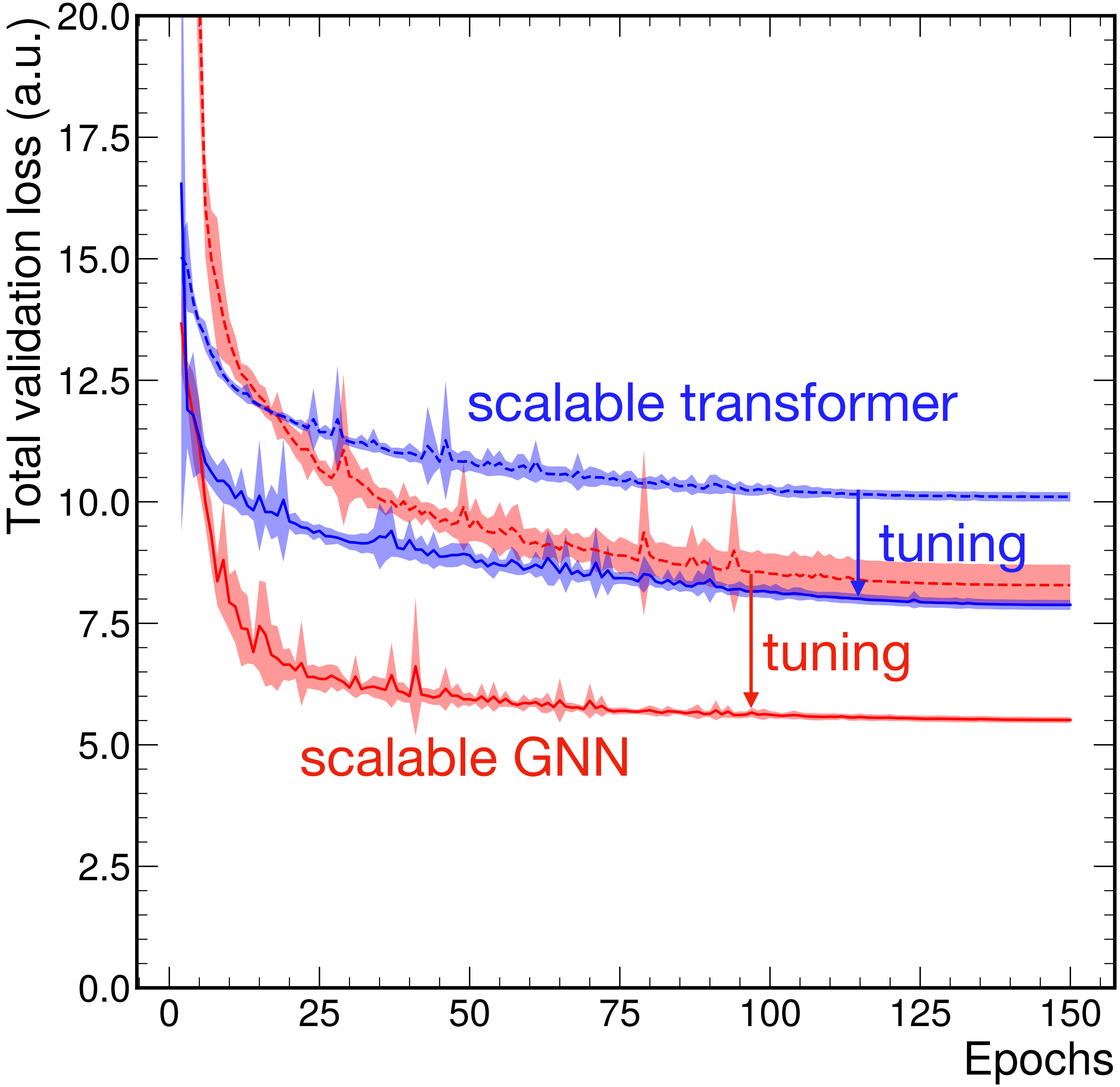
Aggregate and build higher-level representations in a learnable way

Alternative: scalable transformer based on the Performer architecture



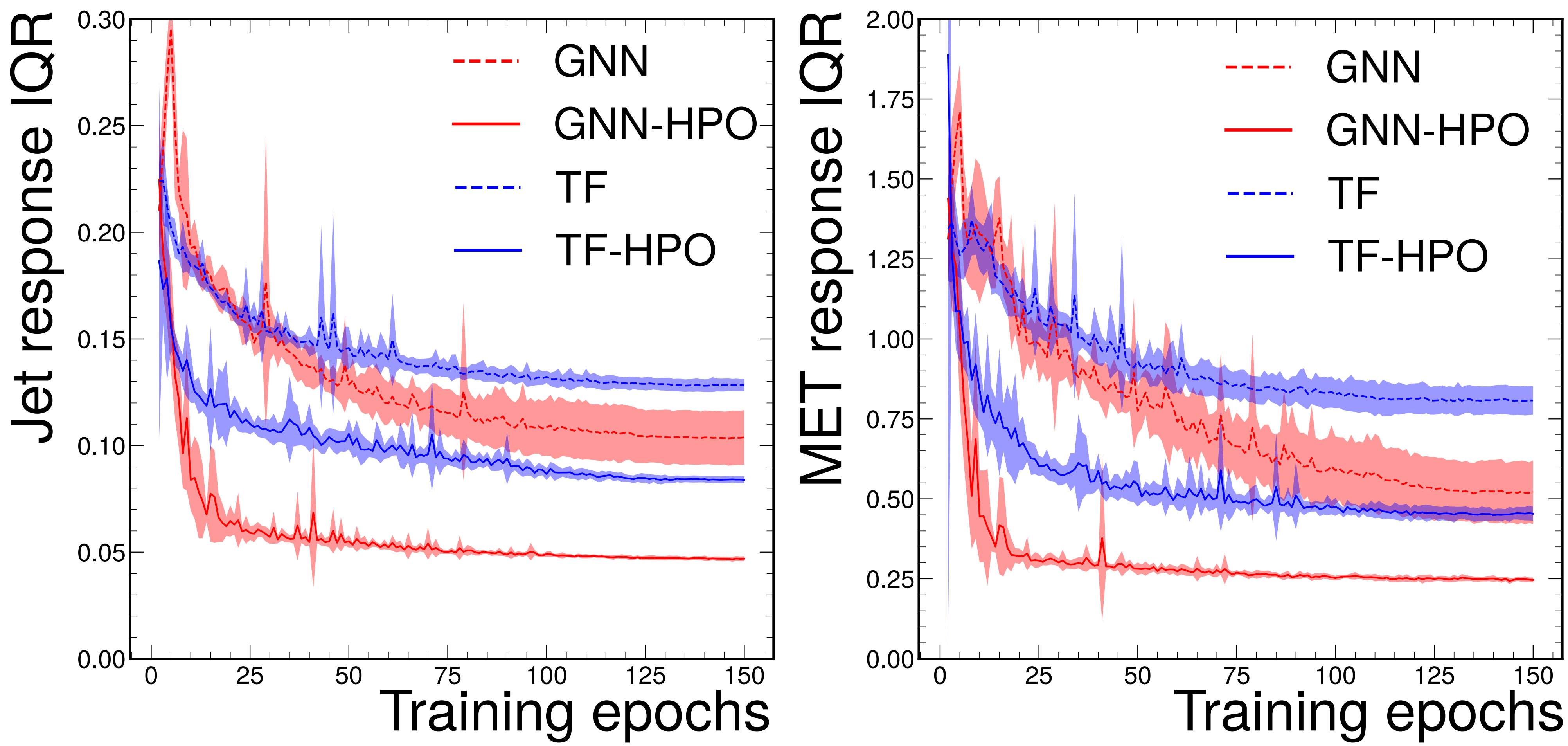
https://blog.research.google/2020/10/rethinking-attention-with-performers.html

Large-scale model comparison and hyperparameter tuning on Jülich HPC



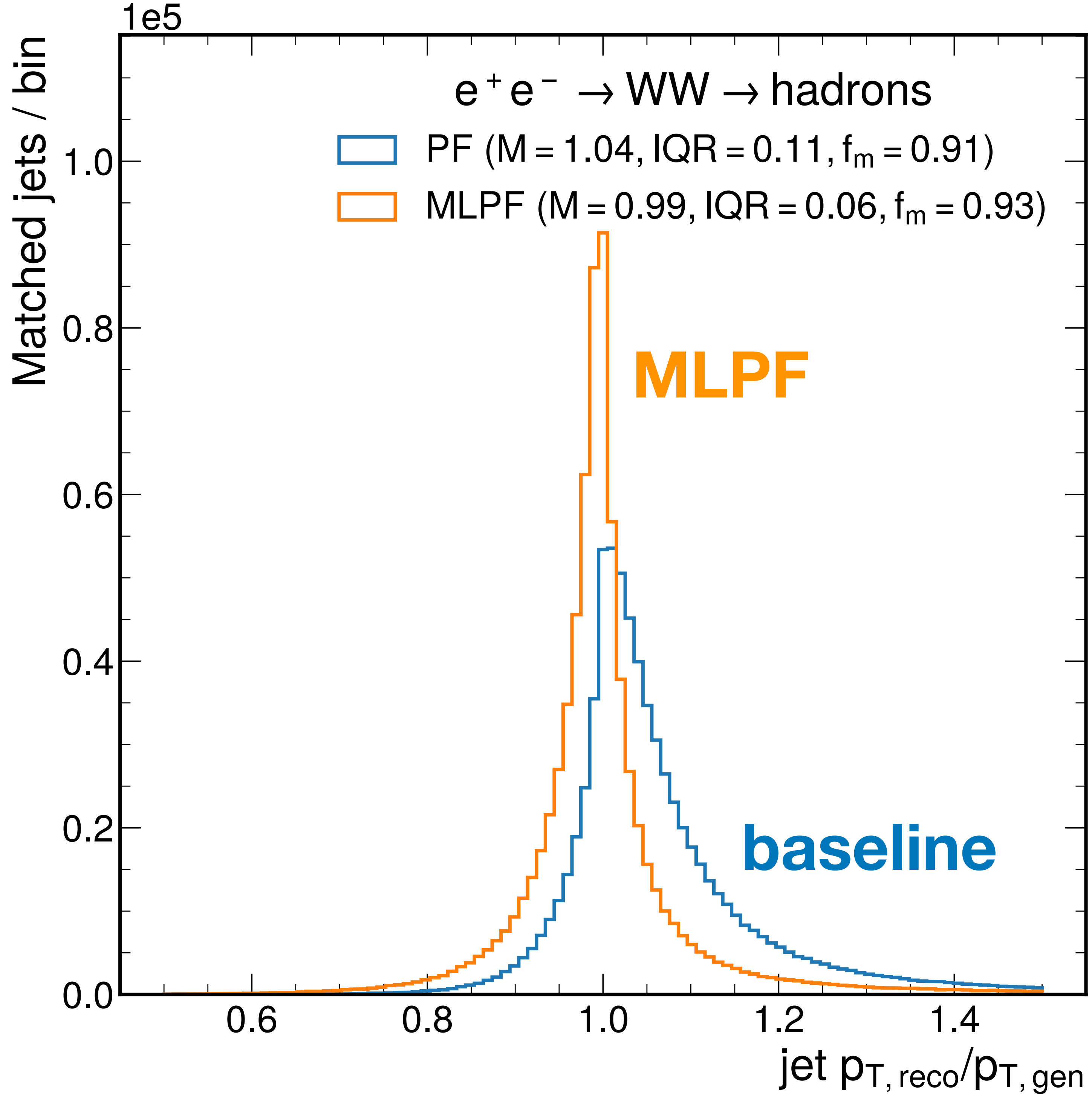
Better-than-SOTA result!

Even though we optimise a particle-based loss…



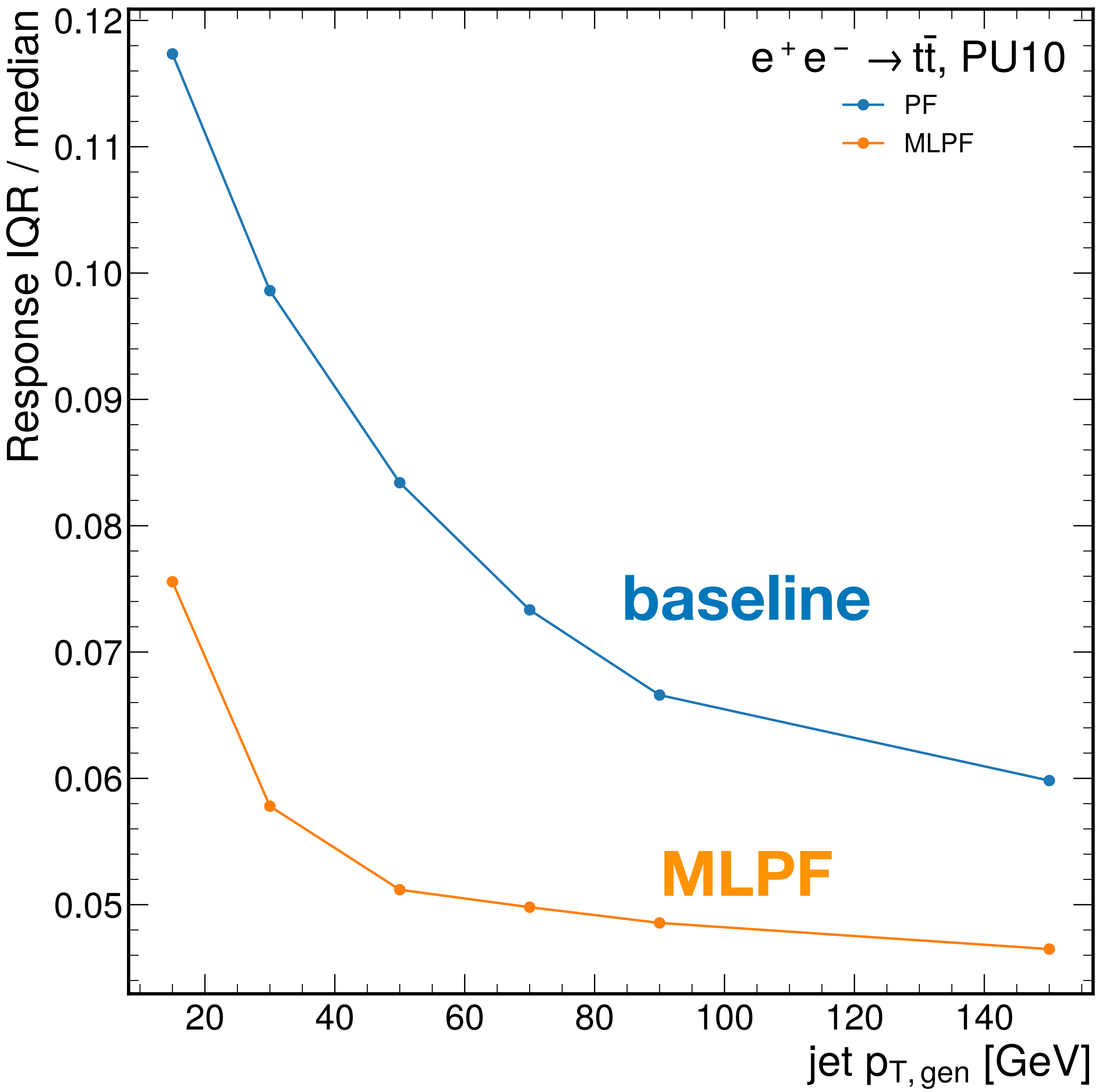
Better-than-SOTA **event reconstruction emerges naturally**

In samples never used in training…



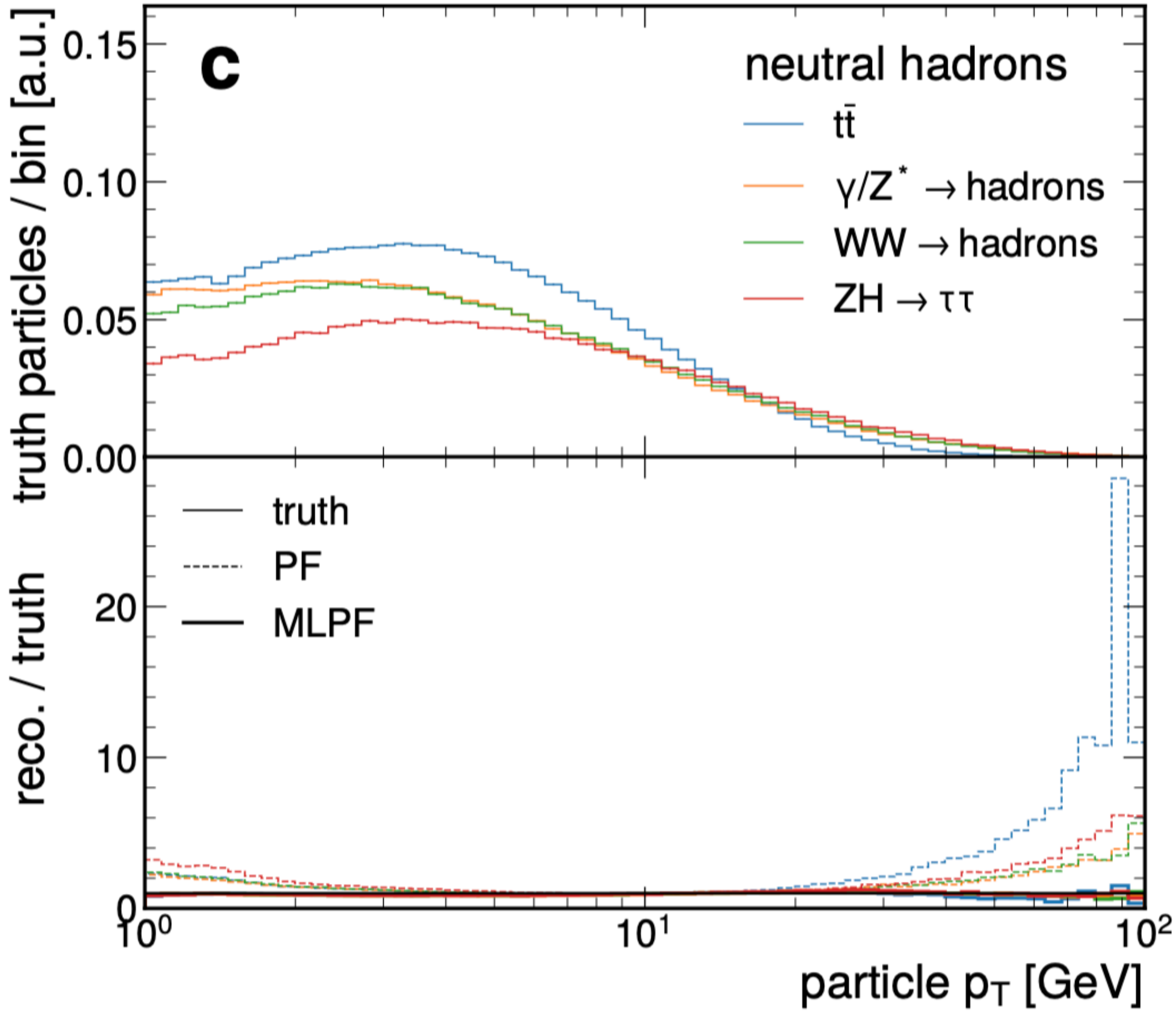
about 50% improvement in jet response width over the baseline

In samples never used in training…



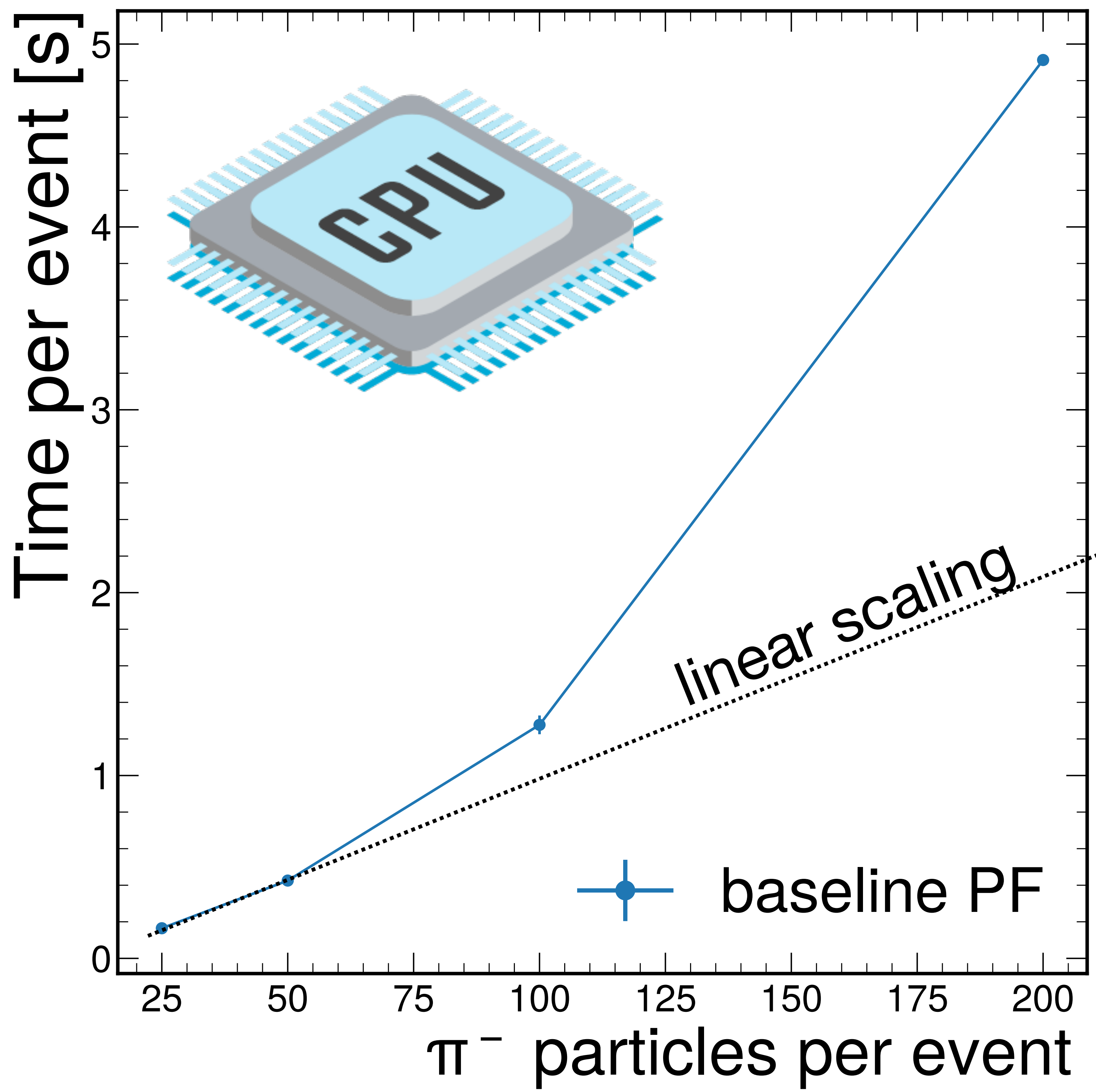
a consistent improvement over the full generated pT spectrum

Datasets are diverse…

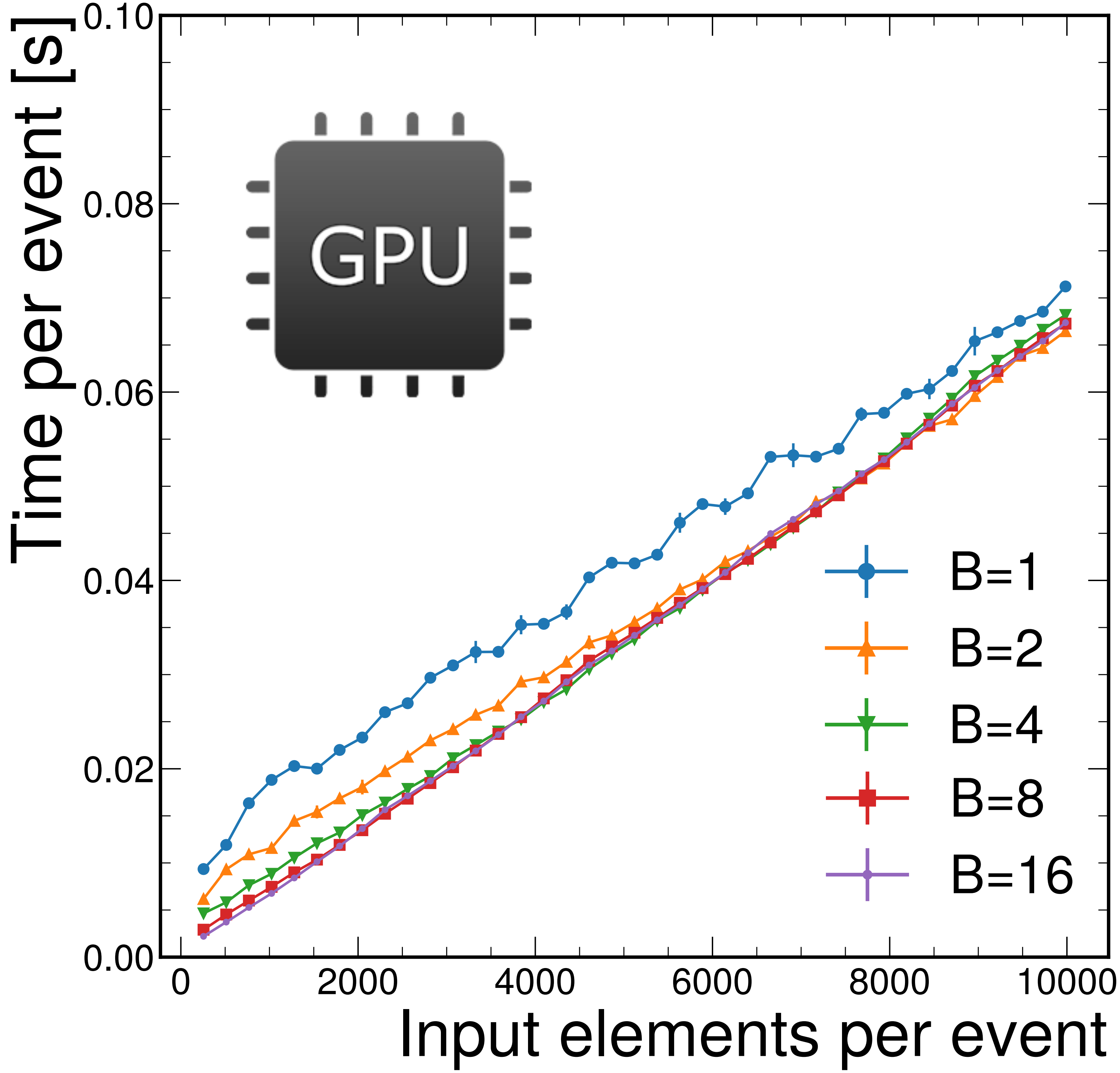


so we have to predict the bulk and the tails well for all particle types

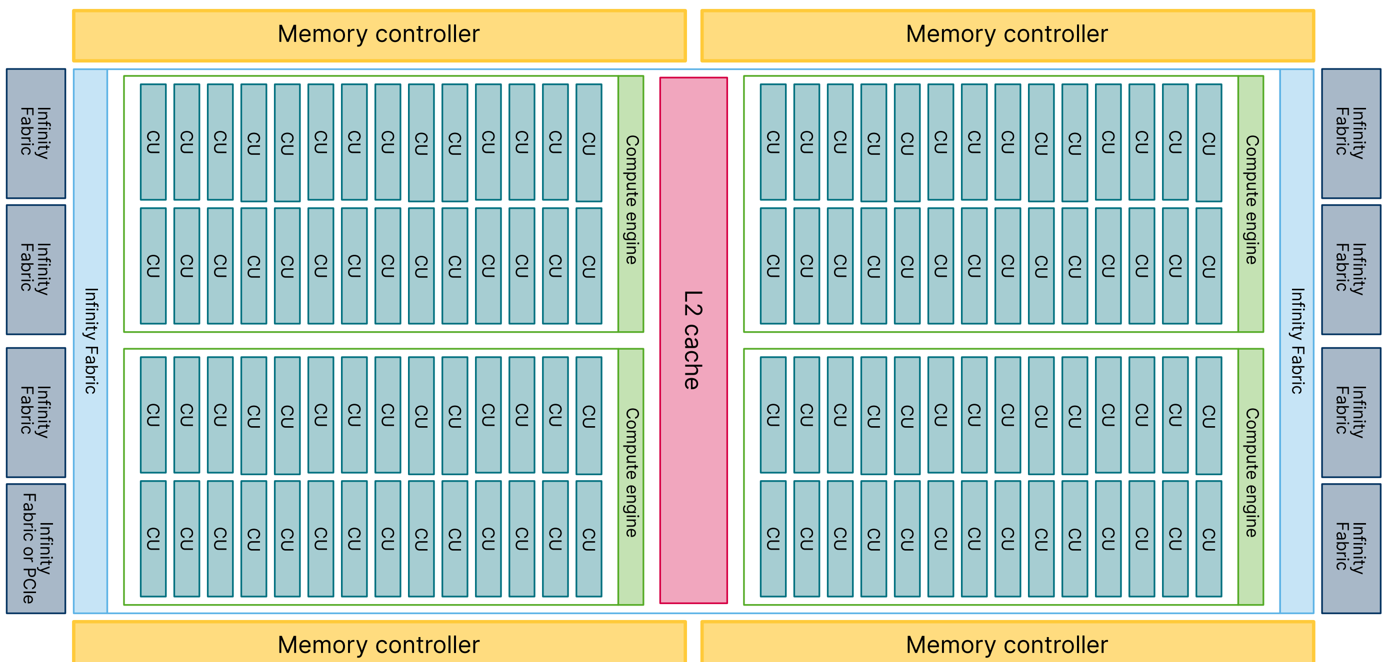
Baseline (untuned) algo runs **only on CPU**, scales ~quadratically, **runtime per event is in seconds**.



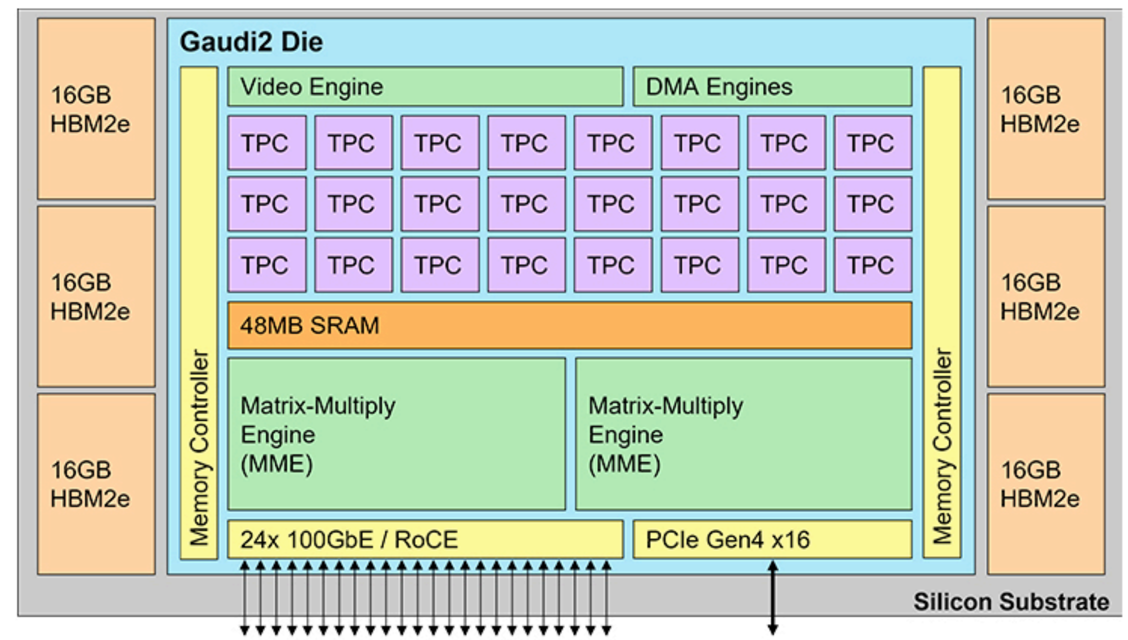
ML model **scales linearly**, runs in **milliseconds per event** on a consumer 8GB GPU



The HPC AI chip landscape is diversifying



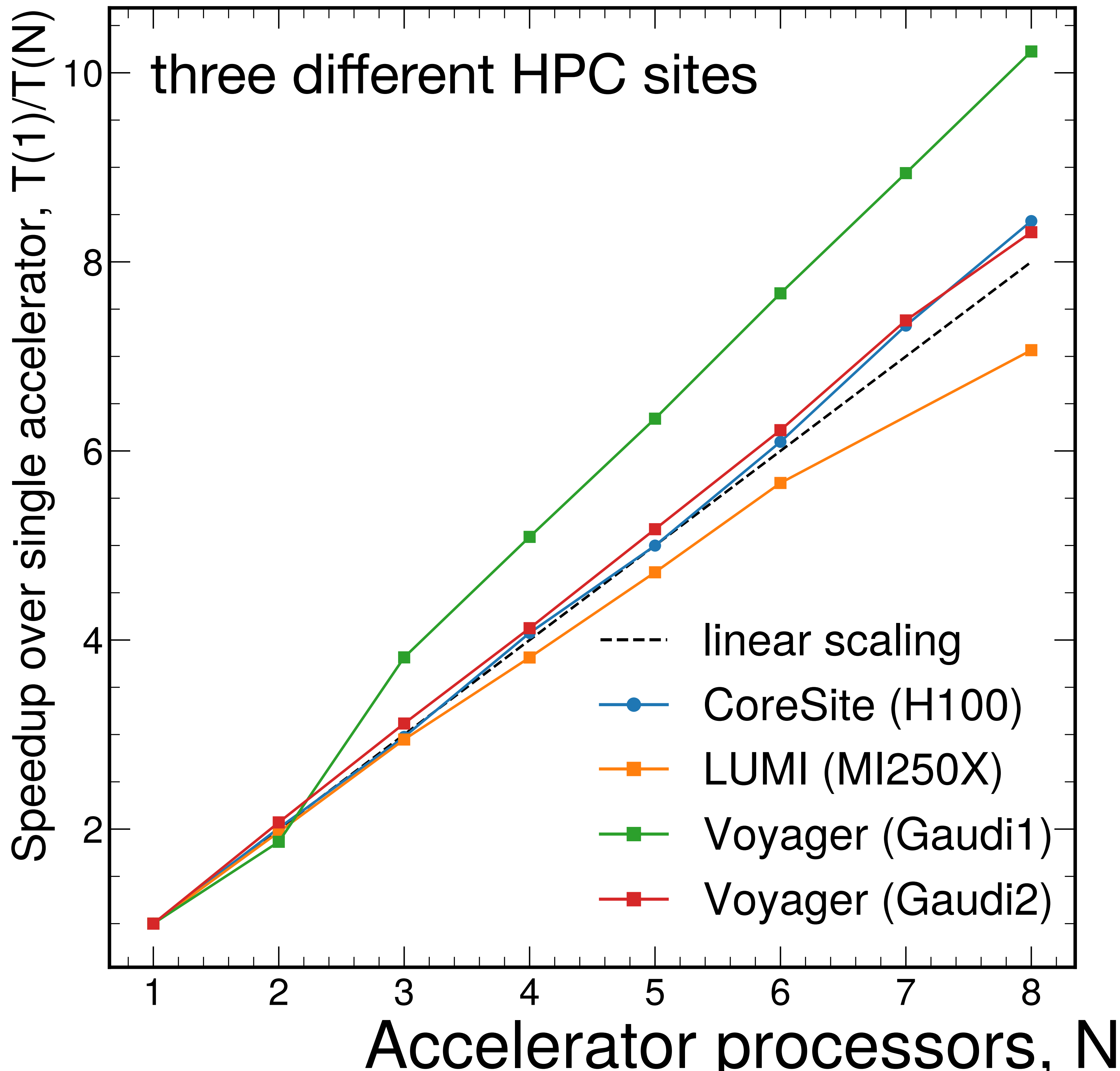
AMD MI250X GPU



Intel Gaudi2 deep learning processor

… we need flexible and portable codes to make use of these resources in the near future!

Portable on CPU, nVidia & AMD GPU, Intel Habana Gaudi chips



Outlook for ML-based reconstruction

pretraining and fine-tuning	SOTA supervised training
reconstruction as a differentiable module	deployment in experimental frameworks
foundation-type models for reconstruction	validation on real data
interpretability	quantisation and pruning
robustness & uncertainty quantification	deployment on real-time inference hardware

32. Reconstructing Particle Tracks in One Go with a Recursive Graph Attention Network

Jay Chan (Lawrence Berkeley ...

11/03/2024, 17:30

Track 2: Data Analysis - ... | Oral | Track 2: Data Analysis - ...

Track reconstruction is a crucial task in particle experiments and is traditionally very computationally expensive due to its combinatorial nature. Many recent developments have explored new tracking algorithms in order to improve scalability in preparation of the HL-LHC. In particular, Graph neural networks (GNNs) have emerged as a promising approach due to

176. New developments and applications of a Deep-learning-based Full Event Interpretation (DFEI) in proton-proton collisions

Felipe Luan Souza De Almeida (Syracuse University ...

12/03/2024, 12:30

Track 1: Computing Tec... | Oral | Track 1: Computing Tec...

The LHCb experiment at the Large Hadron Collider (LHC) is designed to perform high-precision measurements of heavy-hadron decays, which requires the collection of large data samples and a good understanding and suppression of multiple background sources. Both factors are challenged by a five-fold increase in the average number of proton-proton

117. Denoising Graph Super-Resolution with Diffusion Models and Transformers for Improved Particle Reconstruction

Nilotpal Kakati (Weizmann Institute ...

12/03/2024, 12:50

Track 2: Data Analysis - ... | Oral | Track 2: Data Analysis - ...

Accurately reconstructing particles from detector data is a critical challenge in experimental particle physics. The detector's spatial resolution, specifically the calorimeter's granularity, plays a crucial role in determining the quality of the particle reconstruction. It also sets the upper limit for the algorithm's theoretical capabilities. Super-resolution techniques

94. Beyond Language: Foundation Models for Collider Physics Data

Anna Hallin (University of Hamb...

12/03/2024, 12:30

Track 2: Data Analysis - ... | Oral | Track 2: Data Analysis - ...

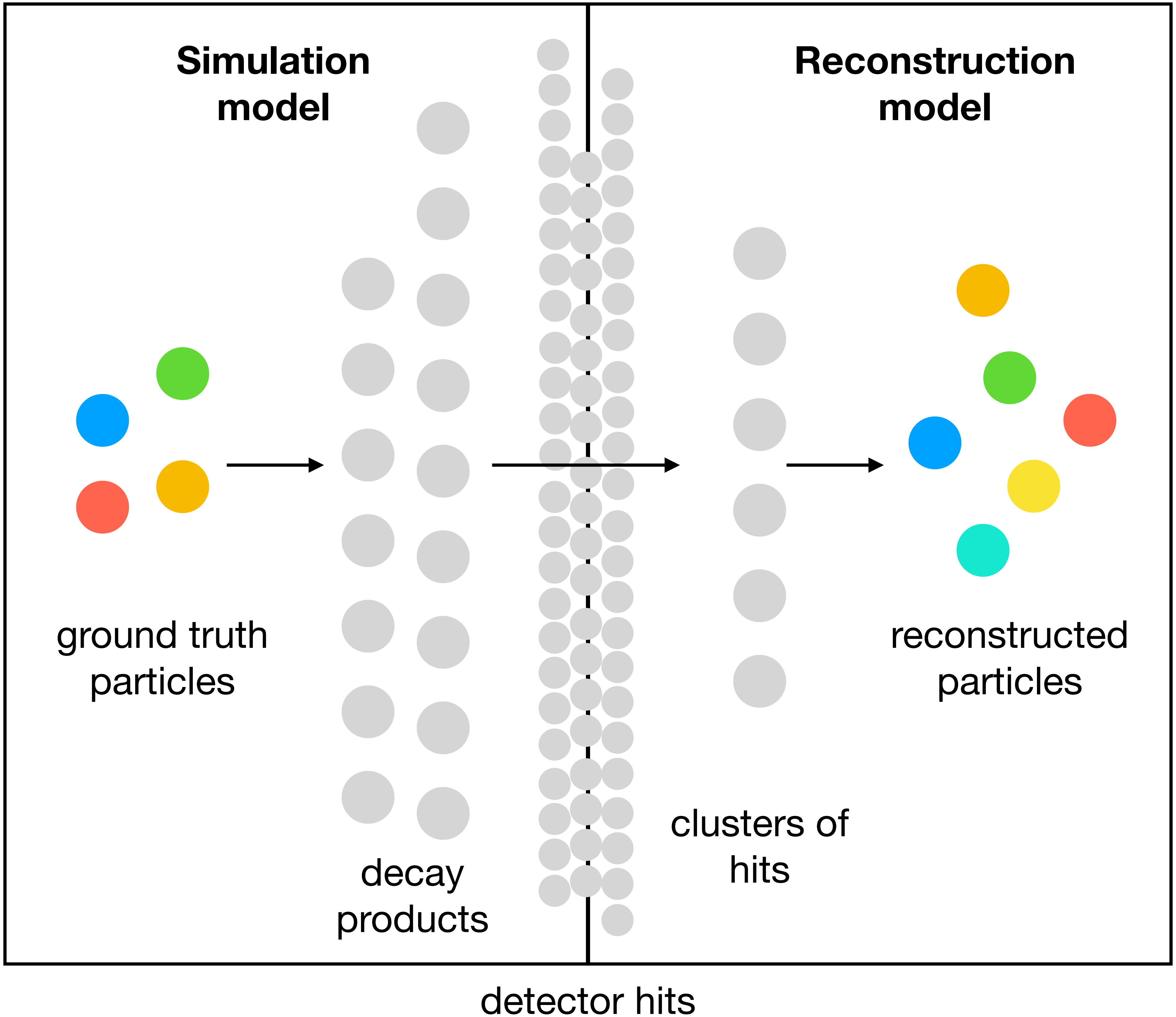
Foundation models have revolutionized natural language processing, demonstrating exceptional capabilities in handling sequential data. Their ability to generalize across tasks and datasets offers promising applications in high energy physics (HEP). However, collider physics data, unlike language, involves both continuous and discrete data types, including four-

ML event reconstruction improves physics

Scalable ML models improve throughput

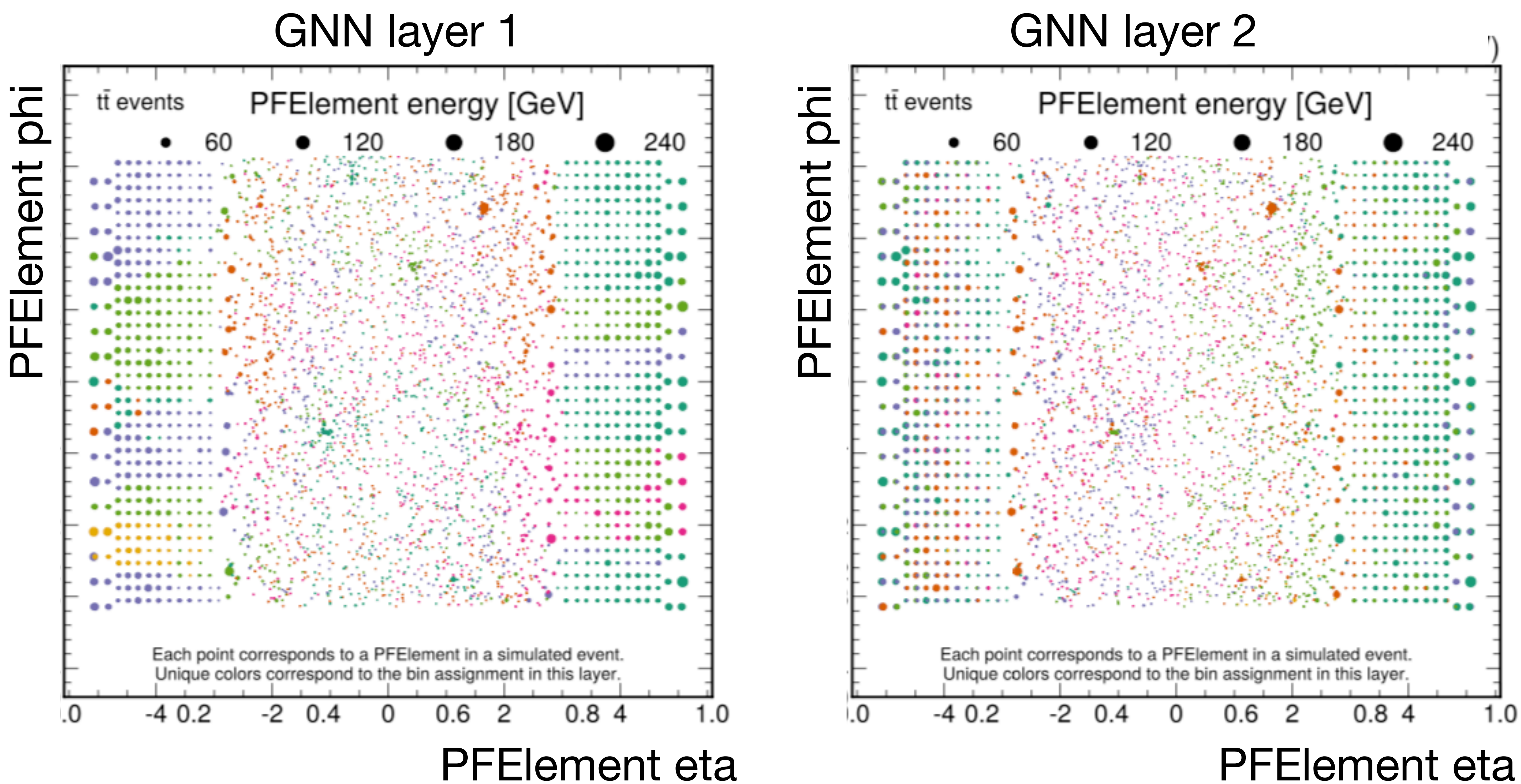
Open datasets & code accelerate research

Backup



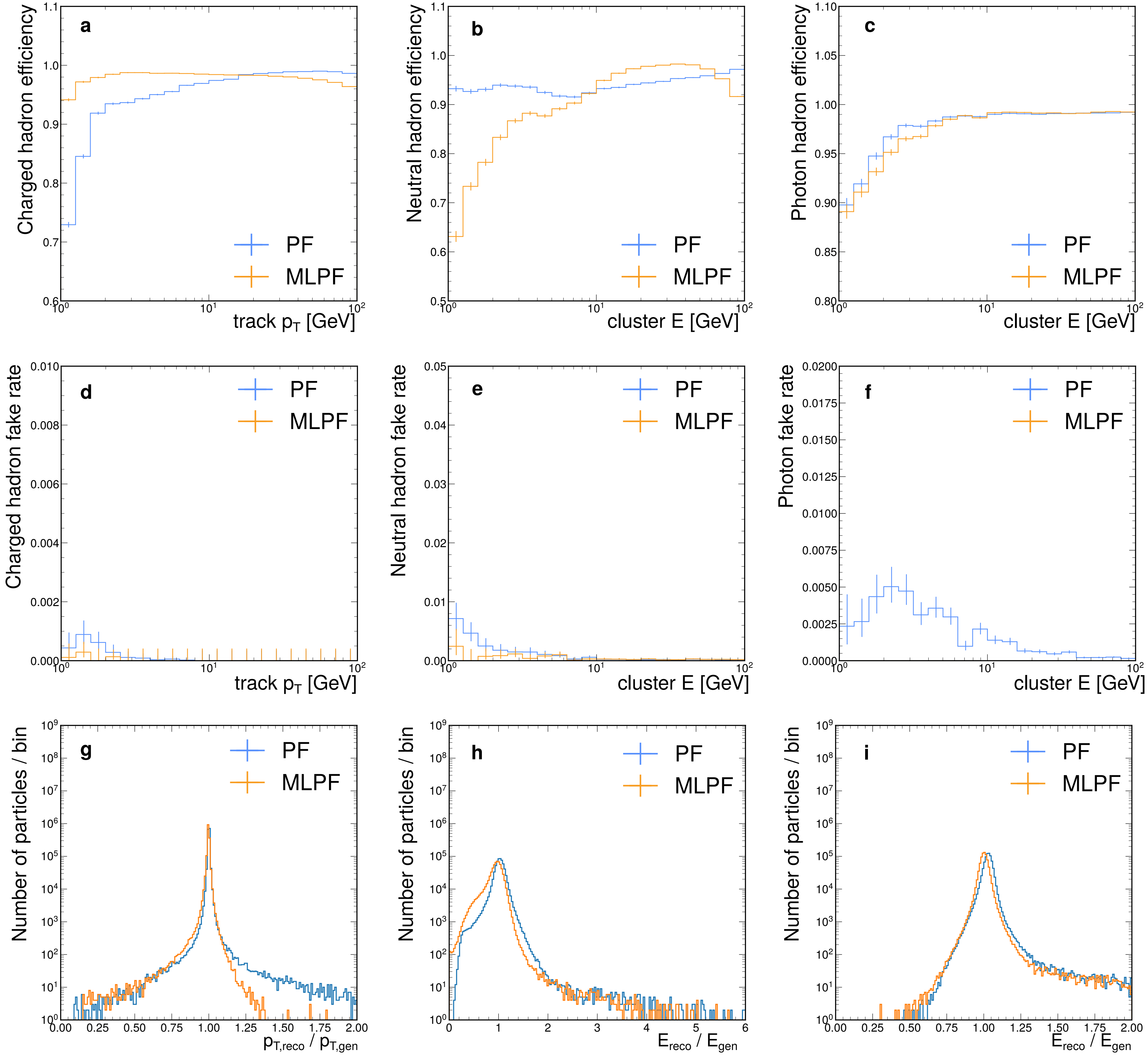
Simulation model
Reconstruction model
ground truth particles
decay products
clusters of hits
reconstructed particles
detector hits

Clustering (graph building) is an internal detail, not a model target.

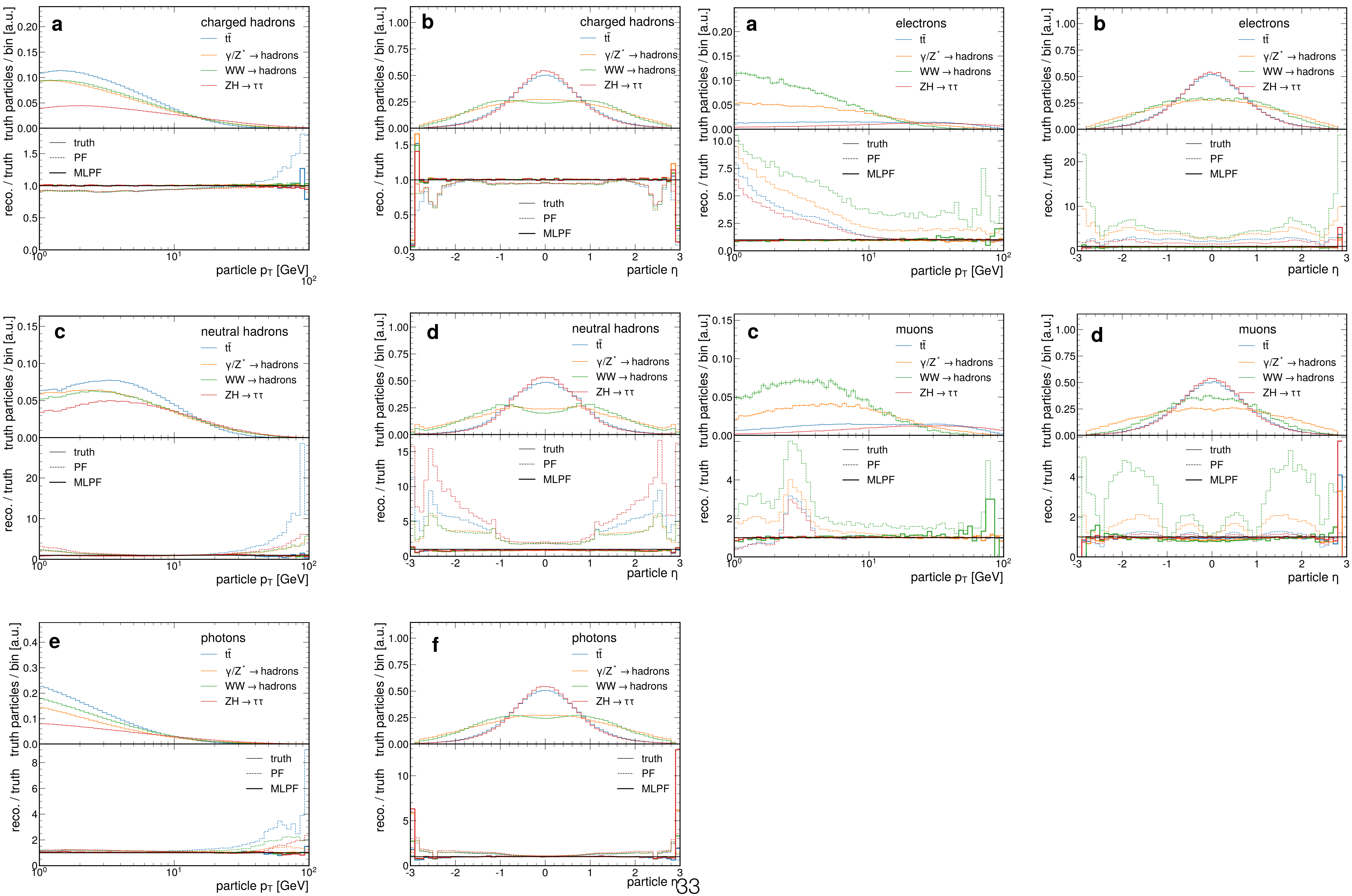


Optimize model for particle reconstruction, not clustering!

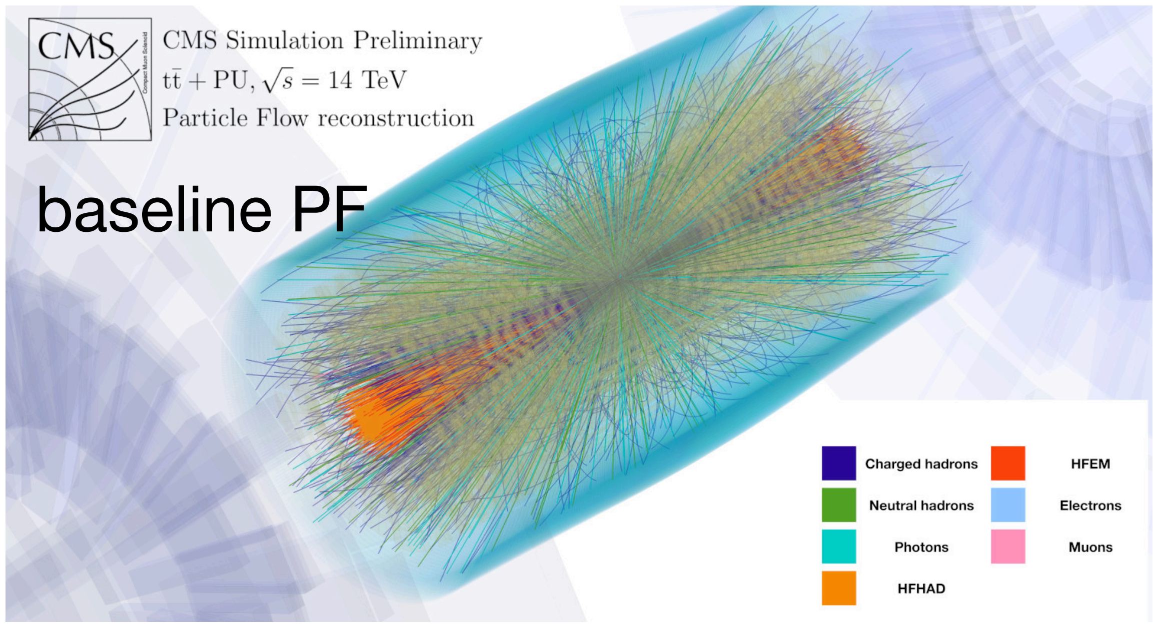
Performance in single particle gun samples is nearly equivalent, as expected

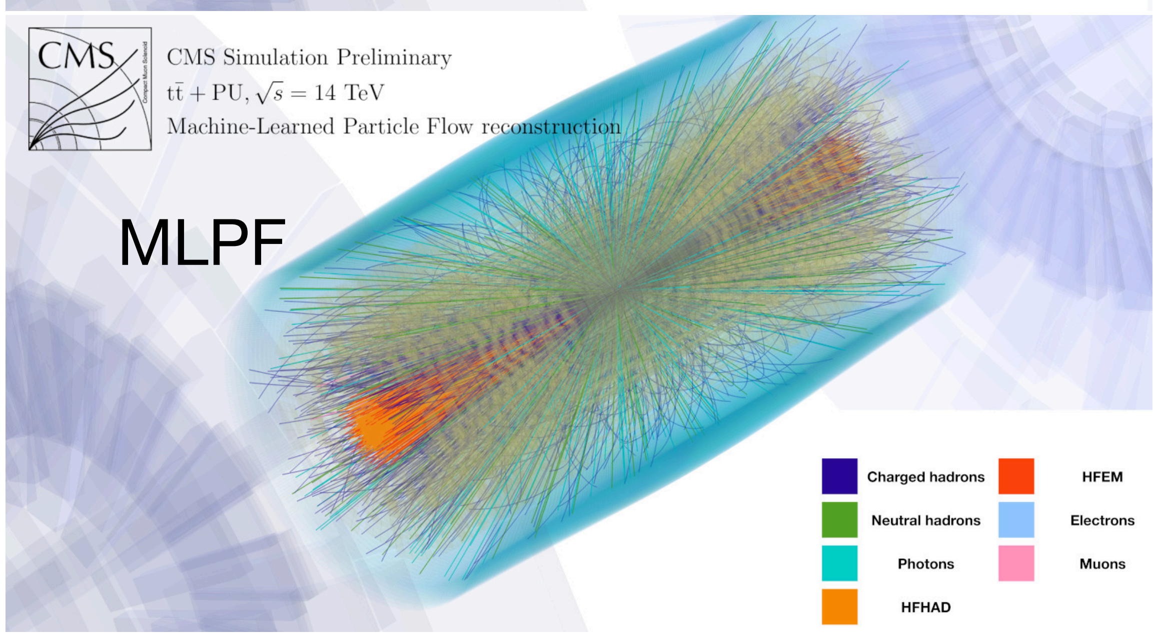


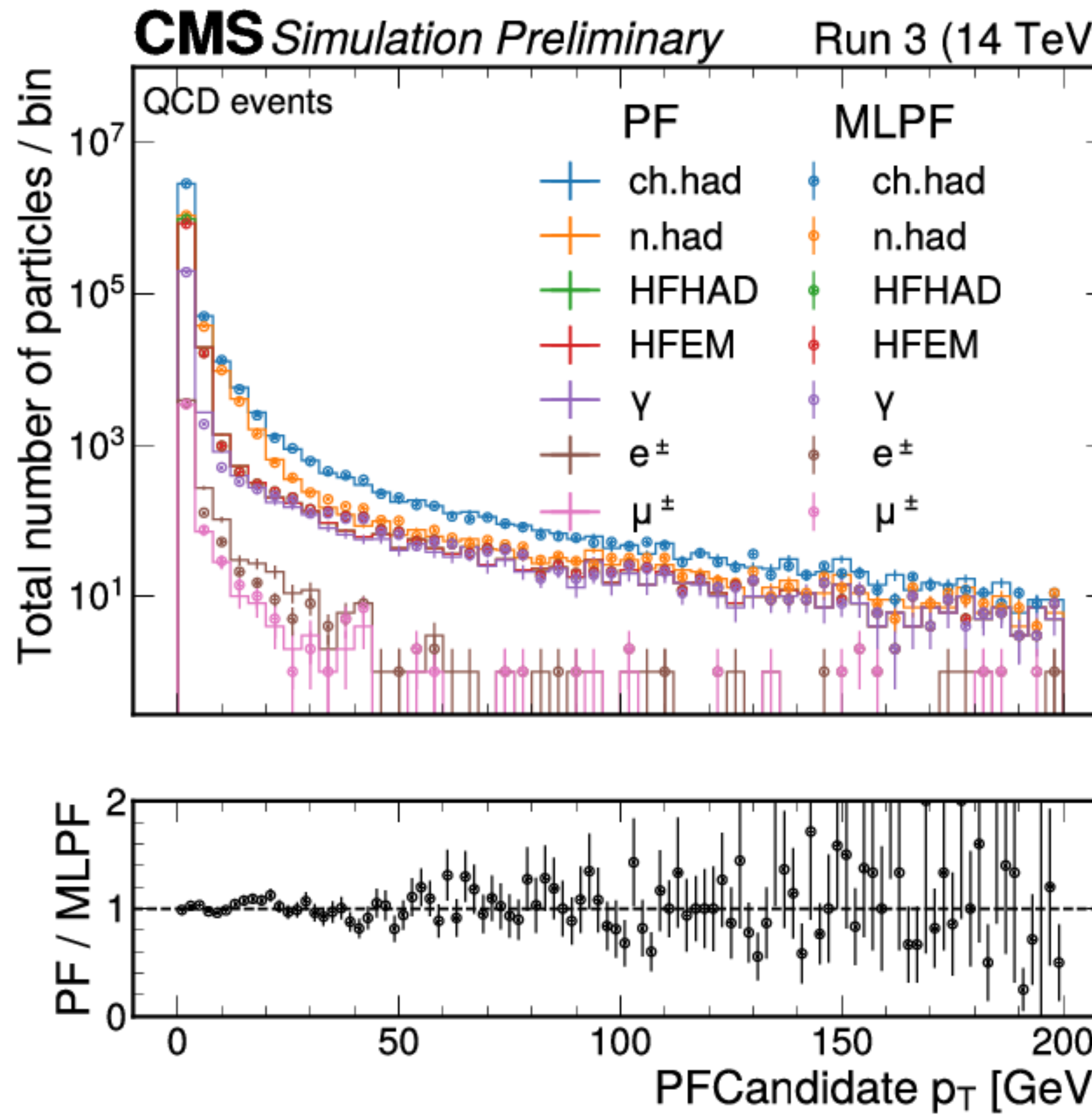
Particle distributions are improved over PF…



Also tested in a real detector (2022), now in the process of updating







JP, Javier Duarte, Farouk Mokhtar, Eric Wulff, Jieun Yoo, Jean-Roch Vlimant, Maurizio Pierini, Maria Girone. Machine Learning for Particle Flow Reconstruction at CMS. ACAT 2021. https://doi.org/10.48550/arXiv.2203.00330, http://cds.cern.ch/record/2792320

Open datasets!

Tracks and calorimeter hits

tracks & calo hits, ML format, ~5GB
- reduced stats, granular events
  - genparticles
  - tracks and calorimeter hits
  - PF candidates
- ML-friendly TFDS format

Charged particle tracking

Raw detector hits

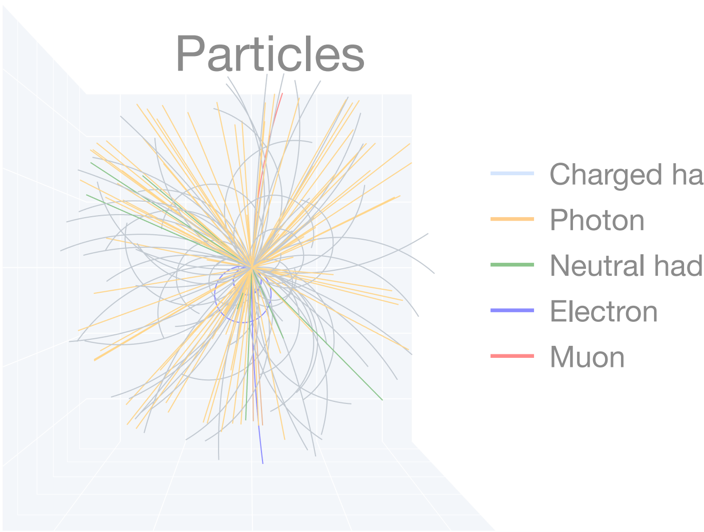
CLIC in EDM4HEP, ~2.6TB
- full stats, full details
- 5 physics samples, ~1M events each
- 7 gun samples, ~100k events each

Calorimeter clustering

Tracks and calorimeter clusters

tracks & clusters, ML format, ~50GB
- full stats, coarse events
  - genparticles
  - tracks and calorimeter clusters
  - PF candidates
- ML-friendly TFDS format

Particles



- https://doi.org/10.5281/zenodo.8260741
- https://doi.org/10.5281/zenodo.8414225
- https://doi.org/10.5281/zenodo.8409592