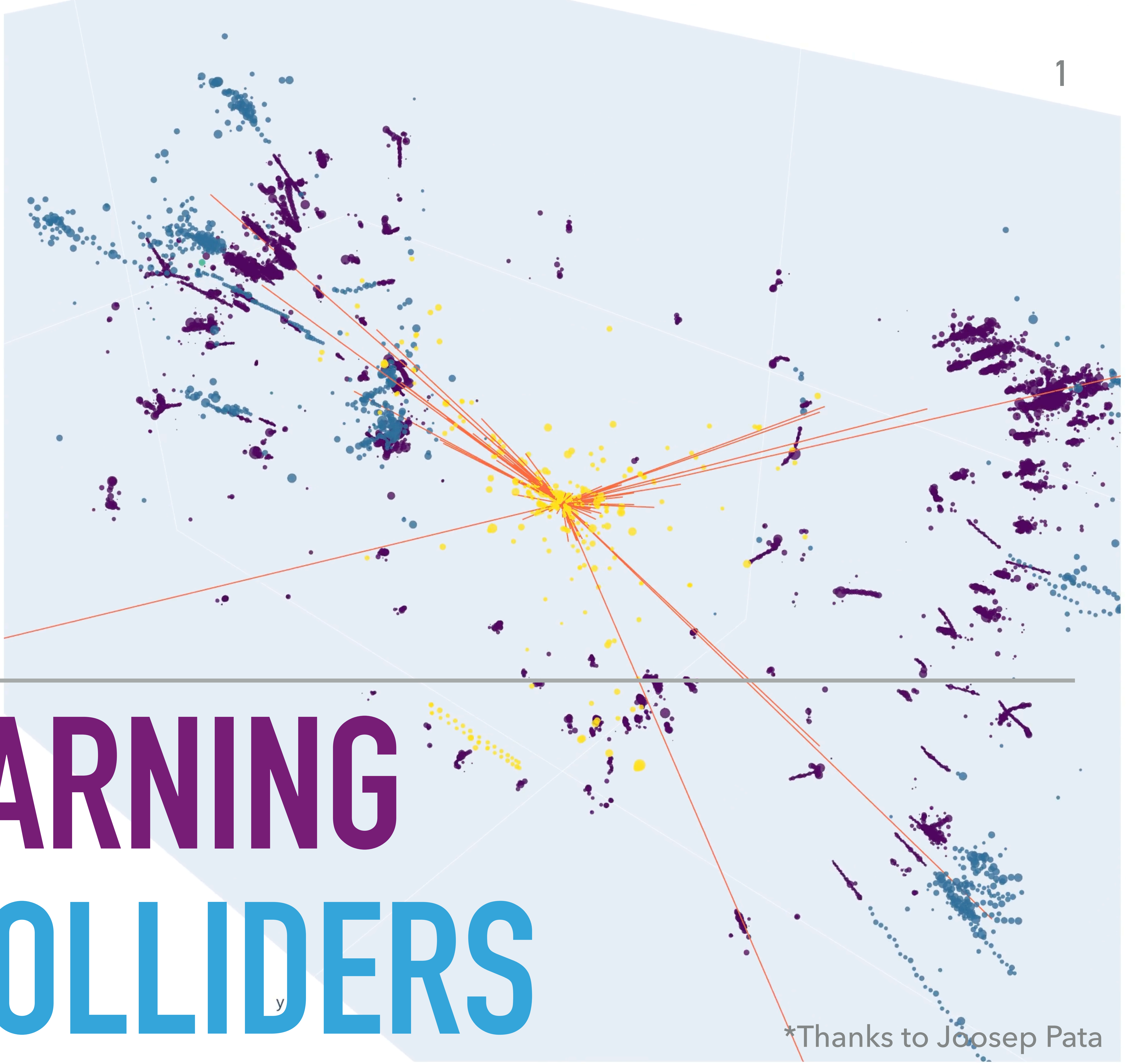
JAVIER DUARTE*

FUTURE OF HIGH ENERGY PHYSICS

ASPEN CENTER FOR PHYSICS

MARCH 29, 2024

MACHINE LEARNING AT FUTURE COLLIDERS

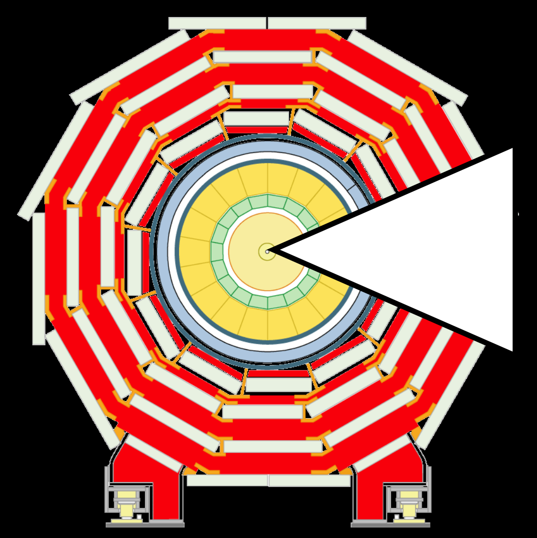


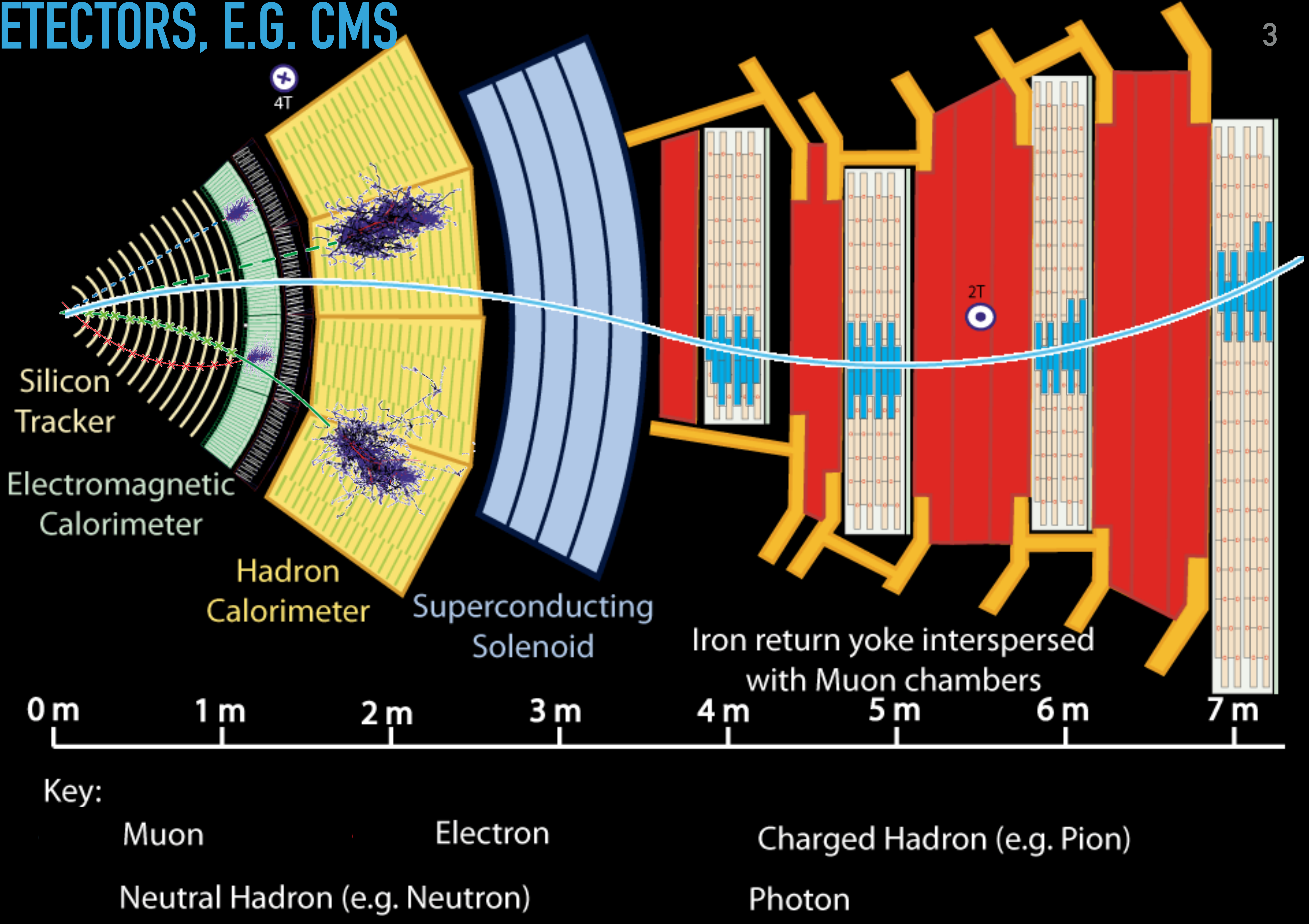
*Thanks to Joosep Pata

INTRODUCTION

- Machine learning has already changed the way we do particle physics *from trigger/data acquisition to event reconstruction, simulation, data analysis, and interpretation*
  - It is an essential and versatile tool that we use to improve existing approaches
  - It enables fundamentally new approaches
- In this talk, I'll describe one thread where ML can shift/inform the paradigm

MULTILAYERED DETECTORS, E.G. CMS



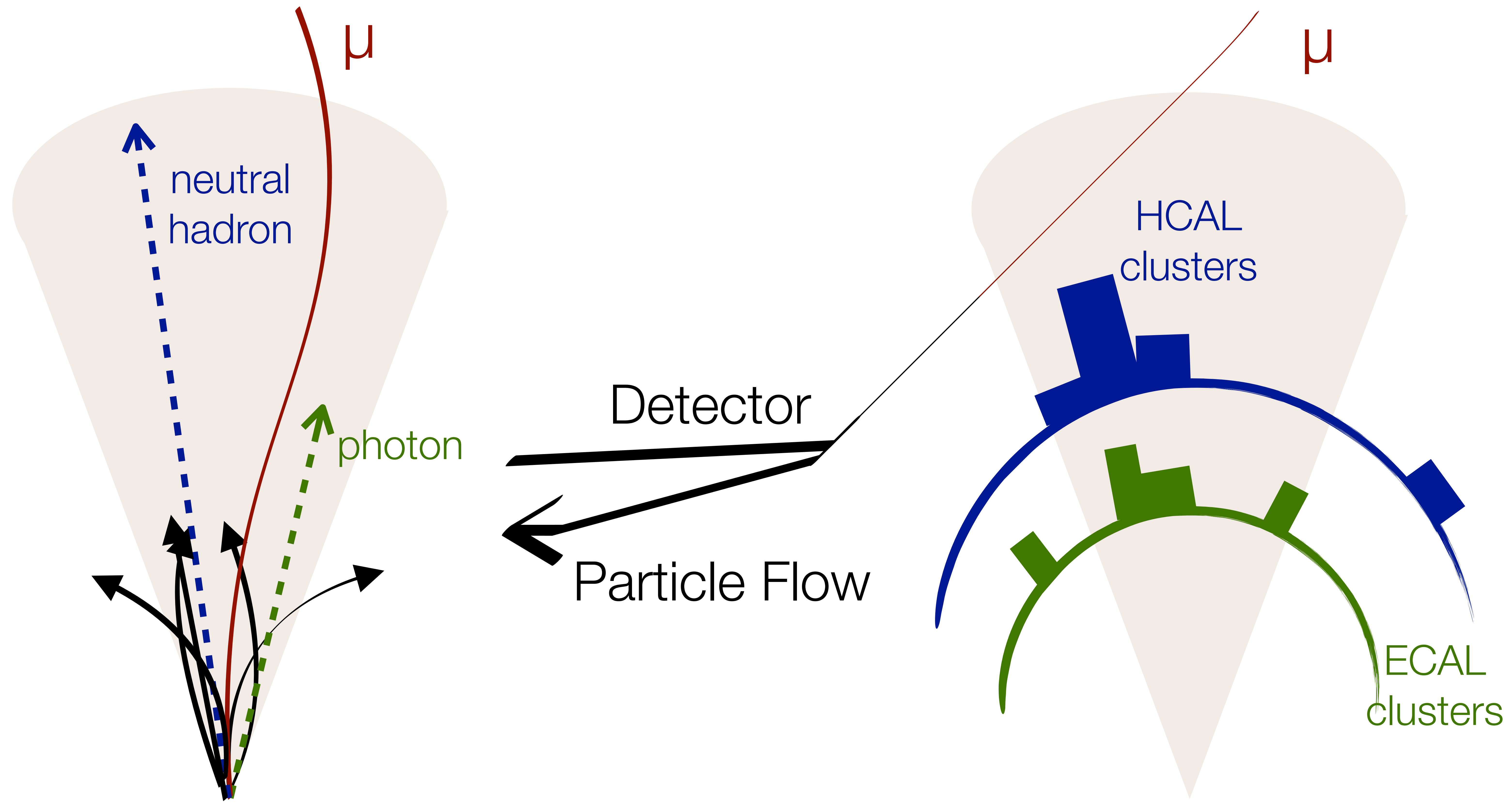


Current and future multilayered detectors…

Require complex reconstruction → particle-flow algorithm

PARTICLE-FLOW RECONSTRUCTION

- Particles interact with detector, leaving energy deposits and tracks
- Efficient combination of info. from complementary detector subsystems to produce a holistic, particle interpretation of the event (that improves on any individual subsystem)



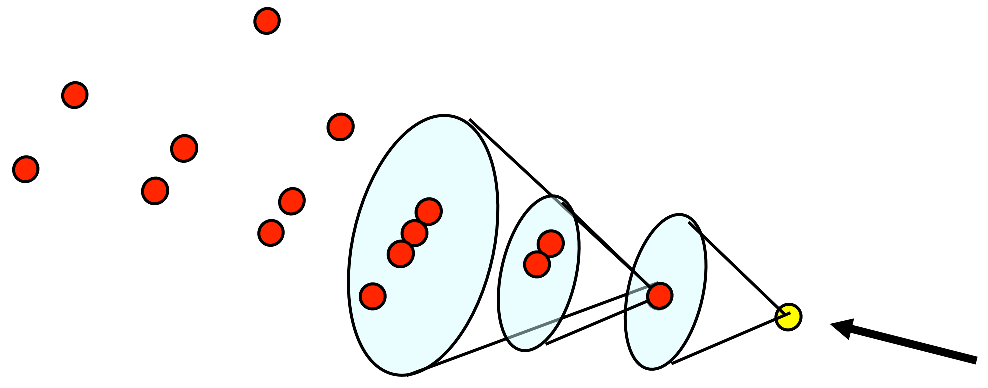
CONVENTIONAL PARTICLE-FLOW, E.G. PANDORA

J. S. Marshall: https://indico.in2p3.fr/event/7691/contributions/42712

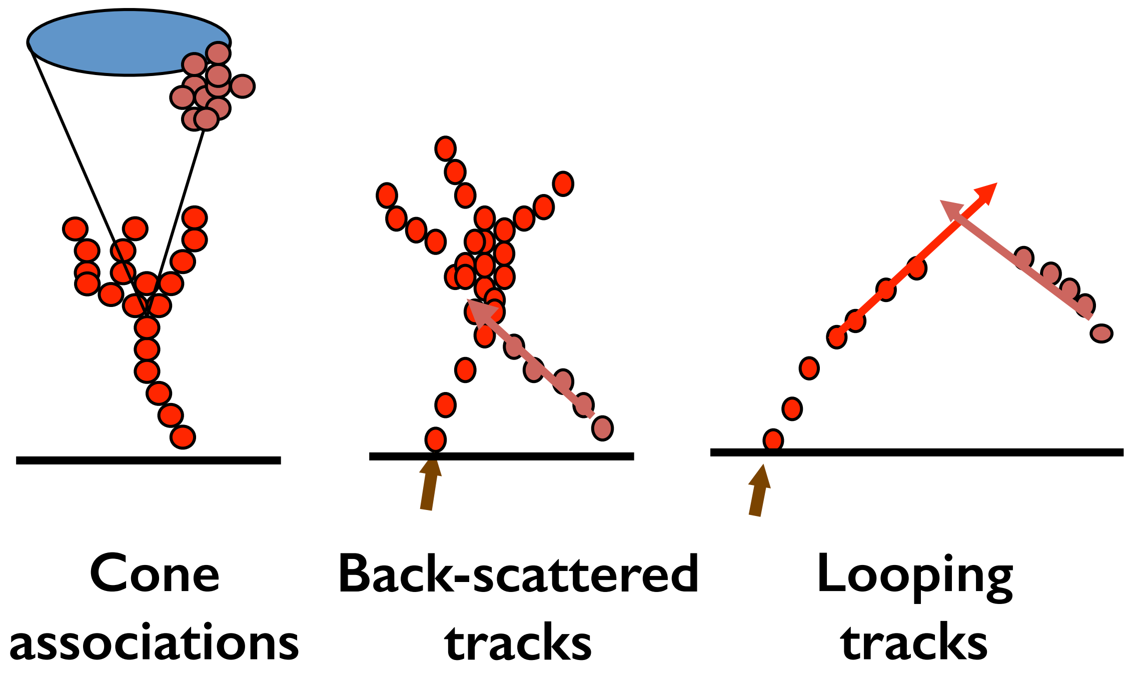
Existing particle-flow algorithms based on complex, hand-tuned heuristics work well

"Basic" reconstruction uses 56 algorithms:

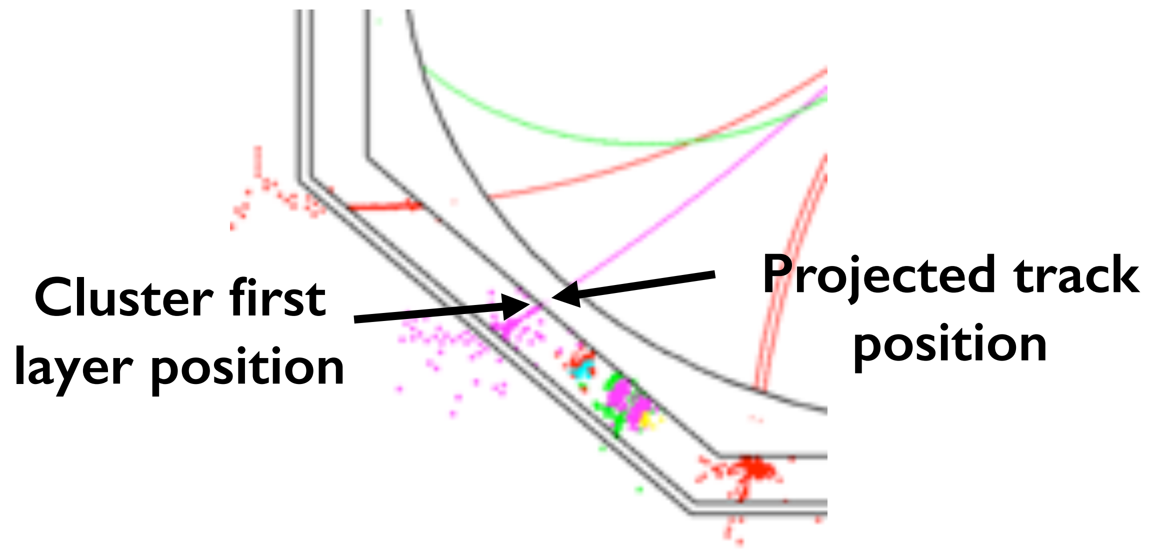
ConeClustering Algorithm



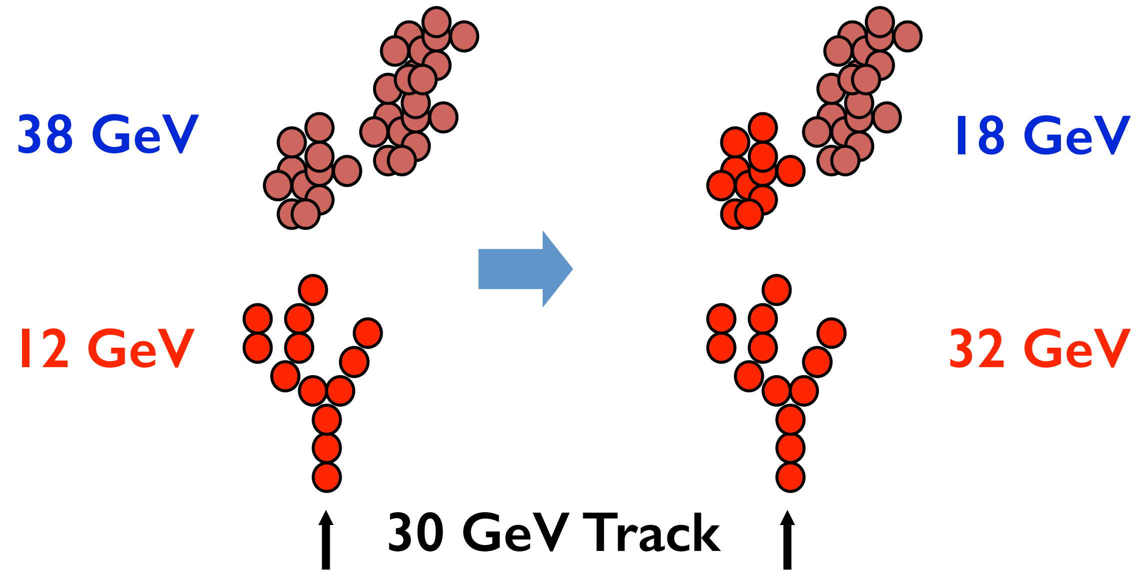
Topological Association Algorithms



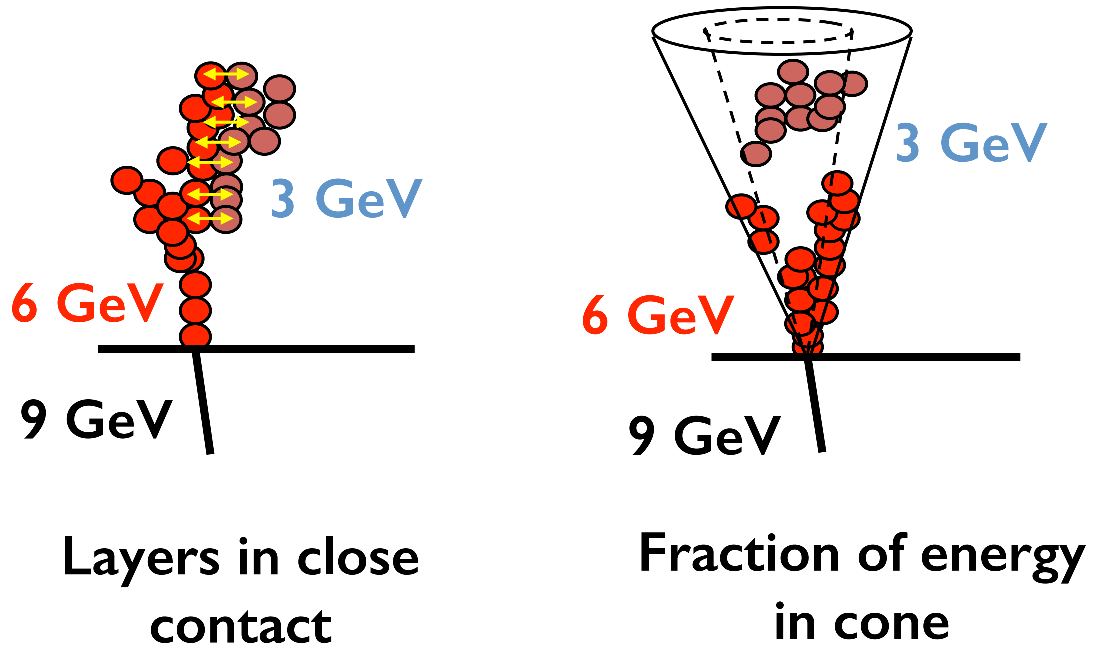
Track-Cluster Association Algorithms



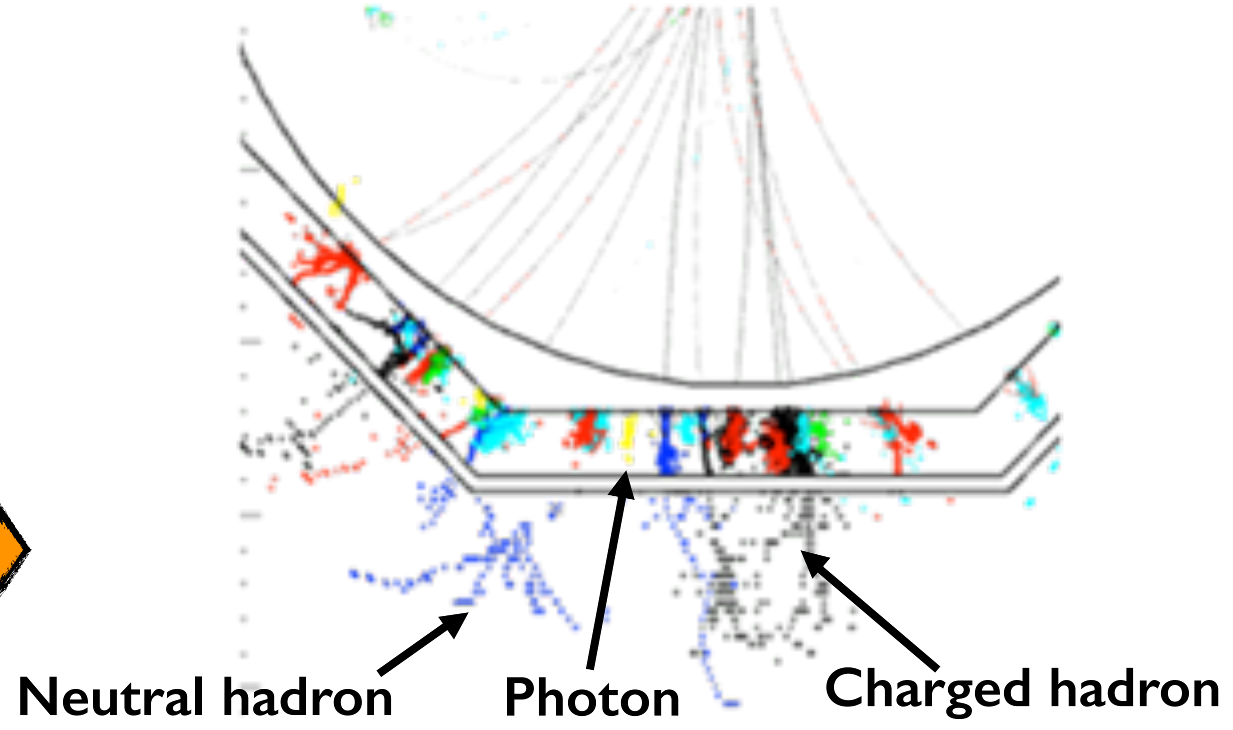
Reclustering Algorithms



Fragment Removal Algorithms



PFO Construction Algorithms

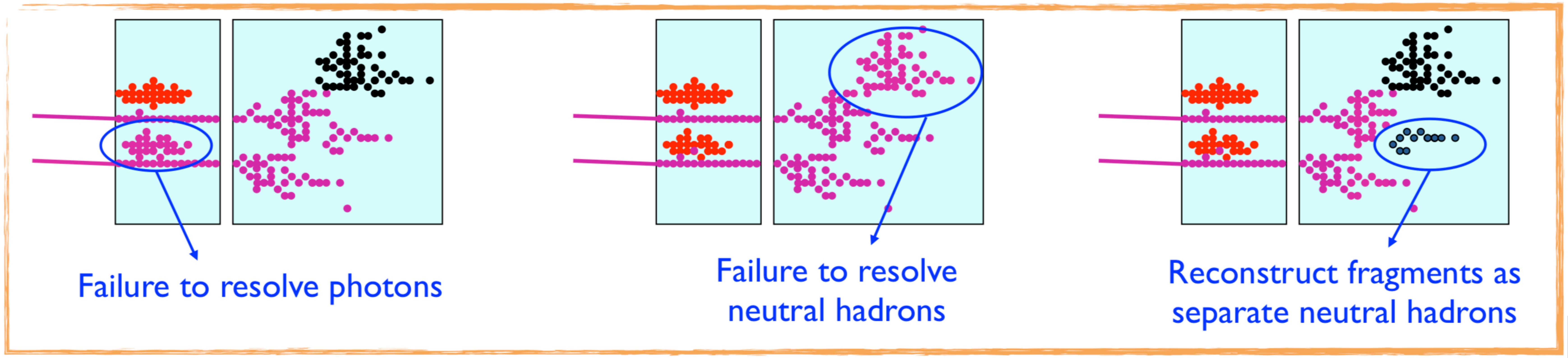


DISADVANTAGES OF CONVENTIONAL PARTICLE-FLOW

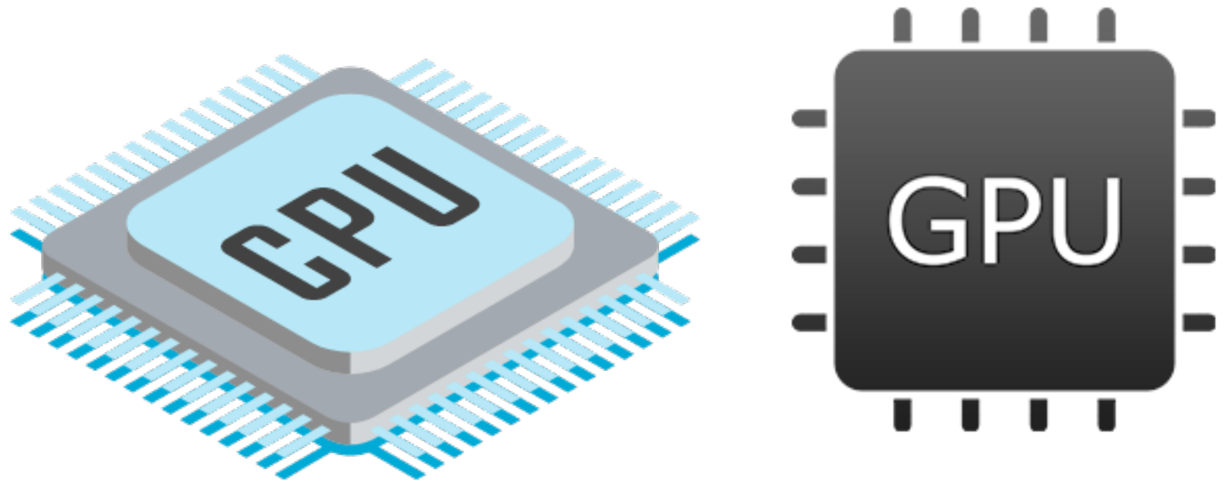
J. S. Marshall: https://indico.in2p3.fr/event/7691/contributions/42712

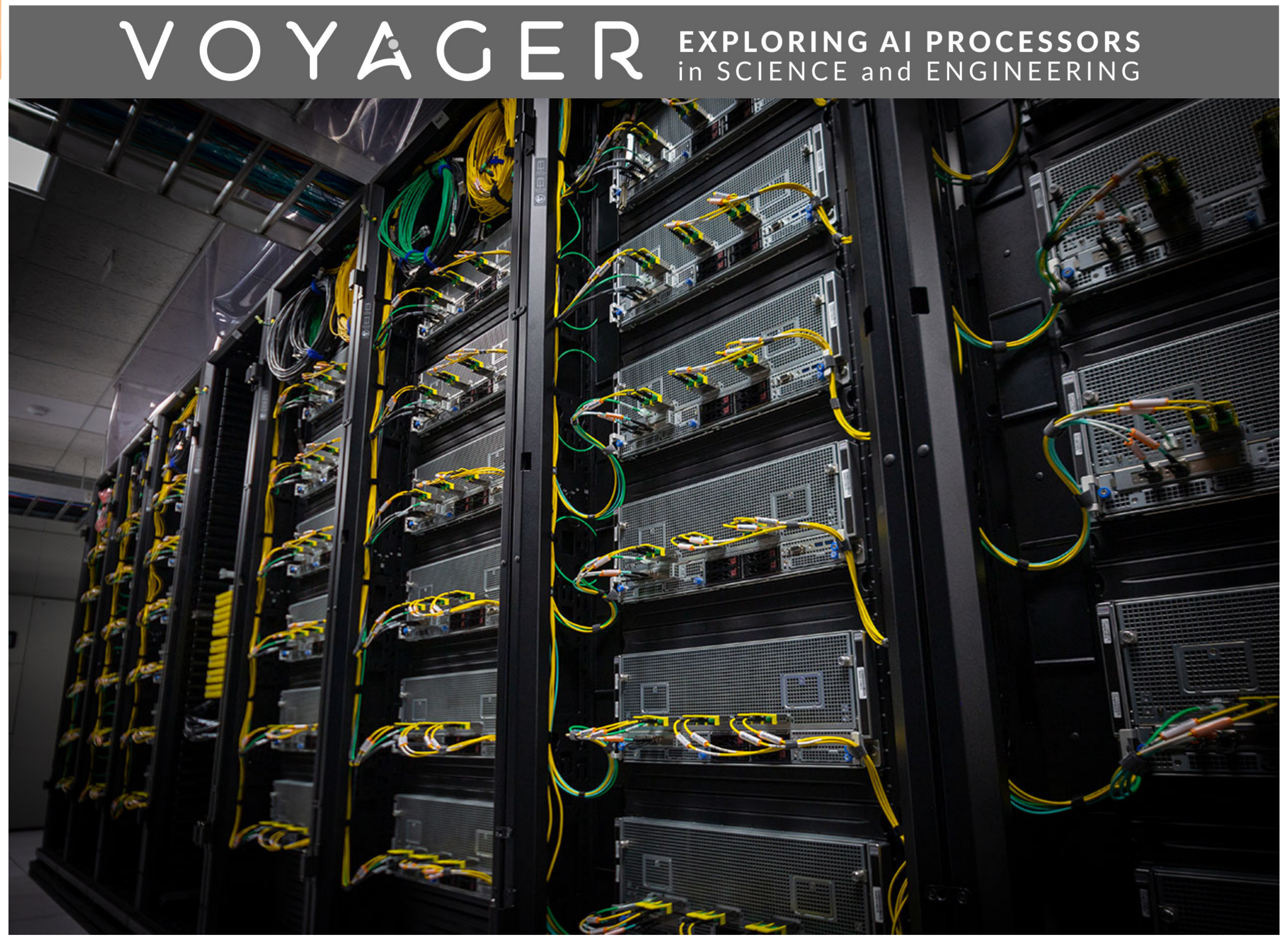
- Our heuristics fail in some ambiguous situations
- Traditional PF algorithms can be tricky to extend, tune, apply to different/new detectors, or port to new computational hardware or HPCs

Types of confusion



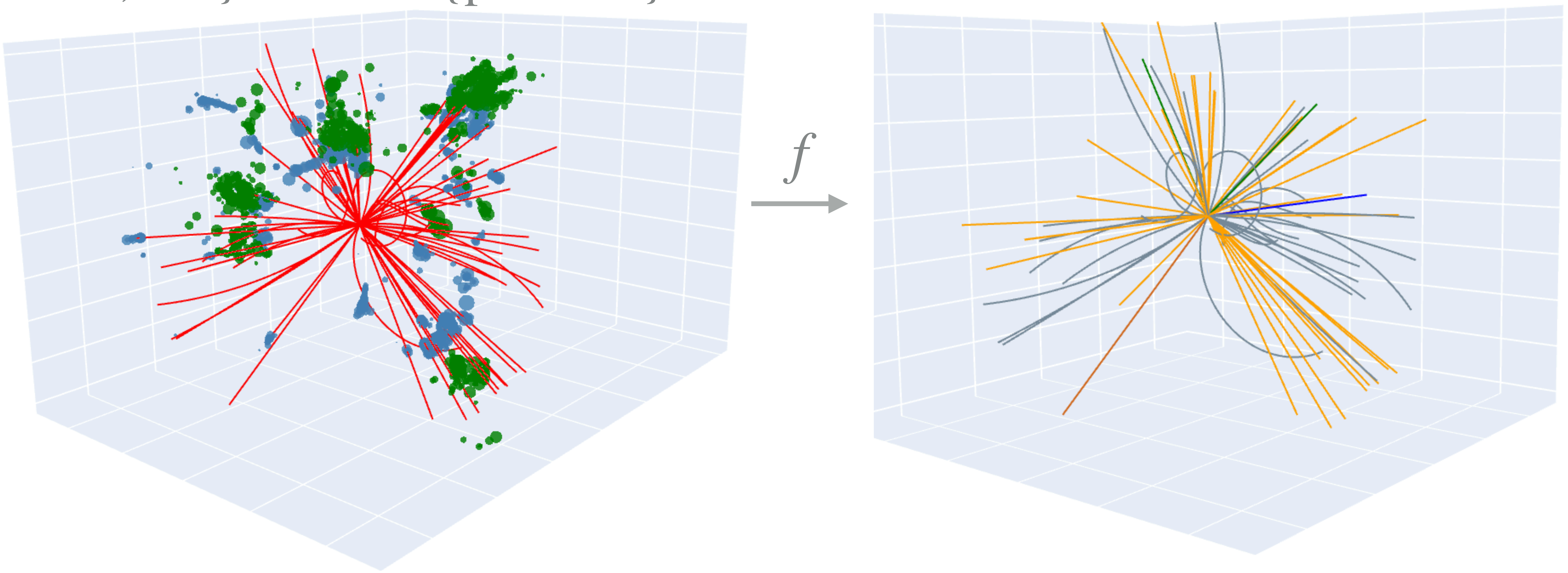
Missing energy | **Missing energy** | **Double counted energy**





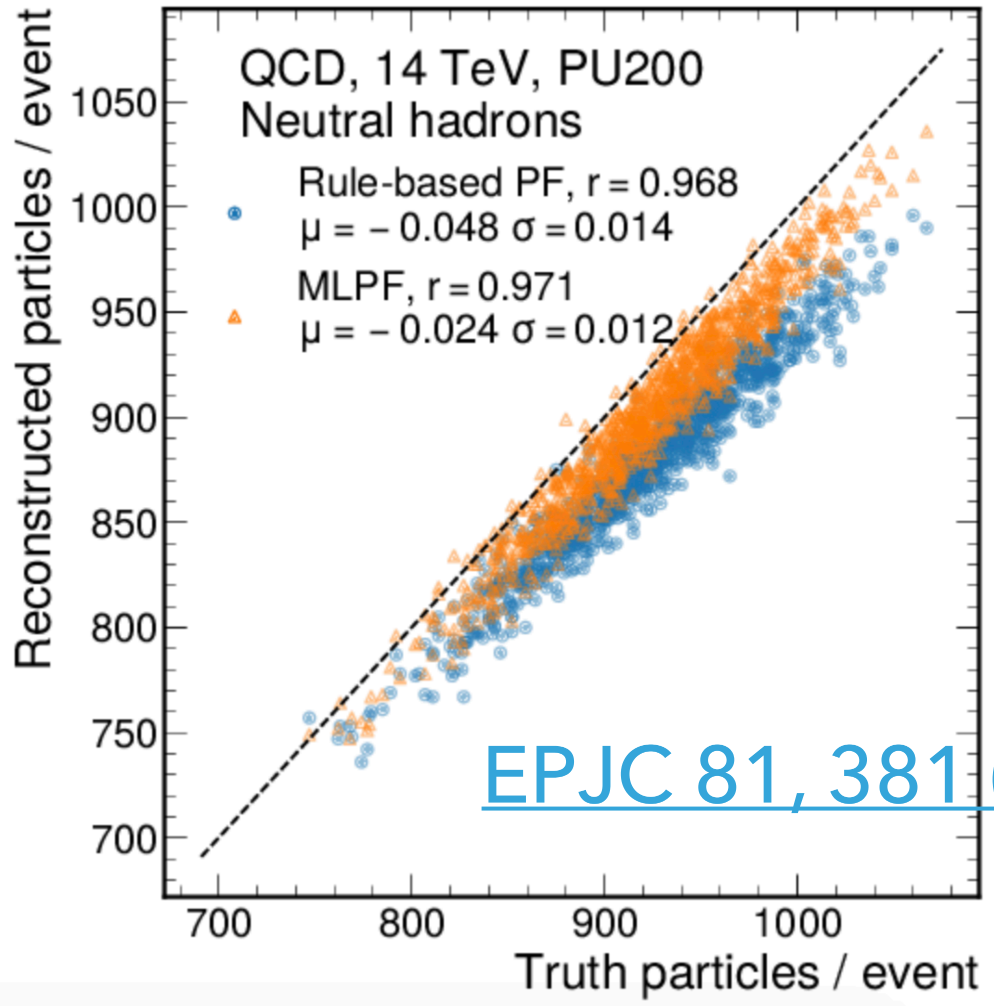
PARTICLE-FLOW AS A MACHINE-LEARNING TASK

- Can we instead formulate PF as an ML task (naturally "tunable" through re-training and portable to new hardware)?
- Learn a "set-to-set" function $f \colon X \rightarrow Y$, where $\{\text{tracks}, \text{clusters}\} \in X$ or $\{\text{tracks}, \text{hits}\} \in X$ and $\{\text{particles}\} \in Y$



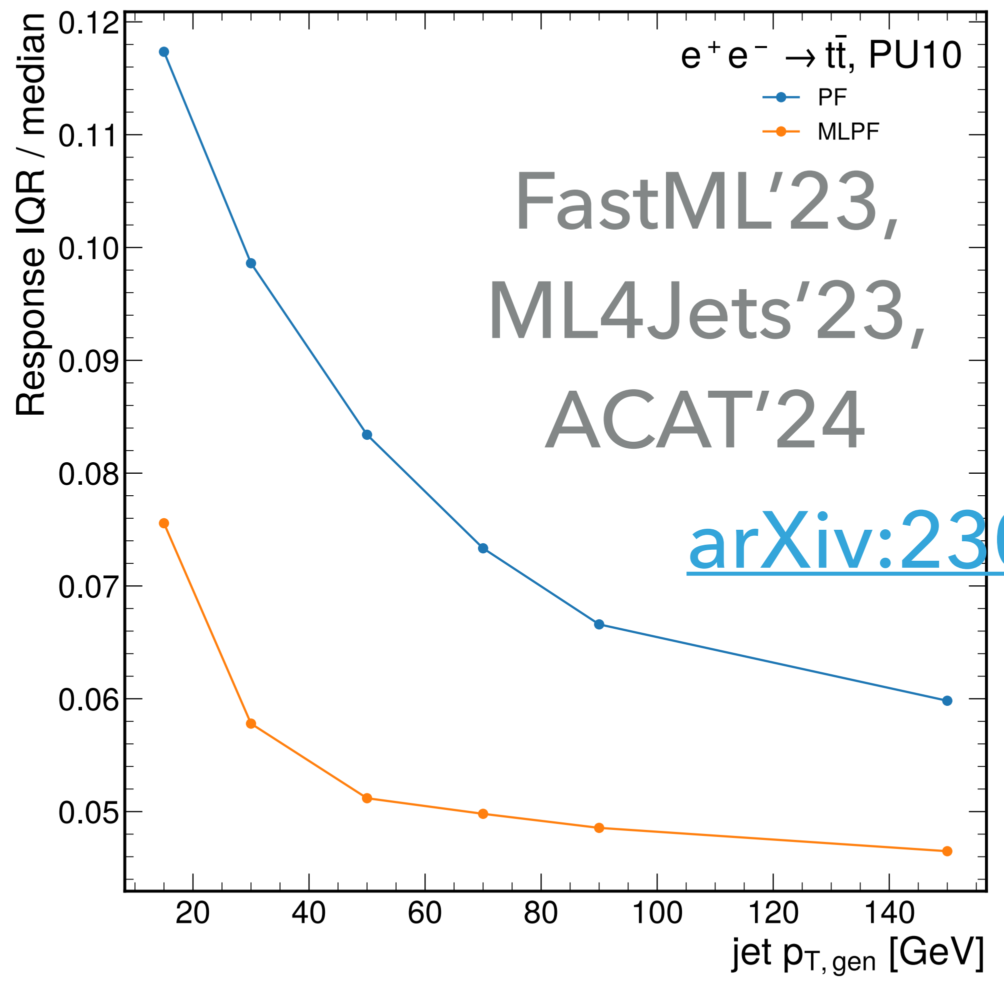
$\{\text{tracks}, \text{hits}\} \in X$ $\{\text{particles}\} \in Y$

MLPF TIMELINE



pp, generic detector,
proof of concept on Delphes sim.

[EPJC 81, 381 (2021)]



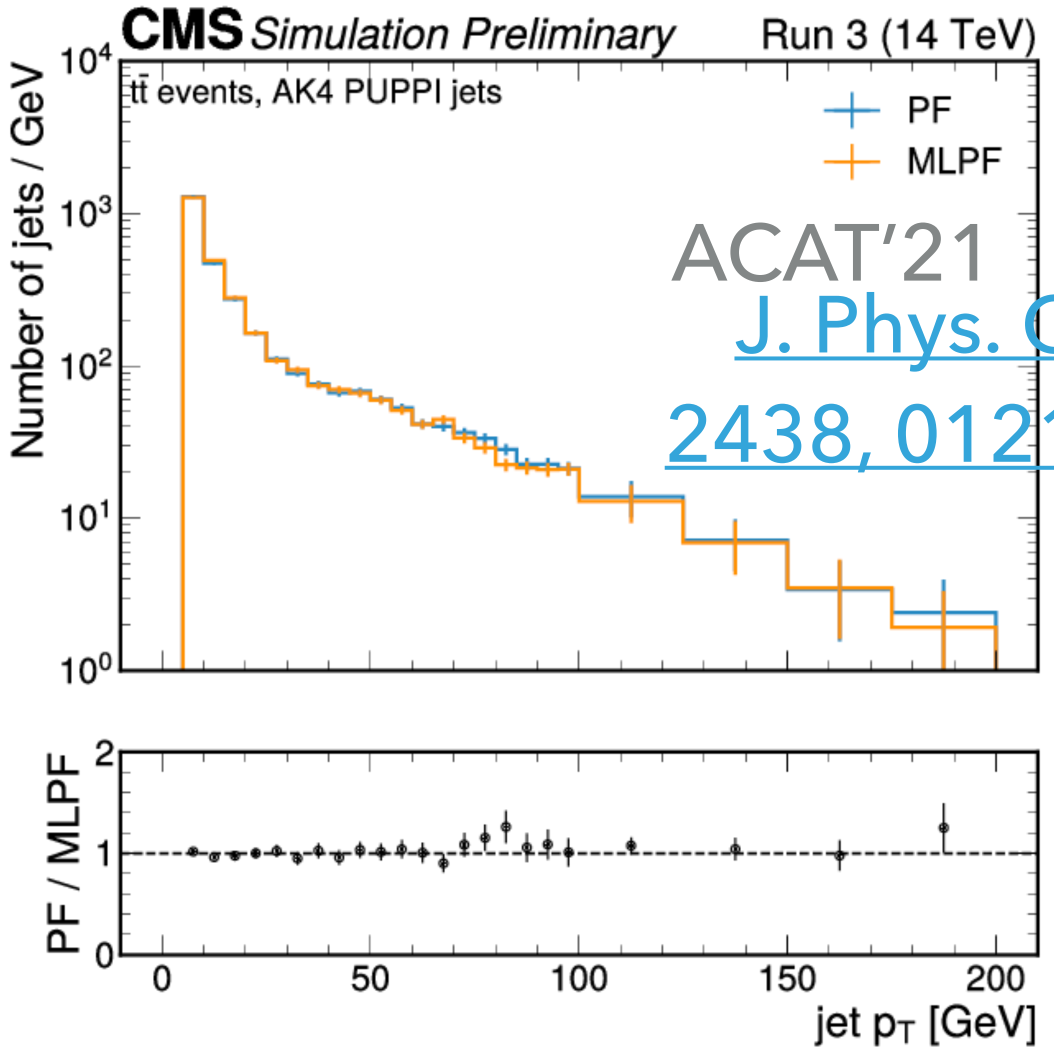
FastML’23,
ML4Jets’23,
ACAT’24

[arXiv:2309.06782]

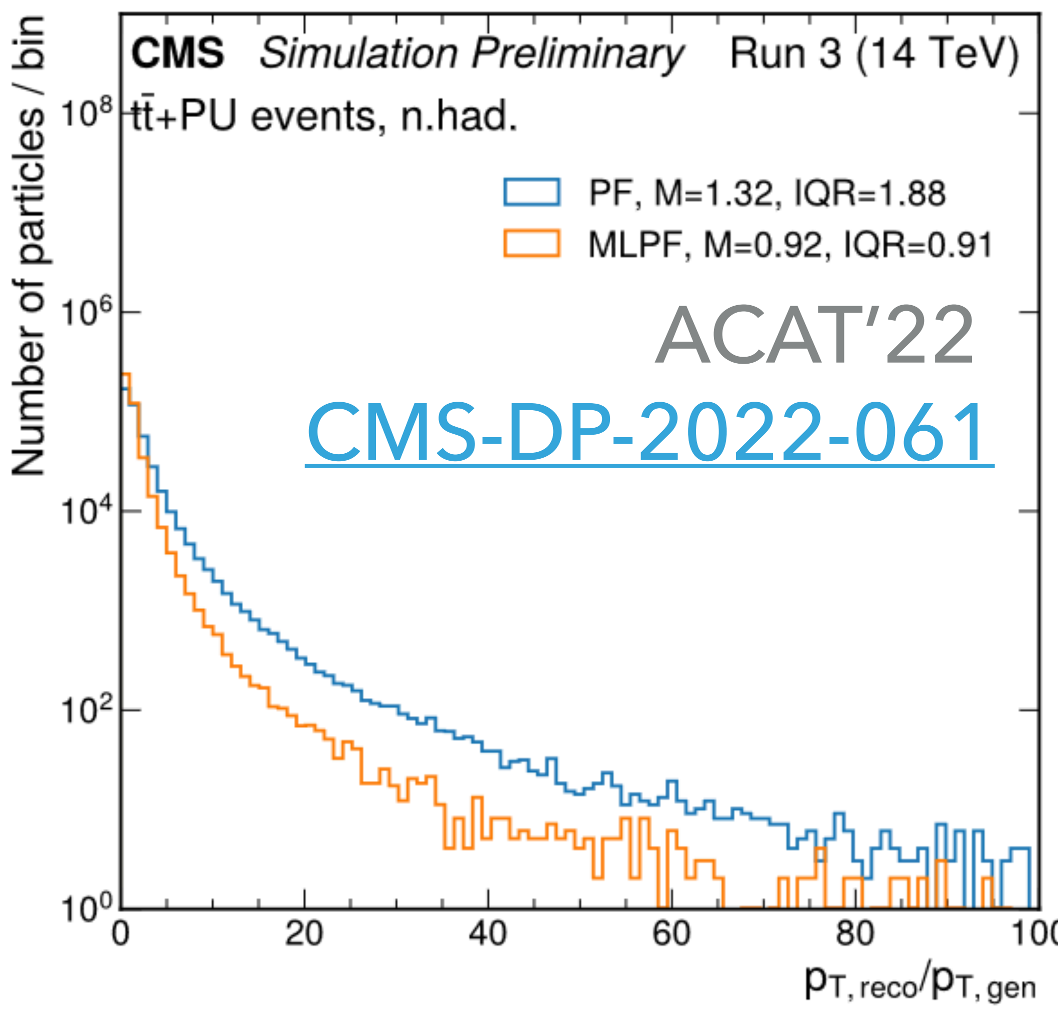
e⁺e⁻, CLIC detector,
event-level SOTA on full sim.

2020 2021 2022 2023 2024

pp, CMS detector,
proof of concept on full sim.



ACAT’21
[J. Phys. Conf. Ser. 2438, 012100 (2023)]



ACAT’22
[CMS-DP-2022-061]

pp, CMS detector,
particle-level SOTA on full sim.

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

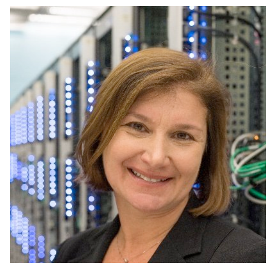
Joosep Pata[1*], Eric Wulff[2], Farouk Mokhtar[3], David Southwick[2], Mengke Zhang[3], Maria Girone[2], Javier Duarte[3]

[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;







OPEN DATASET FOR ML RECONSTRUCTION STUDIES

doi:10.5281/zenodo.8260741

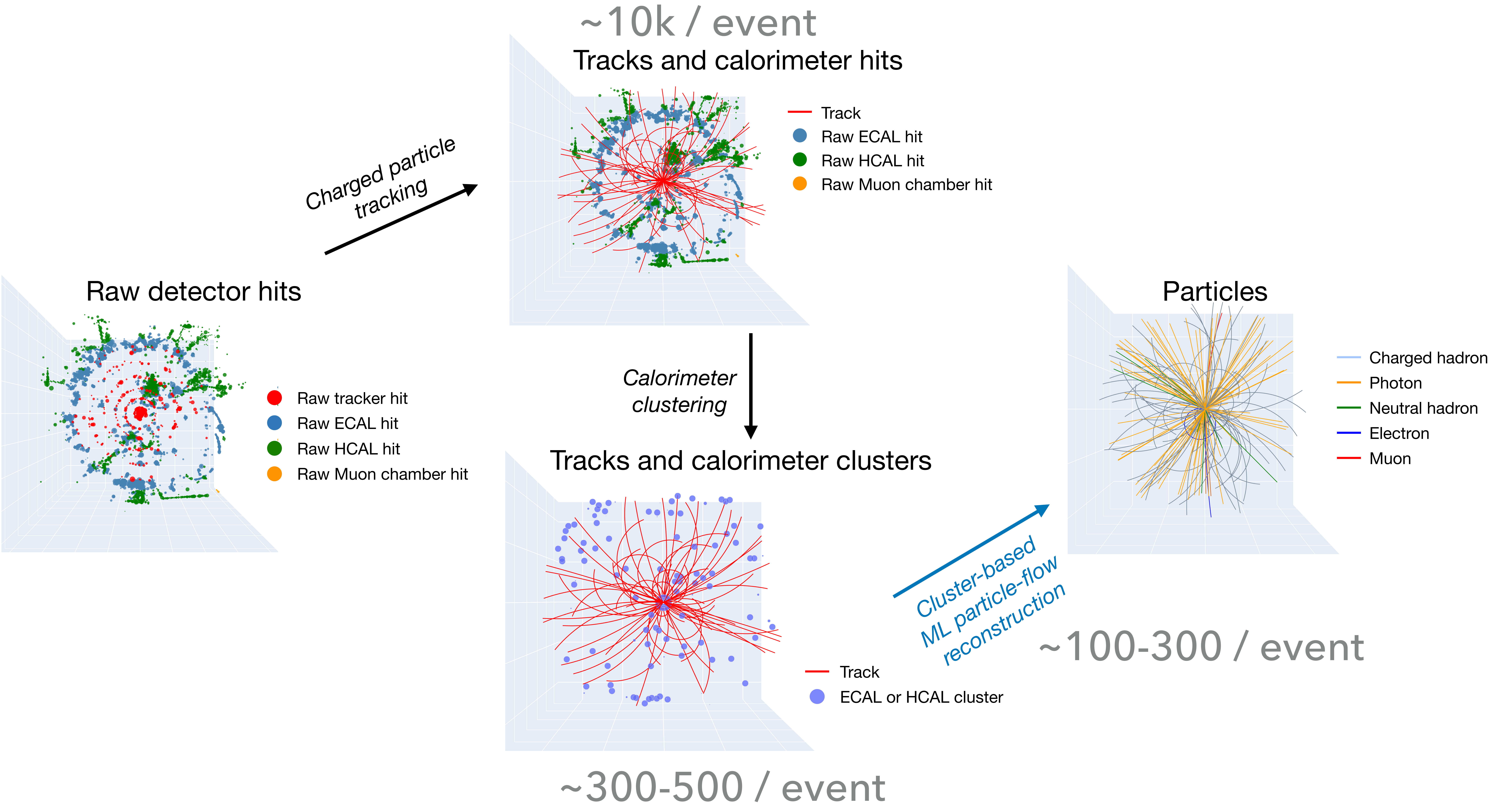
- Gen. particles, reco. tracks and calorimeter hits, reco. Pandora PF particles in EDM4HEP format
- CLIC detector (CLIC_o3_v14) simulation with Geant4, reco. with Marlin interfaced via Key4HEP including Pandora PF reco.
- Processes generated with Pythia8 at $\sqrt{s} = 380\,\mathrm{GeV}$
  - $e^+e^- \rightarrow t\bar{t}, q\bar{q}, ZH(\tau\tau), WW, t\bar{t} + \mathrm{PU10}$
  - Single-particle: $e^{\pm}, \mu^{\pm}, K_L^0, n, \pi^{\pm}, \gamma$ between $[1{,}100]\,\mathrm{GeV}$
- 2.5 TB, 6 million events in total

Particle Flow Reconstruction

Scalable Neural Network Models and Terascale Datasets

https://www.coe-raise.eu/od-pfr

OPEN DATASET FOR ML RECONSTRUCTION STUDIES

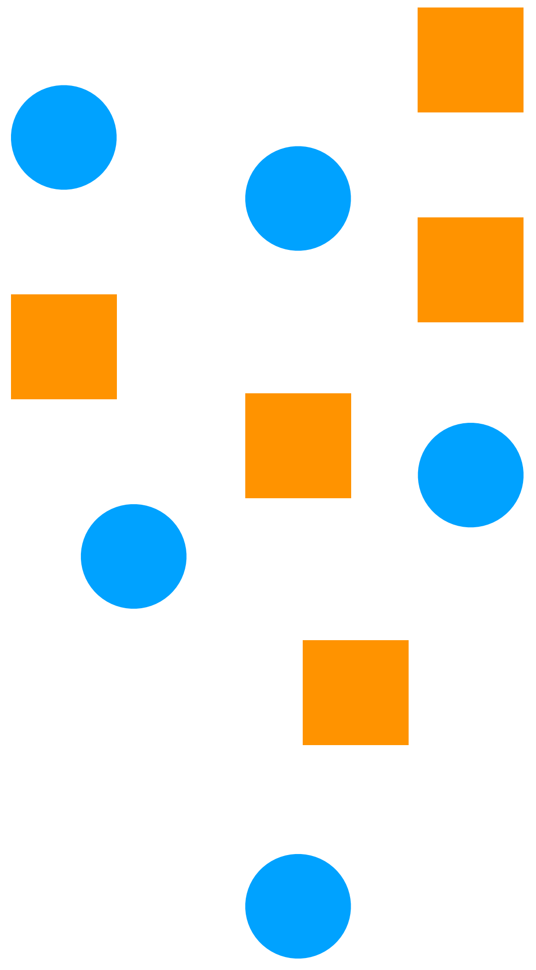


DATA PREPROCESSING FOR NODE PREDICTION

Eur. Phys. J. C 81, 381 (2021)

- Formulate task as node prediction with data preprocessing
- Input set arranged in (arbitrarily ordered) matrix

Input set $X = \{x_i\}$

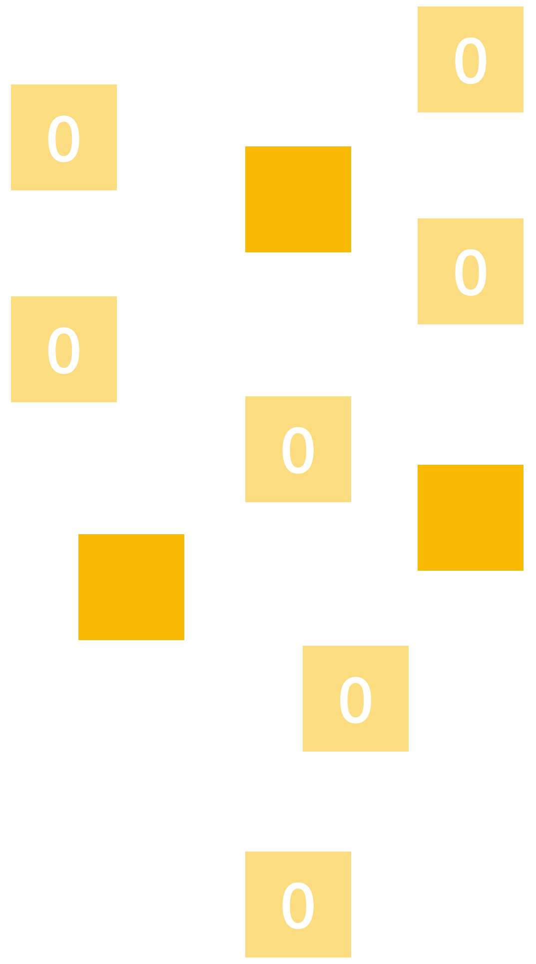


$$x_i^{\text{track}} = \left[p_{\text{T}}, \eta, \phi, \chi^2, N_{\text{dof}}, \tan\lambda, D_0, \Omega = \text{sign}(q)/R, Z_0\right]$$

$$x_i^{\text{cluster}} = \left[E_{\text{T}}, \eta, \phi, E_{\text{ECAL}}, E_{\text{HCAL}}, x, y, z, N_{\text{hit}}, \sigma_x, \sigma_y, \sigma_z\right]$$

- Target set zero-padded to same size $|X| = |Y|$, with each output particle arranged in same array position as best-matched input element

Target set $Y = \{y_i\}$

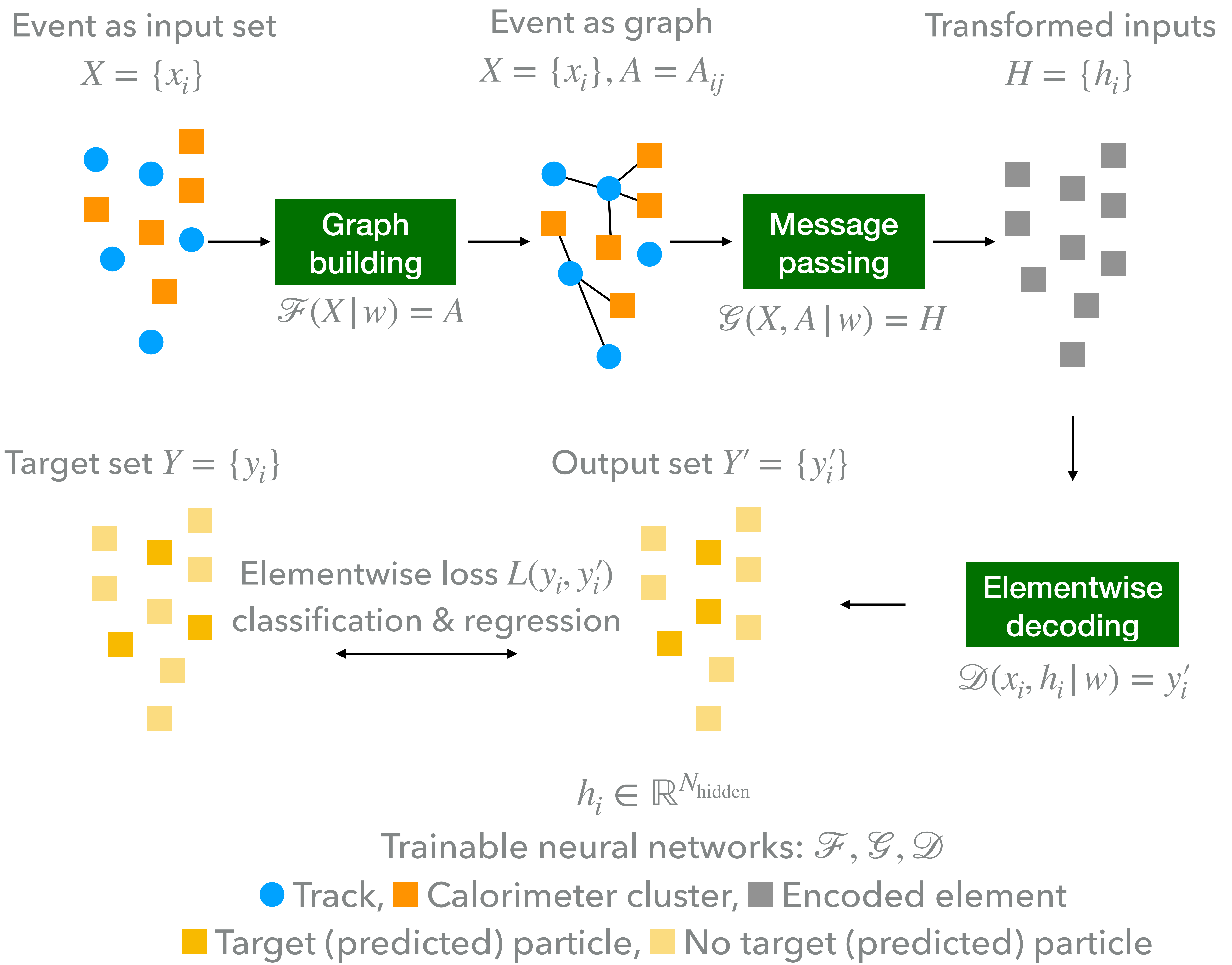


$$y_i = \left[\text{PID}, p_{\text{T}}, E, \eta, \phi, q\right]$$

$\text{PID} \in \{\text{none}, \text{charged hadron}, \text{neutral hadron}, \gamma, e^{\pm}, \mu^{\pm}\}$

GRAPH NEURAL NETWORK APPROACH

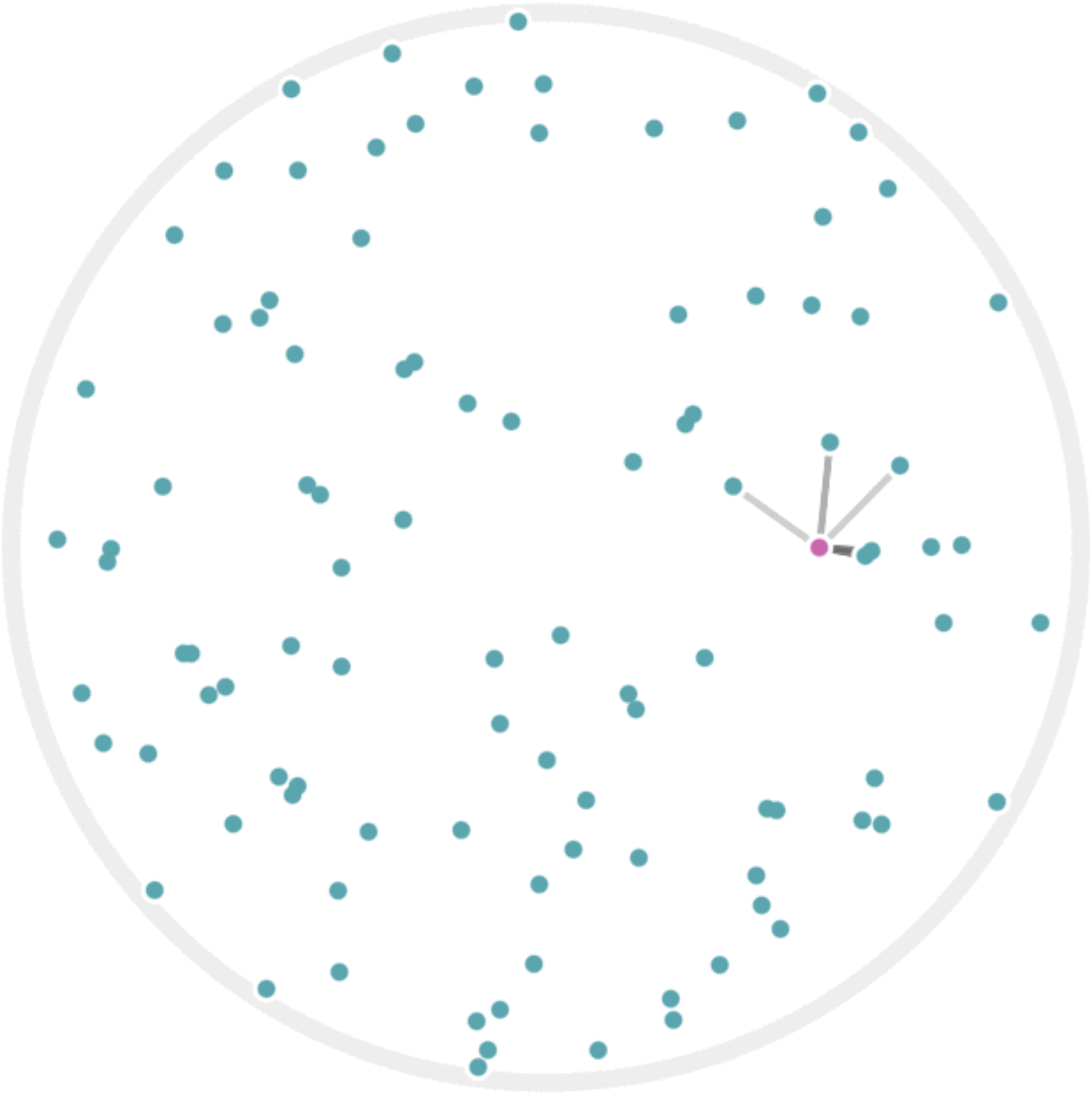
- Convert input set to a locally, **sparsely connected** graph
- Message-passing NN to transform features
- Decode transformed inputs elementwise
- (During training) Compare to target set, optimize weights



GRAPH BUILDING COMPLEXITY

T. Neylon, https://unboxresearch.com/articles/lsh_post1.html

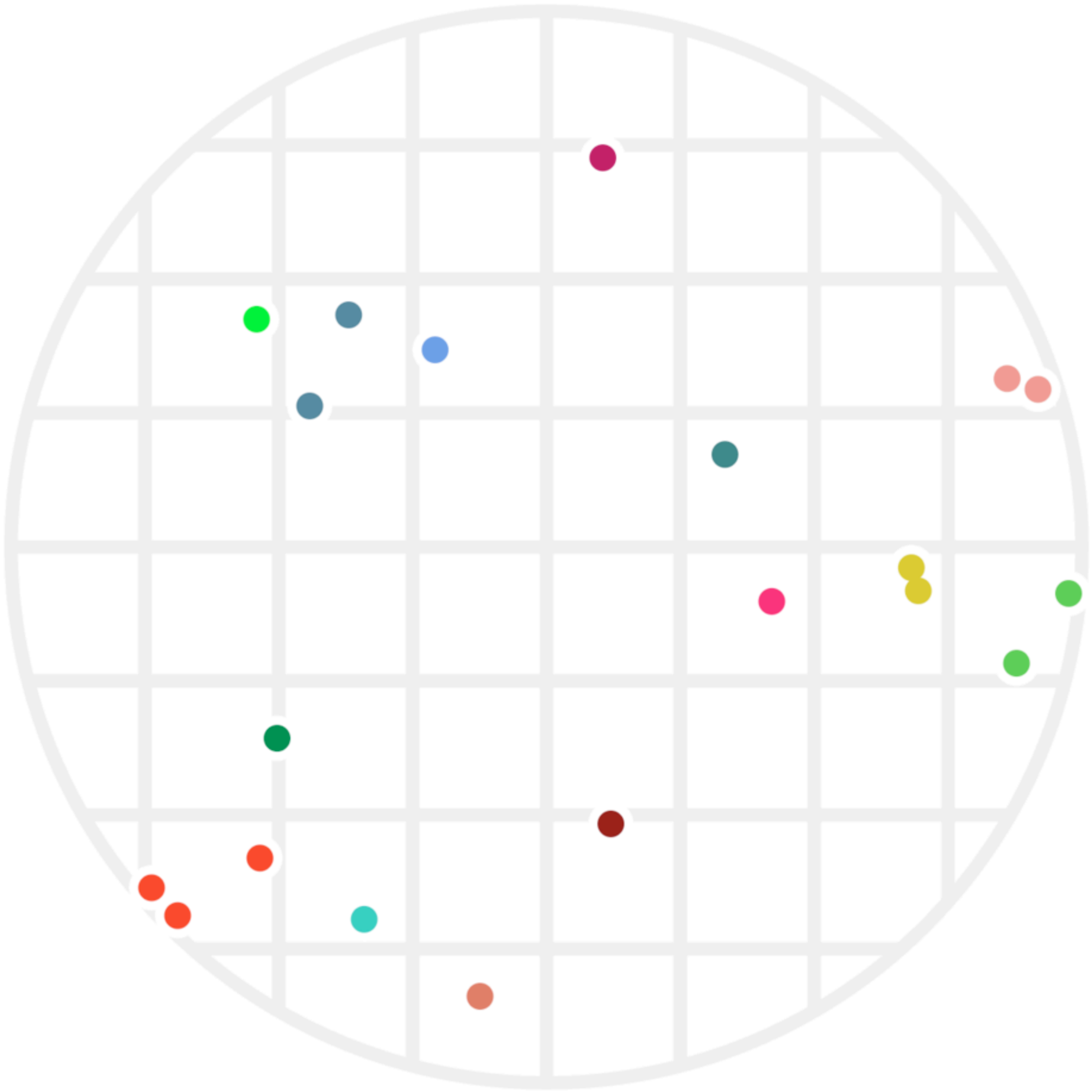
Naive nearest neighbors graph building: need to compare each pair of particles, $\mathcal{O}(N^2)$ complexity



$\mathcal{O}(N^2)$ complexity plagues other SOTA ML approaches like *transformers*

LOCALITY-SENSITIVE HASHING

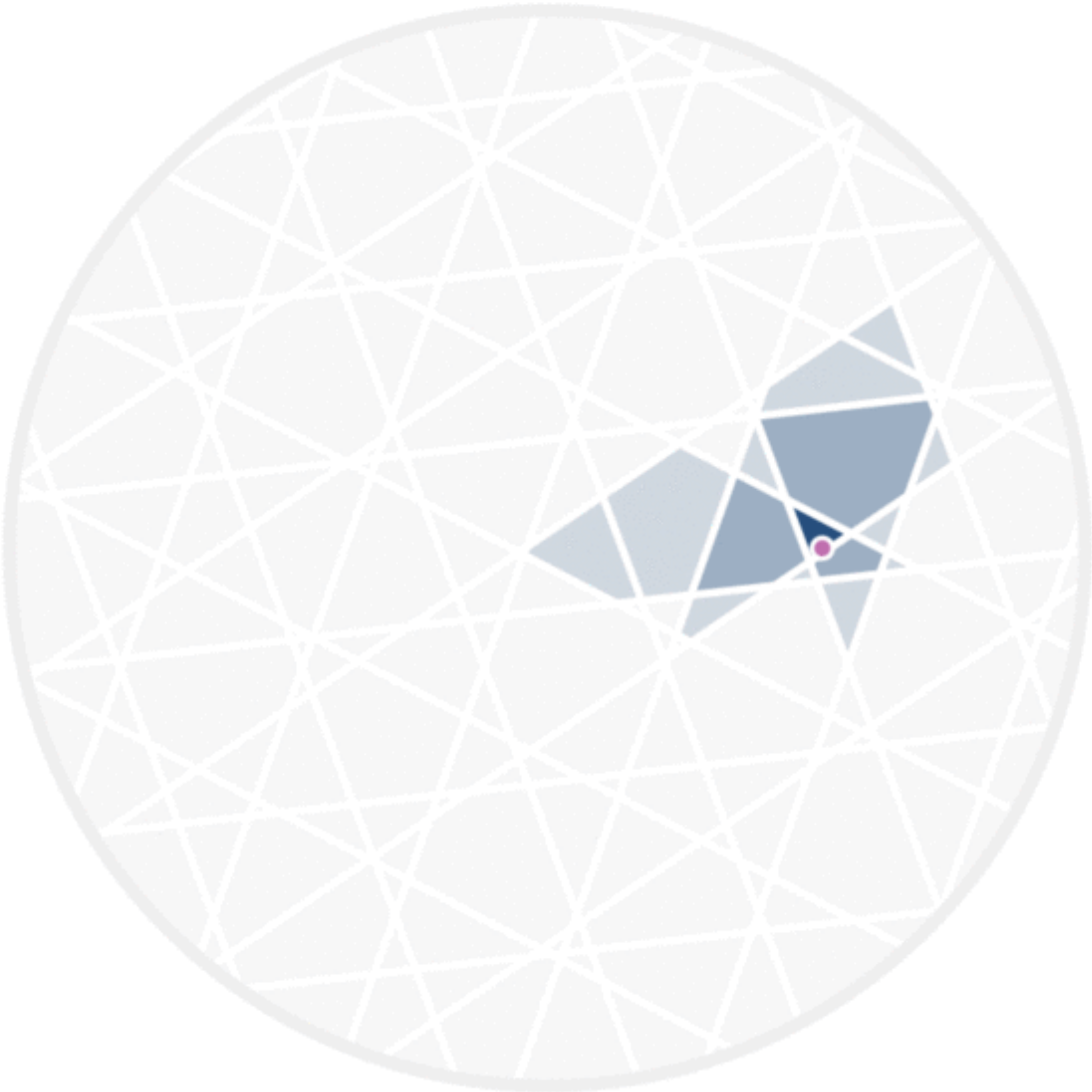
Divide space into bins, particles are *connected* if they are in the same bin



Hash function: particle features to bin index

T. Neylon, https://unboxresearch.com/articles/lsh_post1.html

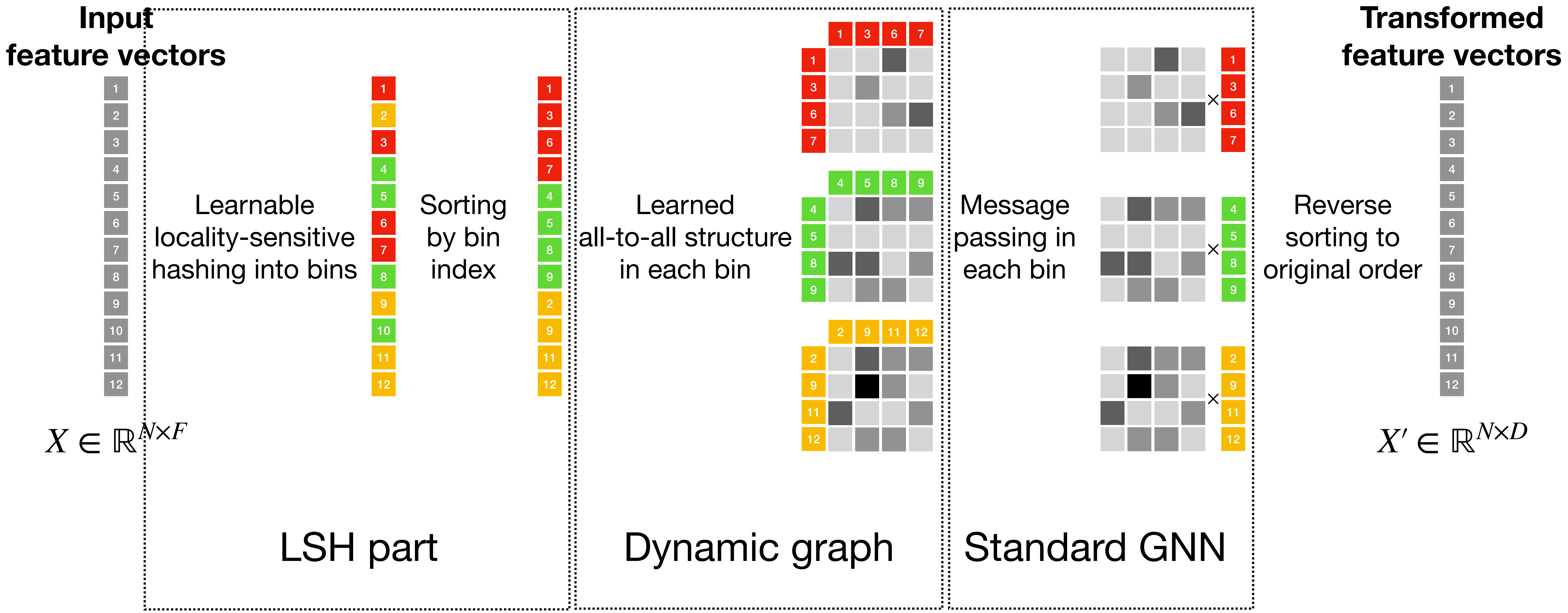
Randomized bins (hash functions) work even better!



5 random hash functions, where darkest blue = 5 hash collisions, lightest blue = 3 hash collisions

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

LSH-BASED GNN



- One layer of scalable GNN based on Reformer [arXiv:2001.04451]
- Can stack them to form multilayered network that learns higher-level representations

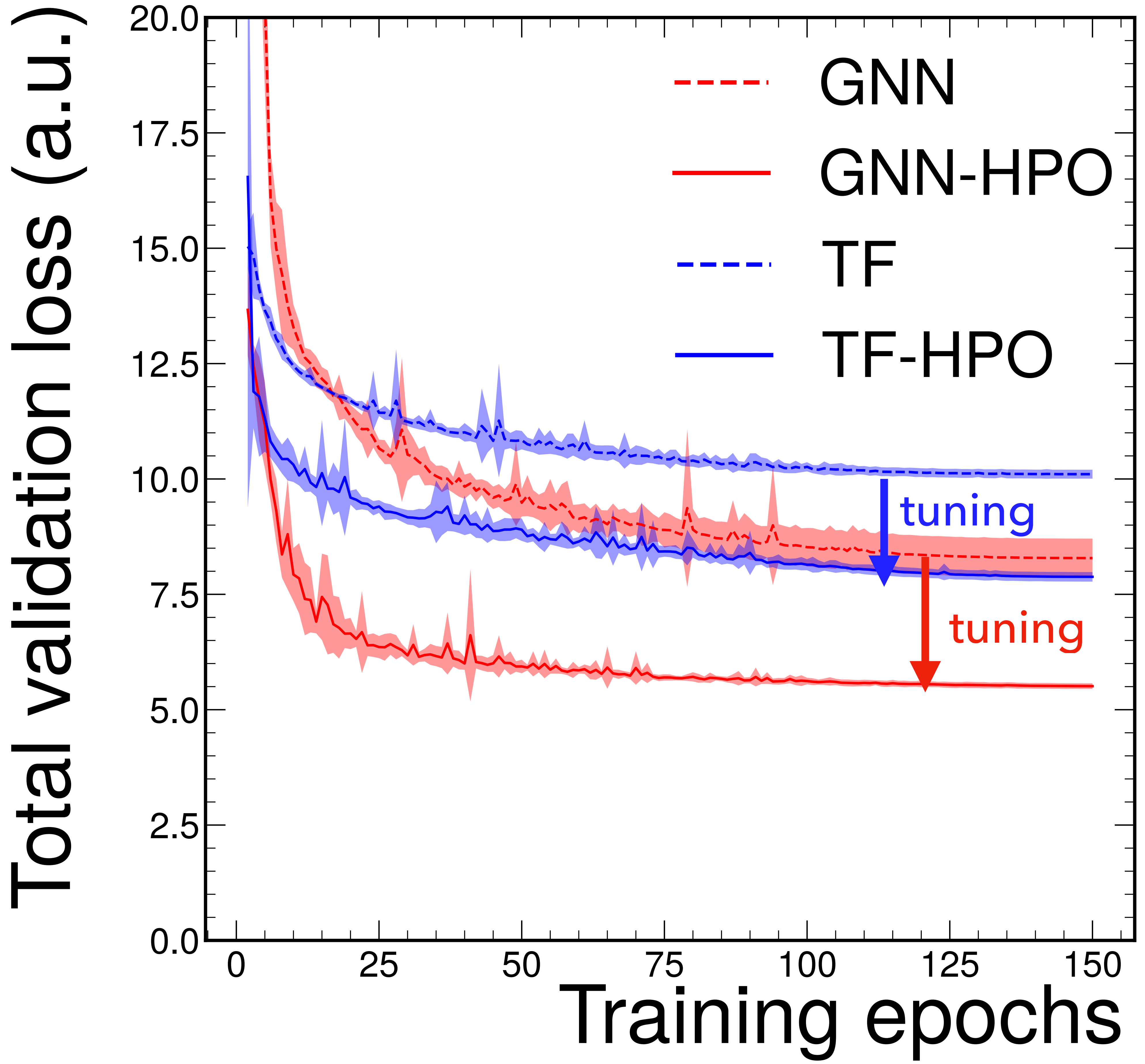
HYPERPARAMETER OPTIMIZATION ON HPC

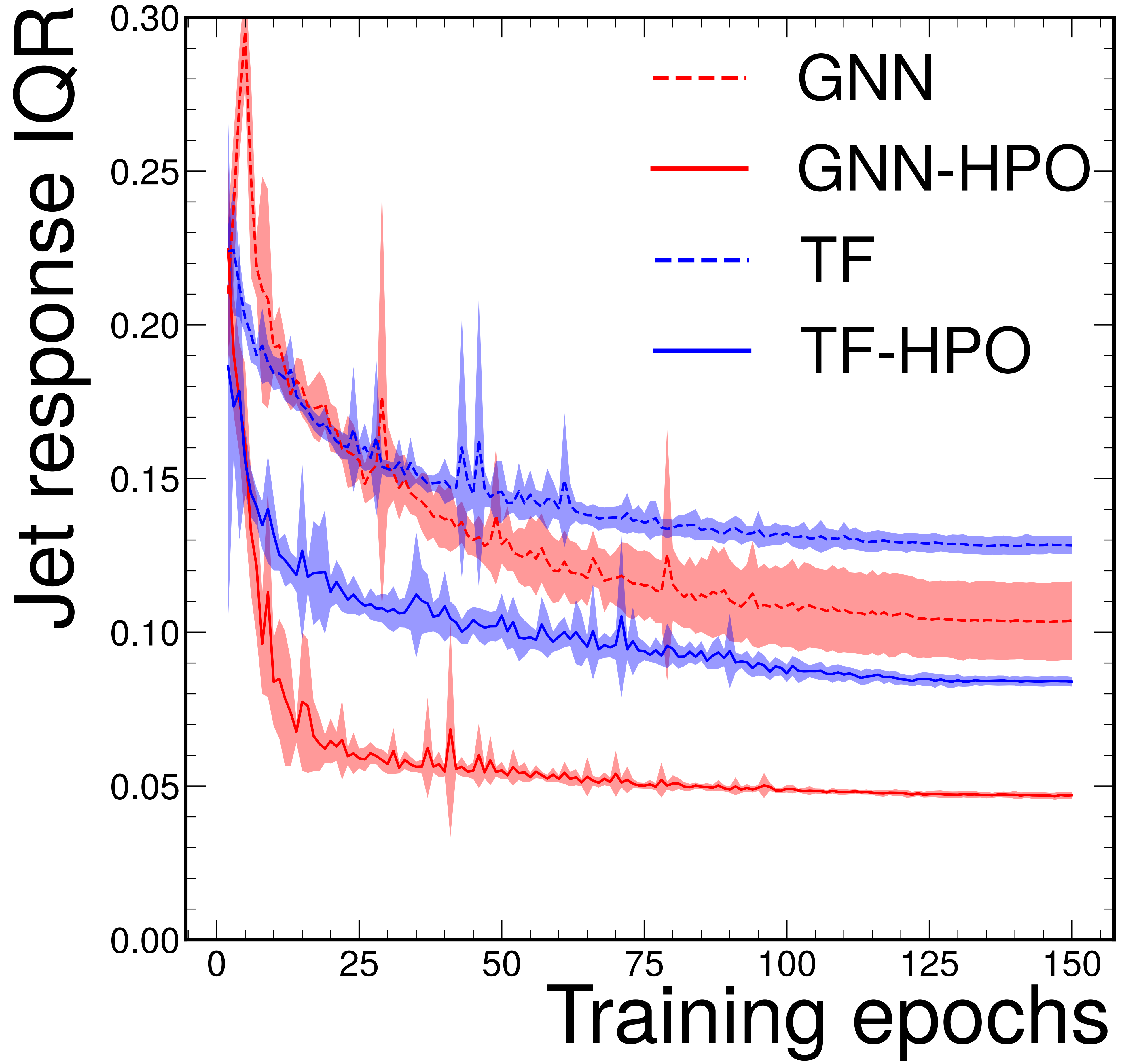
- Many hyperparameters to tune, e.g. number of layers, hidden dimension of each layer, and LSH bin size
- Requires large compute

JURECA Supercomputer at Jülich Supercomputing Centre

IMPACT OF TUNING

- Tuning improves particle-level performance dramatically (trained on $q\bar{q}, t\bar{t}$)
- Though we optimize a particle-level loss, also achieve better jet/MET resolution

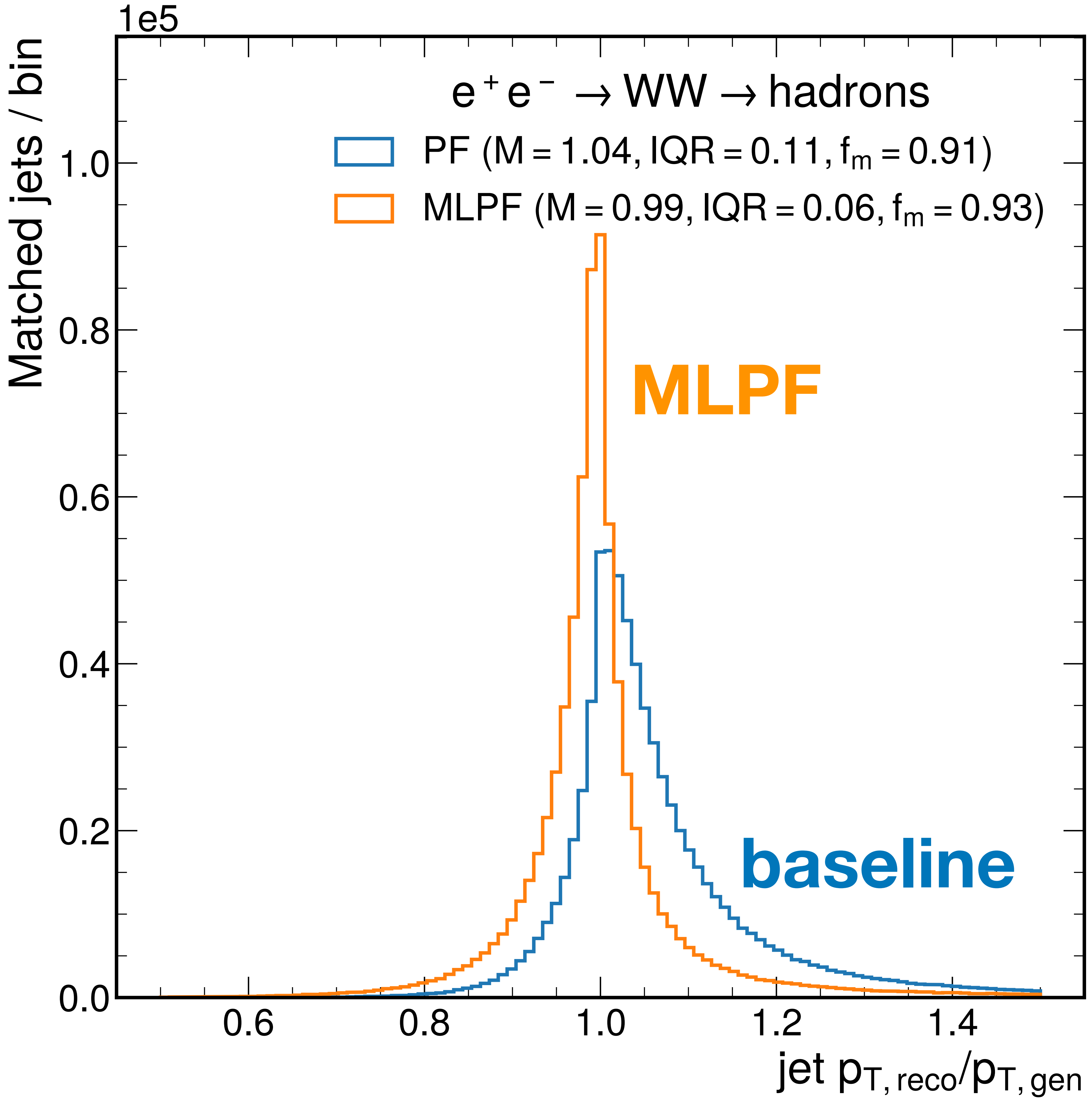


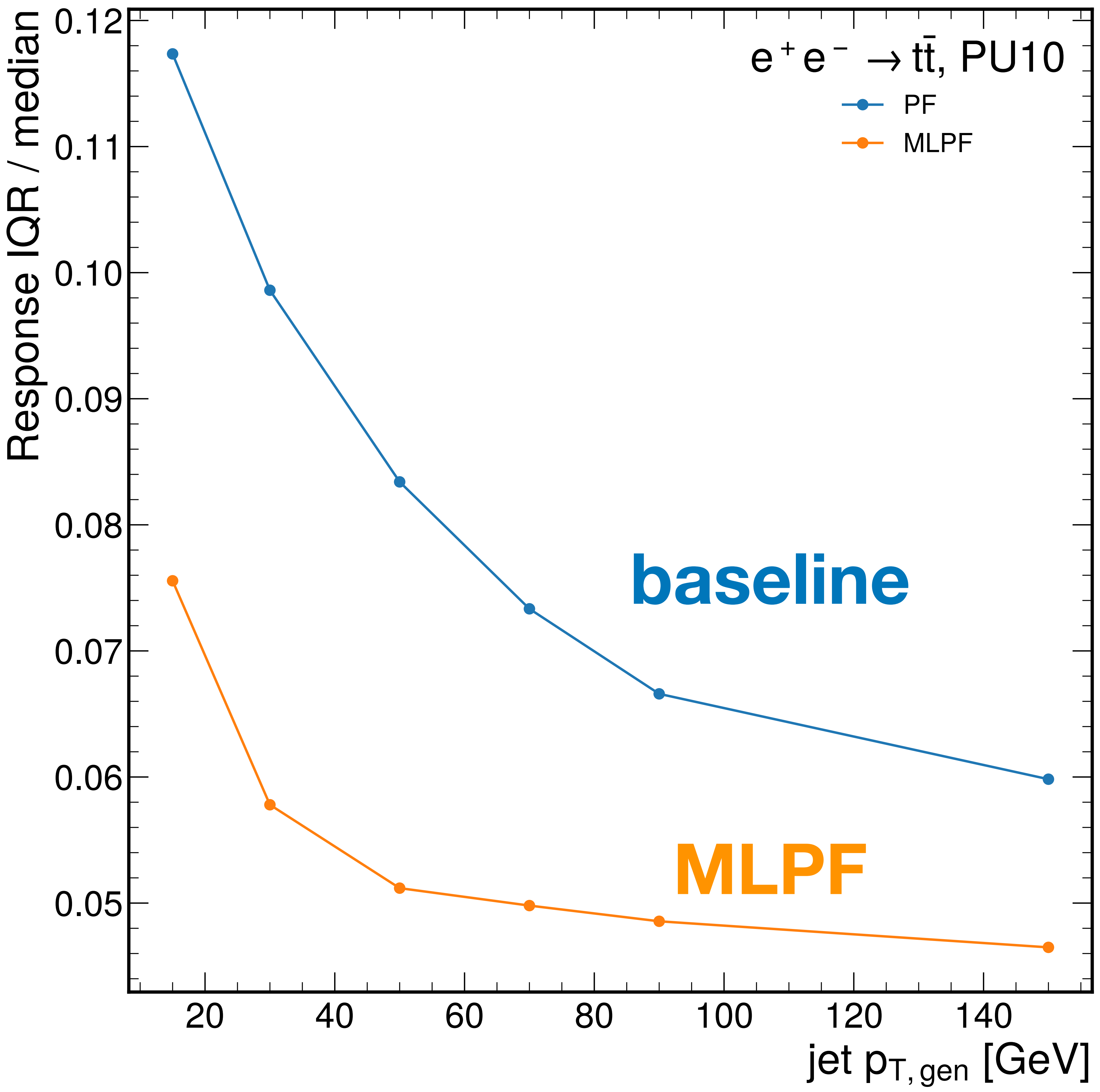


PERFORMANCE AND GENERALIZATION

- Generalizes to samples (e.g., $e^+e^- \rightarrow WW \rightarrow \mathrm{hadrons}$) never used in training
- ~50% improvement in jet response width over the baseline*

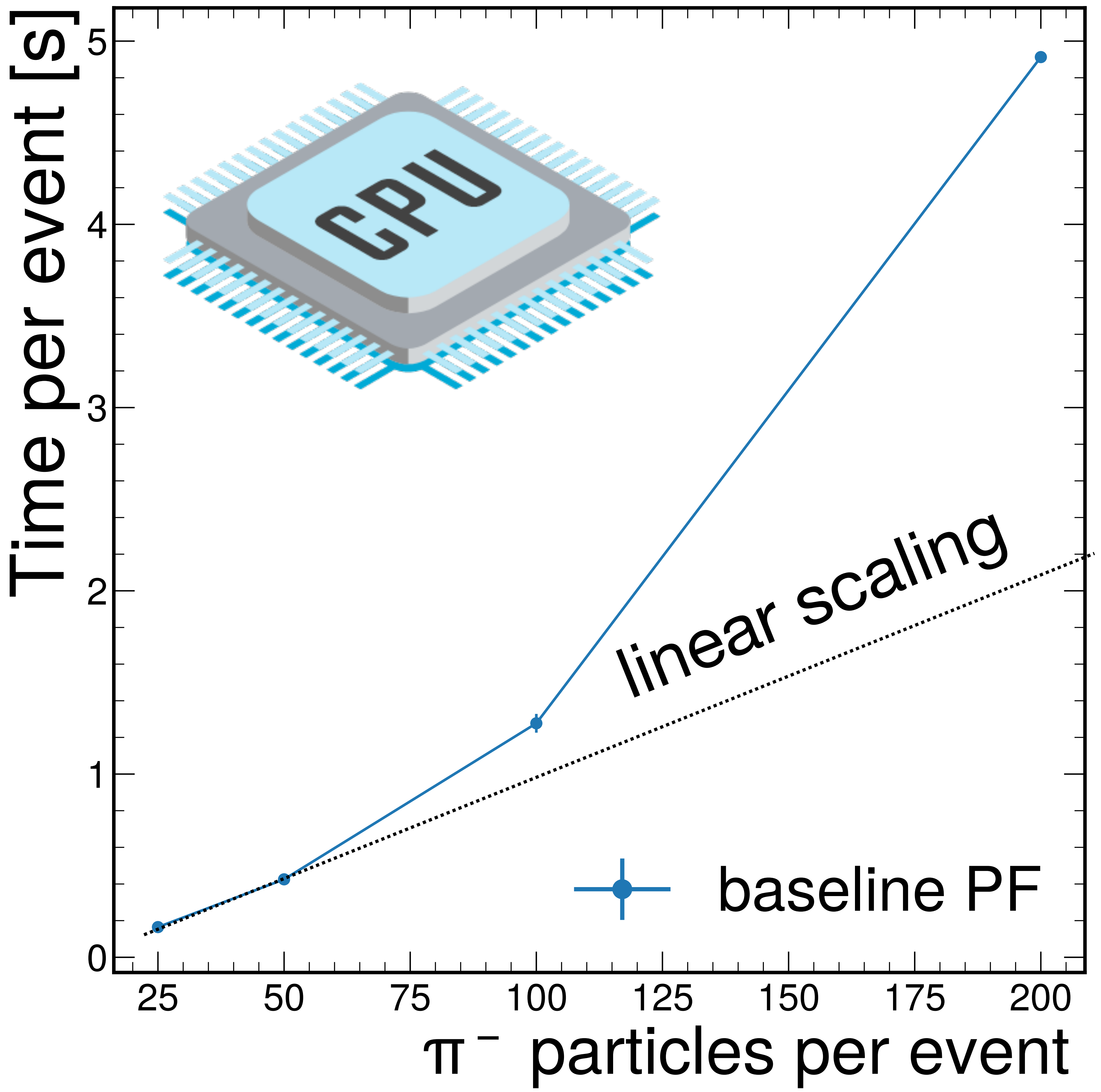
*Defined with gen. particle status = 1

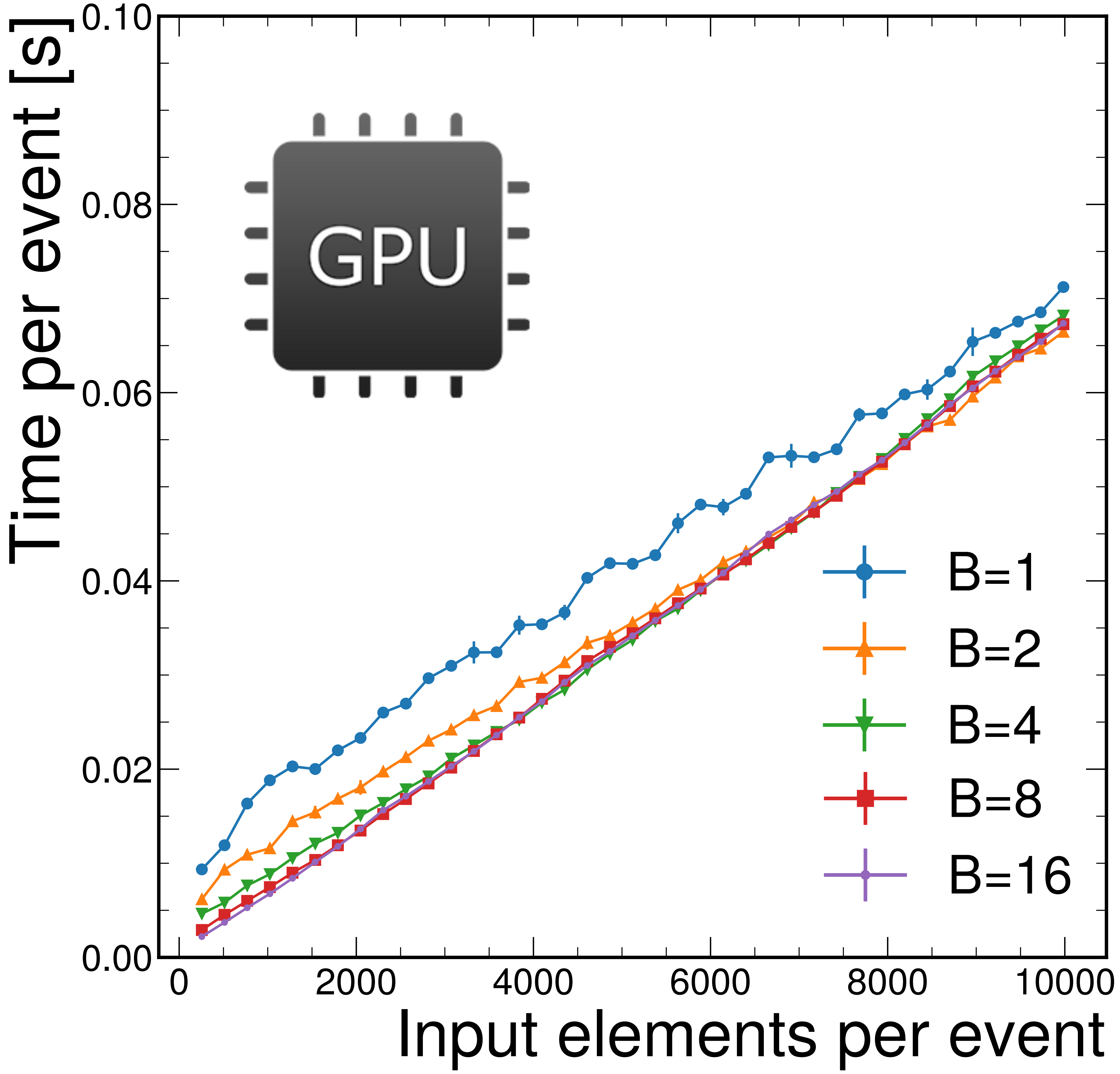




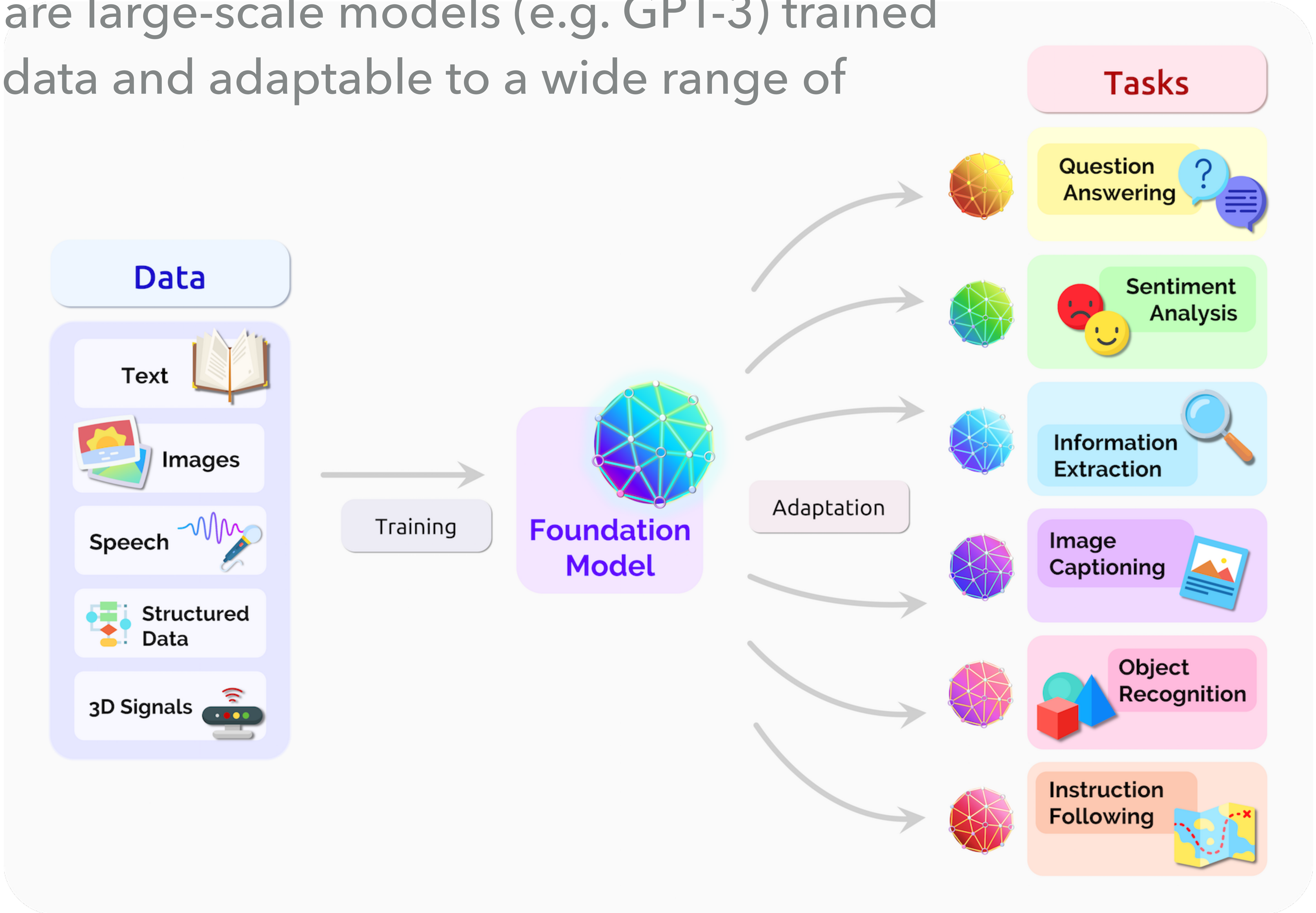
SCALING

- 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 8 GB GPU





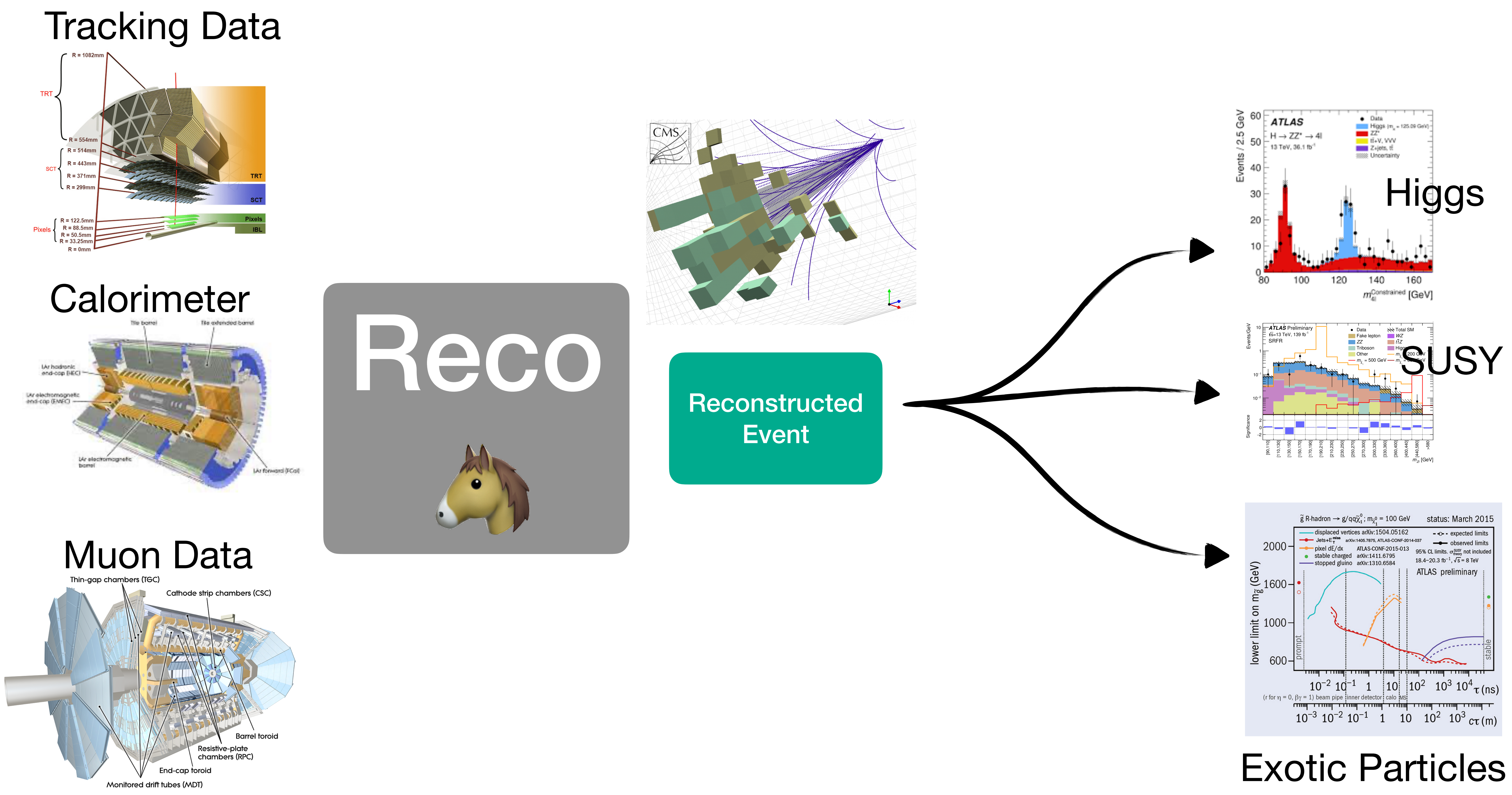
- "Foundation models" are large-scale models (e.g. GPT-3) trained on broad multimodal data and adaptable to a wide range of downstream tasks



FOUNDATION MODELS IN HEP

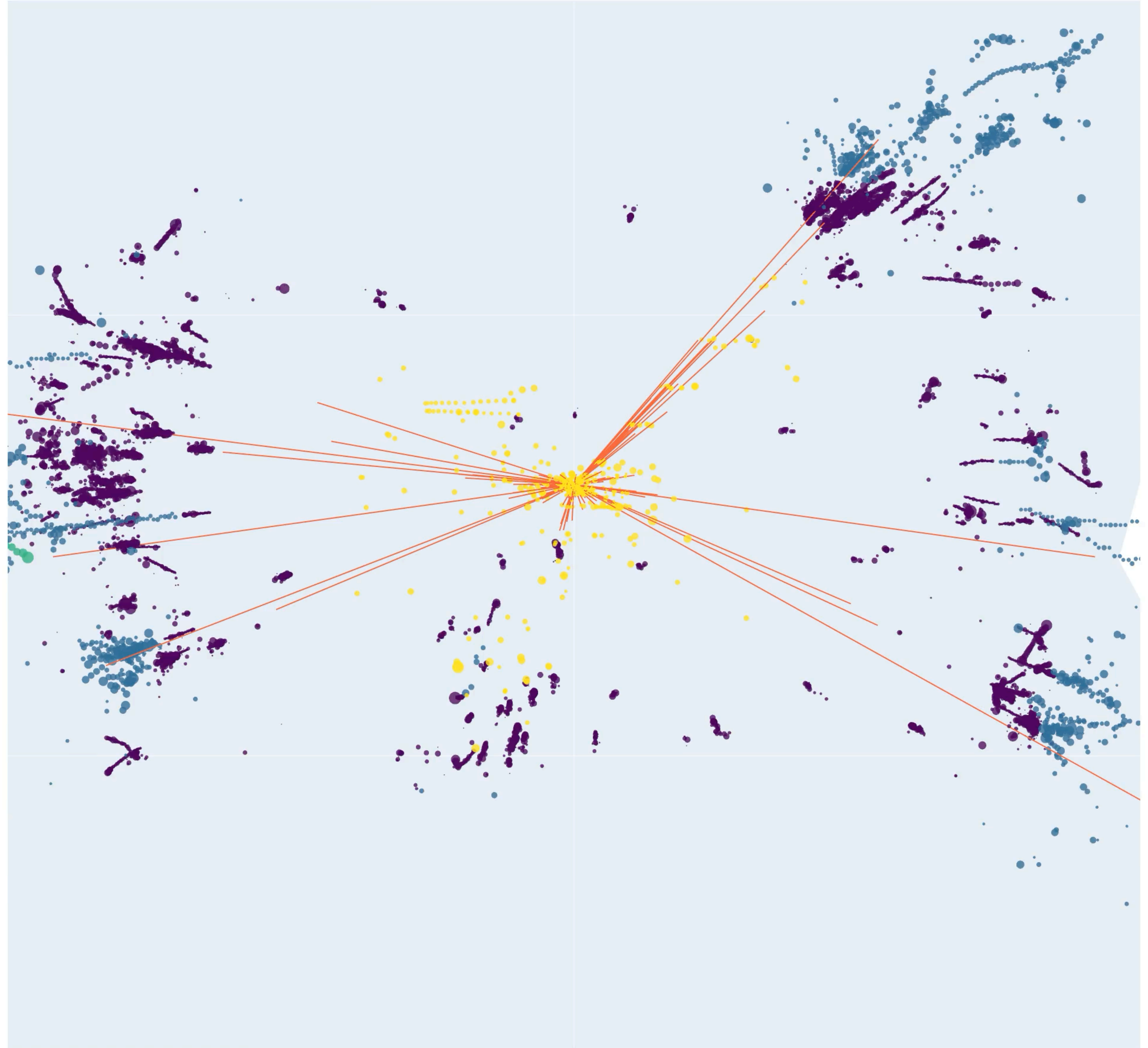
L. Heinrich, https://indico.cern.ch/event/1330797/contributions/5776144/

- Reconstruction in HEP is analogous to a foundation model
- With ML-based reconstruction, can take this analogy more literally and fine-tune reconstruction for different needs, e.g. analysis or new detector concepts



SUMMARY

- ML-based event reconstruction improves physics performance at future colliders
- End-to-end optimization can enable **new paradigms**, e.g. fine-tuning ML-based reconstruction for different use cases (analysis, detector concepts, etc.)
- Scalable ML models improve computational performance
- Open datasets and code accelerate research



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;

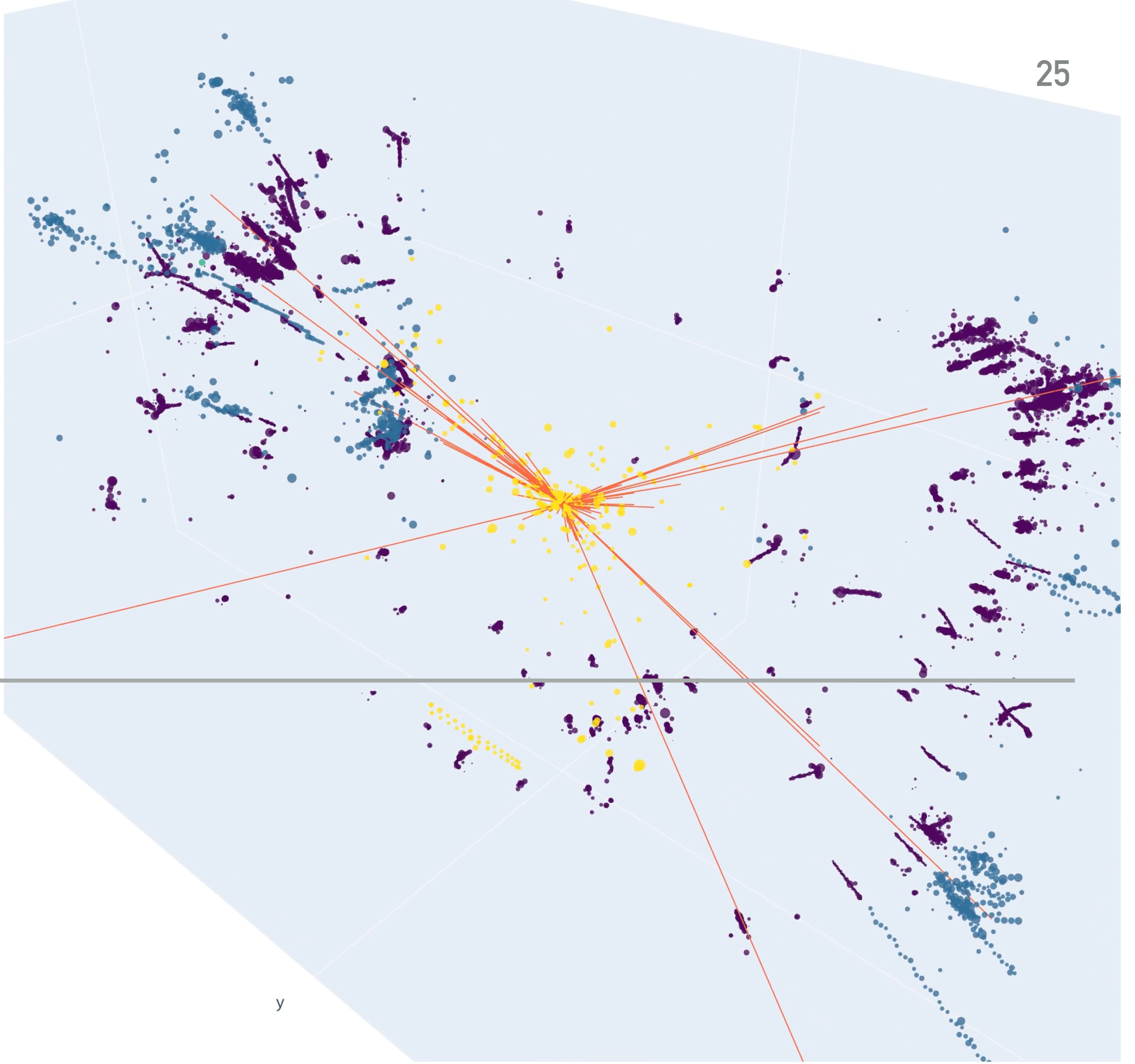
JAVIER DUARTE

FUTURE OF HIGH ENERGY PHYSICS

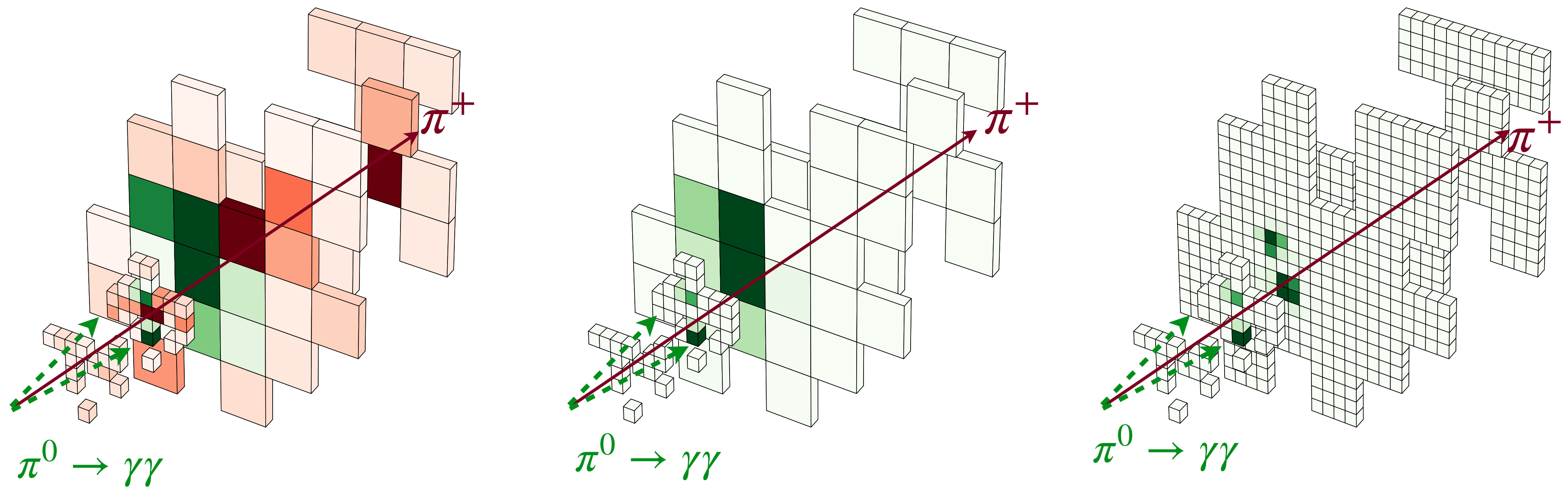
ASPEN CENTER FOR PHYSICS

MARCH 29, 2024

BACKUP



RELATED WORK

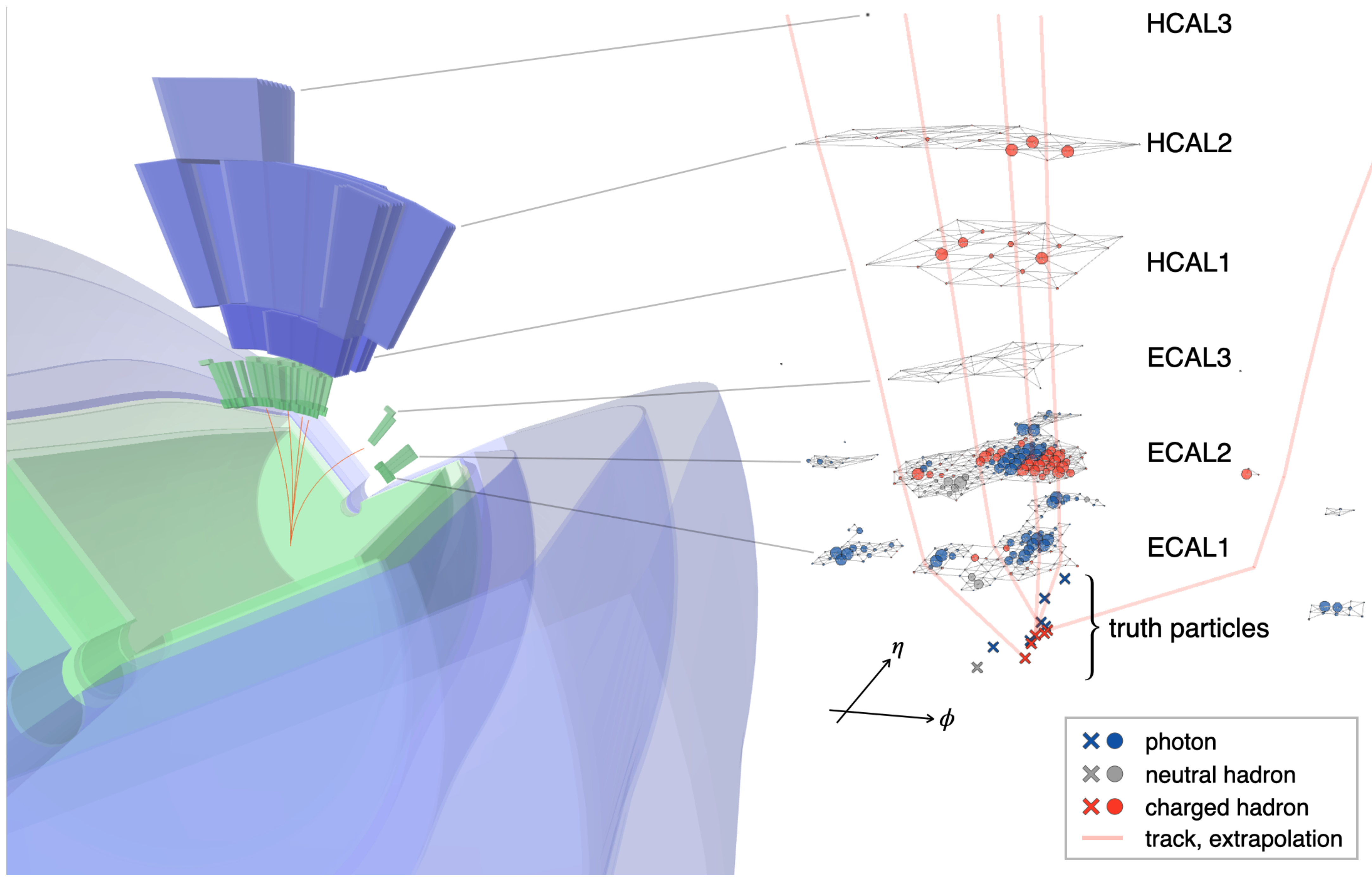


Towards a Computer Vision Particle Flow ⋆

Francesco Armando Di Bello[a,3], Sanmay Ganguly[b,1], Eilam Gross[1], Marumi Kado[3,4], Michael Pitt[2], Lorenzo Santi [3], Jonathan Shlomi[1]

[1]Weizmann Institute of Science, Rehovot 76100, Israel
[2]CERN, CH 1211, Geneva 23, Switzerland
[3]Università di Roma Sapienza, Piazza Aldo Moro, 2, 00185 Roma, Italy e INFN, Italy
[4]Université Paris-Saclay, CNRS/IN2P3, IJCLab, 91405, Orsay, France

arXiv:2003.08863



Reconstructing particles in jets using set transformer and hypergraph prediction networks

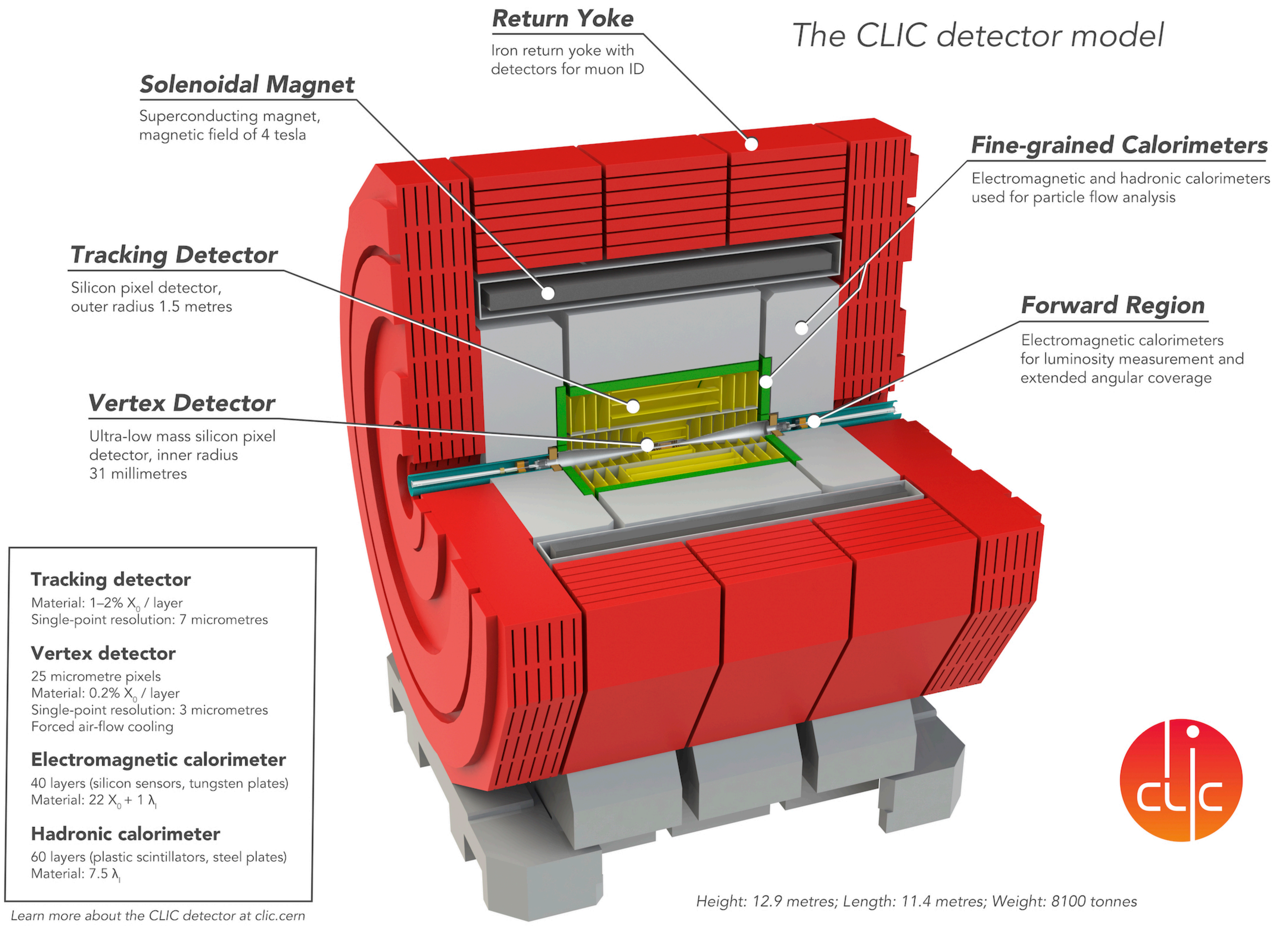
Francesco Armando Di Bello [1,a], Etienne Dreyer [2,b], Sanmay Ganguly [3], Eilam Gross [2], Lukas Heinrich [4], Anna Ivina [2], Marumi Kado [5,6], Nilotpal Kakati [2,c], Lorenzo Santi [6], Jonathan Shlomi [2], Matteo Tusoni [6]

[1] INFN and University of Genova
[2]Weizmann Institute of Science
[3]ICEPP, University of Tokyo
[4]Technical University of Munich
[5]Max Planck Institute for Physics
[6]INFN and Sapienza University of Rome

arXiv:2212.01328

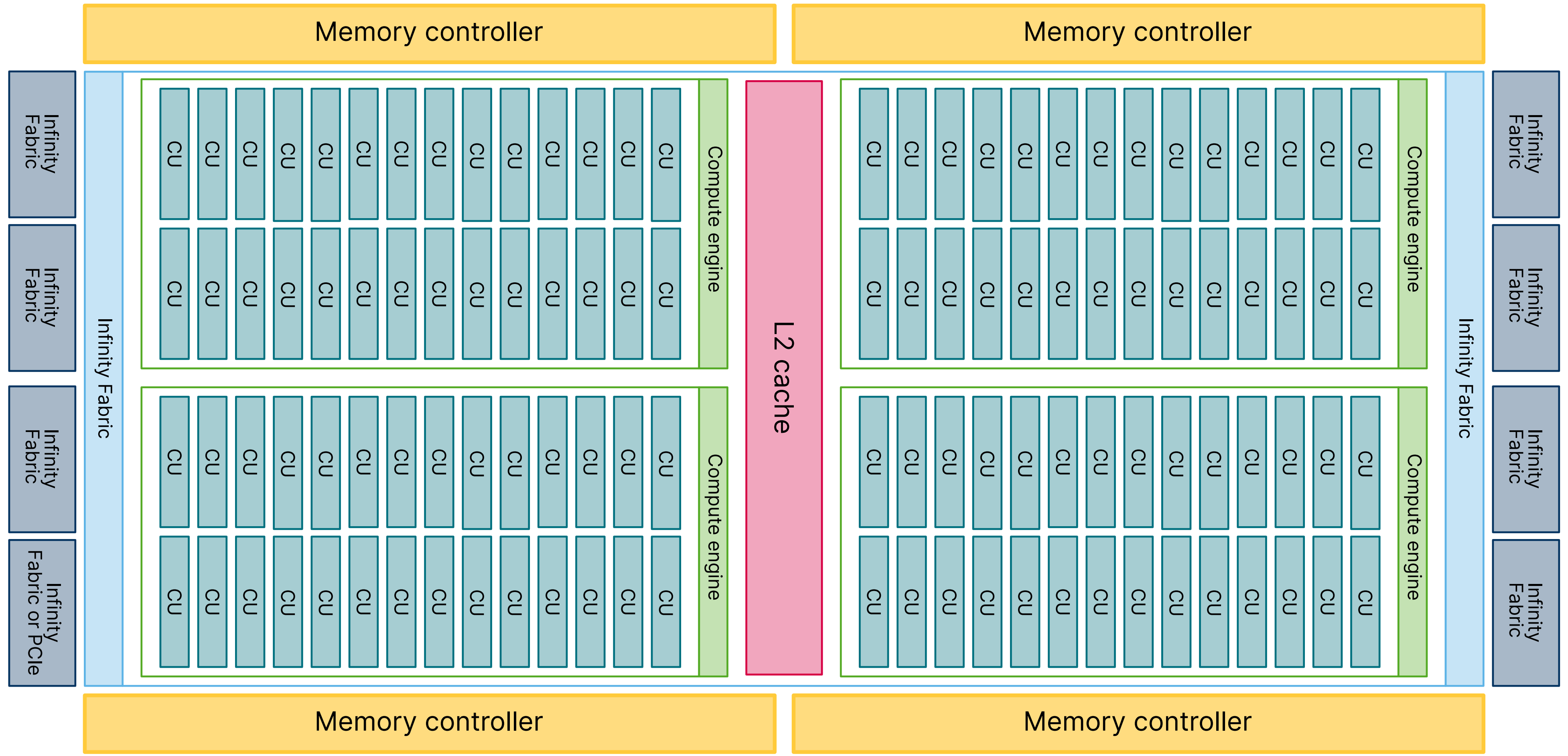
CLIC DETECTOR MODEL

- CLIC detector (CLIC_o3_v14)

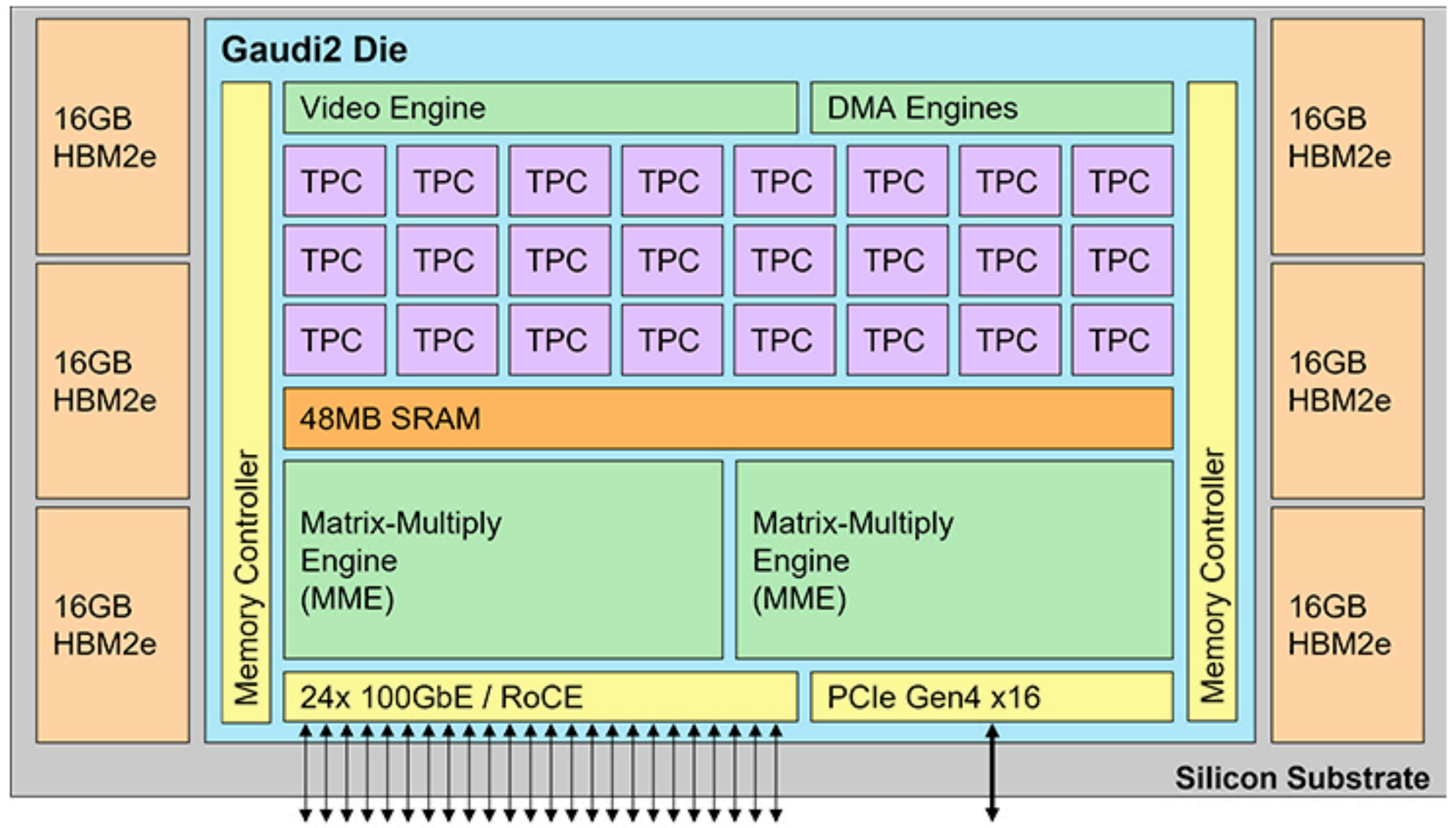


HPC AI CHIPS

The HPC AI chip landscape is diversifying



AMD MI250X GPU

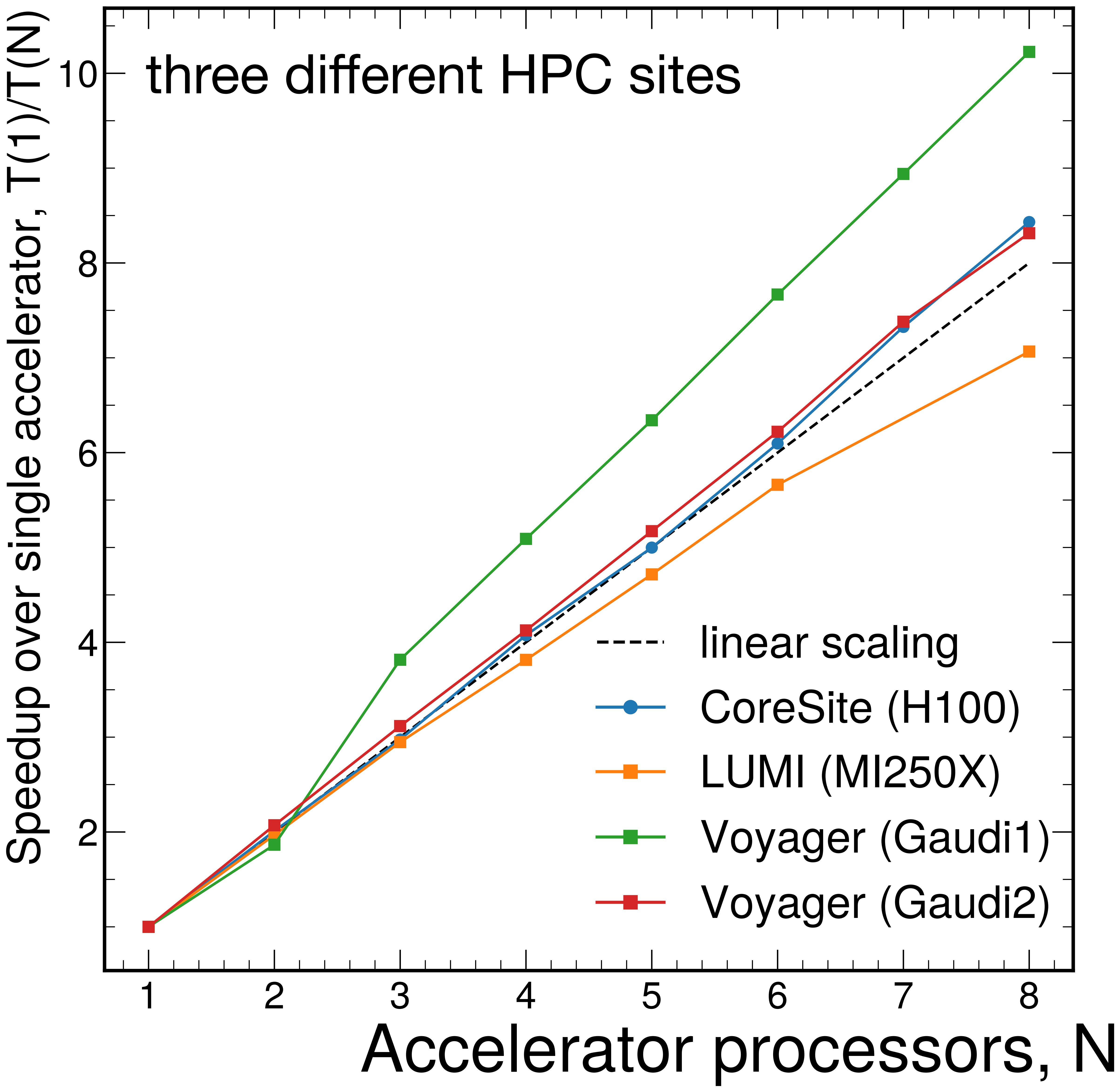


Intel Gaudi2 deep learning processor

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

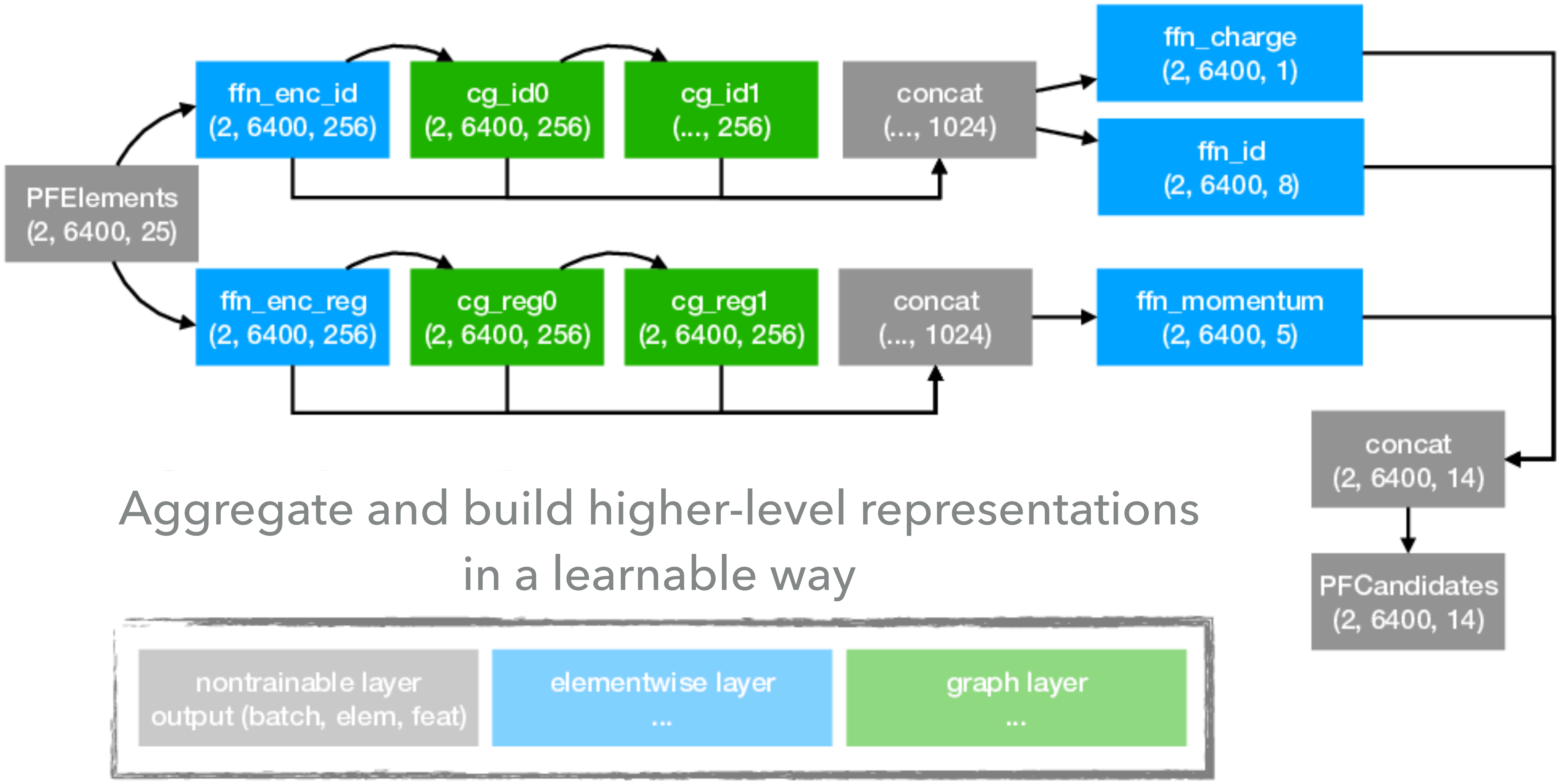
PORTABILITY

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



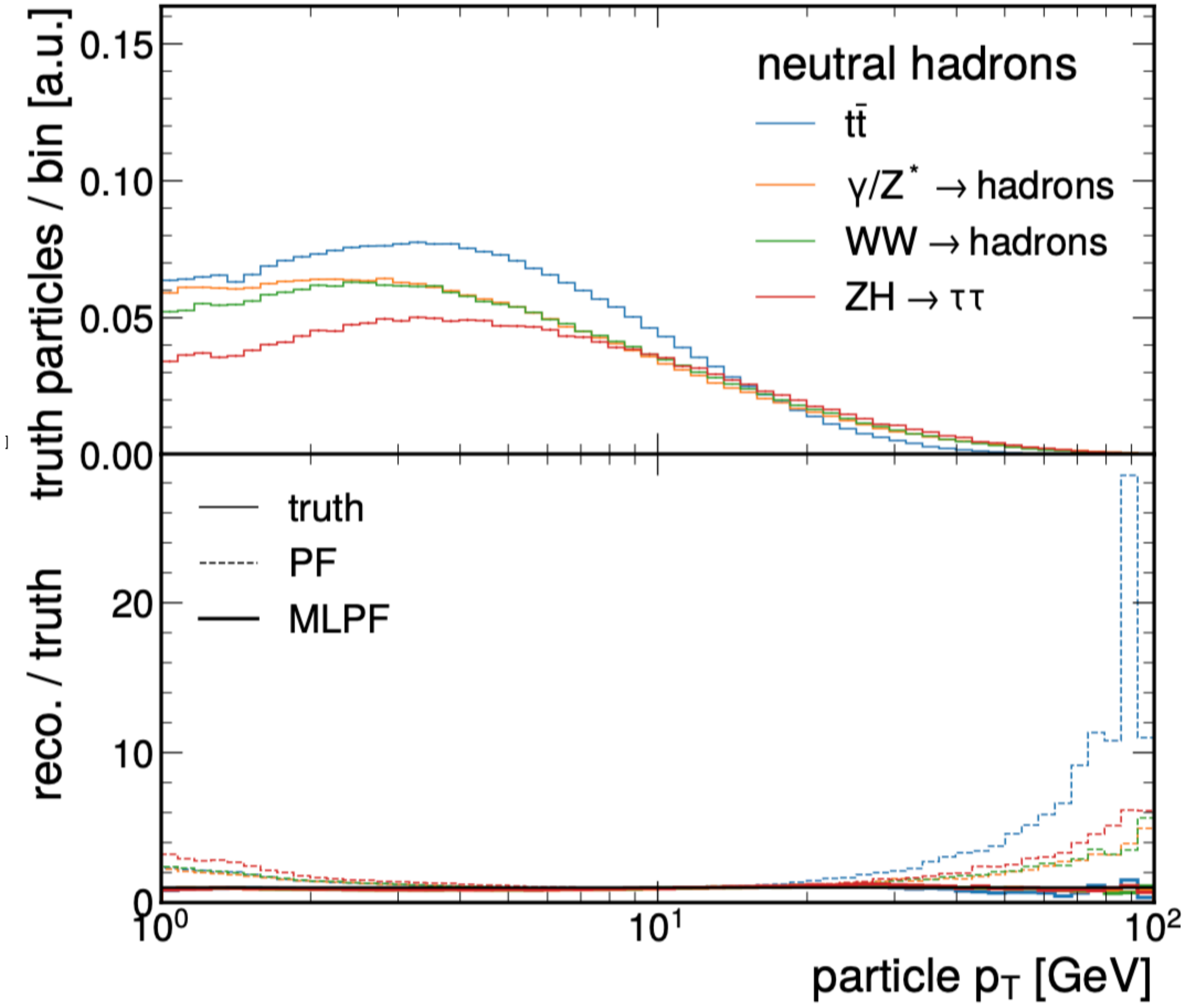
STACKING GNN-LSH

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



BULK AND TAILS

- Datasets are diverse so we have to predict the bulk and the tails well for all particle types



PERFORMER-BASED TRANSFORMER

- Alternative: scalable transformer based on the Performer architecture [arXiv:2009.14794]

One layer of kernel-based self attention with the FAVOR mechanism.