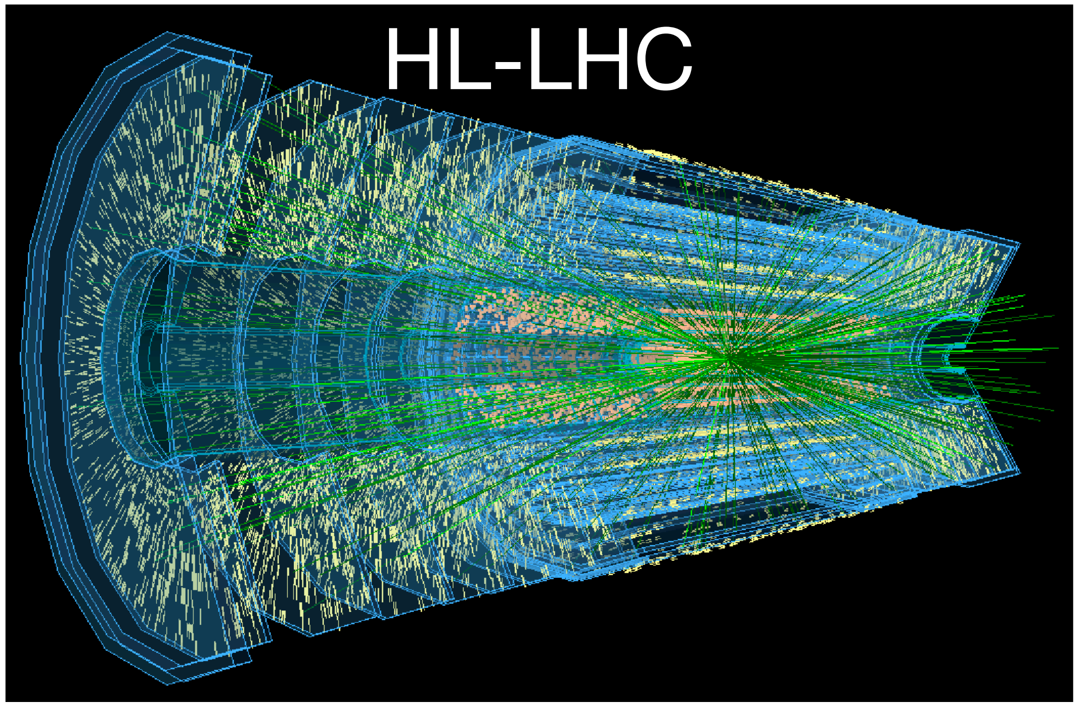
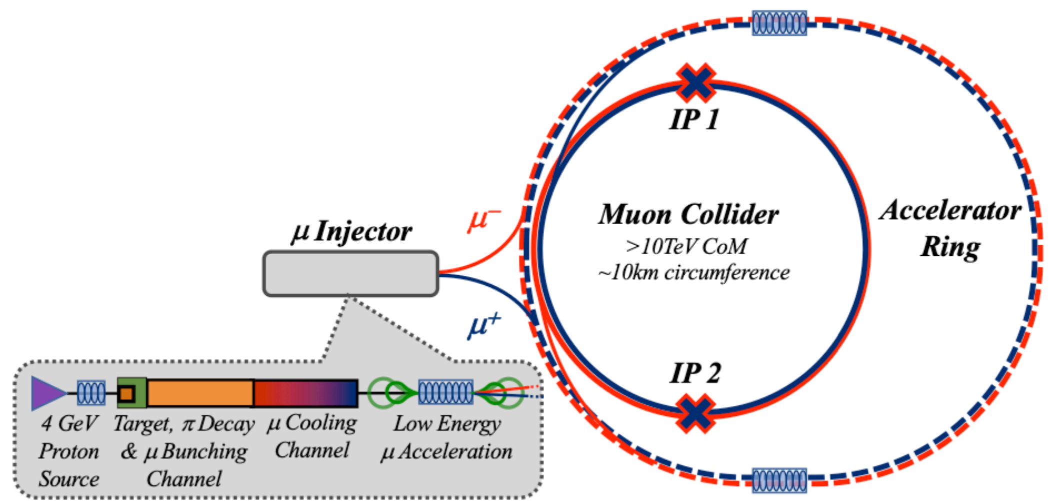
ML & computing

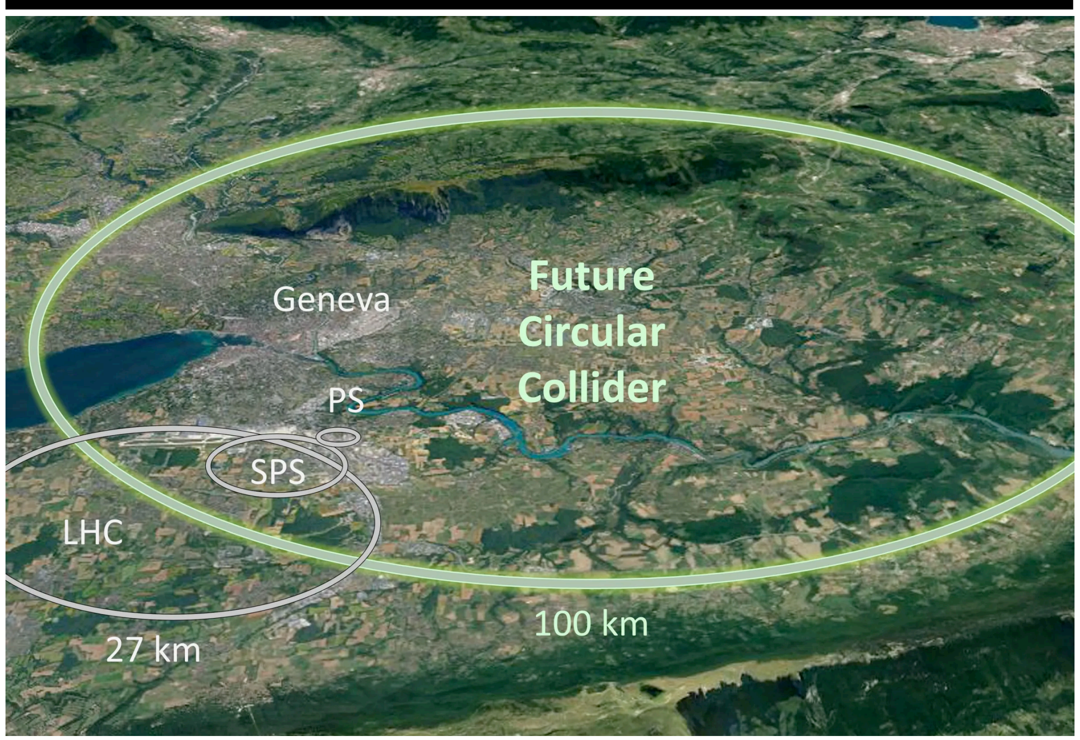
Joosep Pata (KBFI)
October 20, 2024

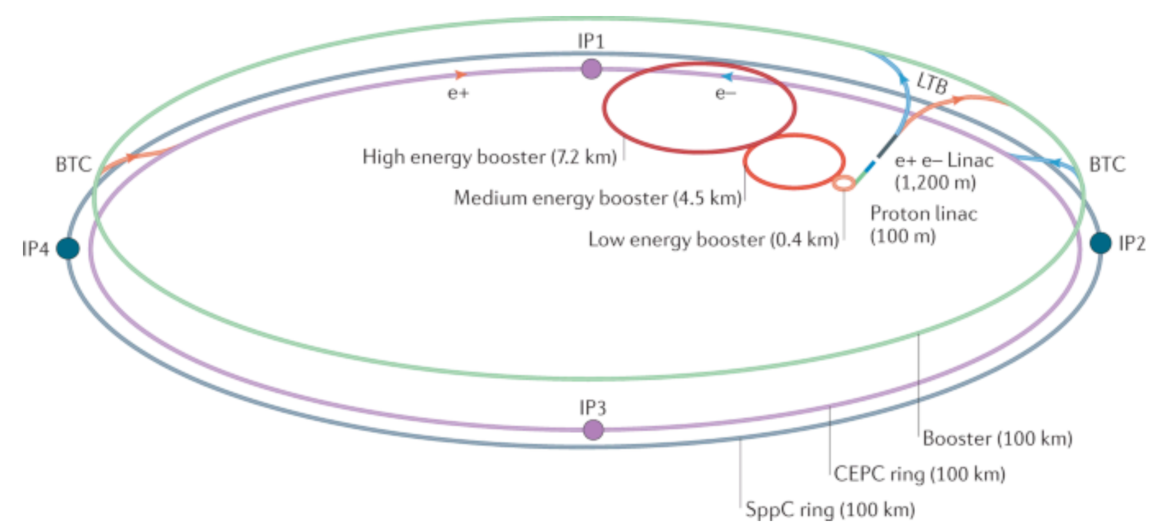
Disclaimer: Due to limited time, not a general overview of everything.
Only a small snapshot of many exciting activities… Personal views.

Different machines under consideration for the future of HEP



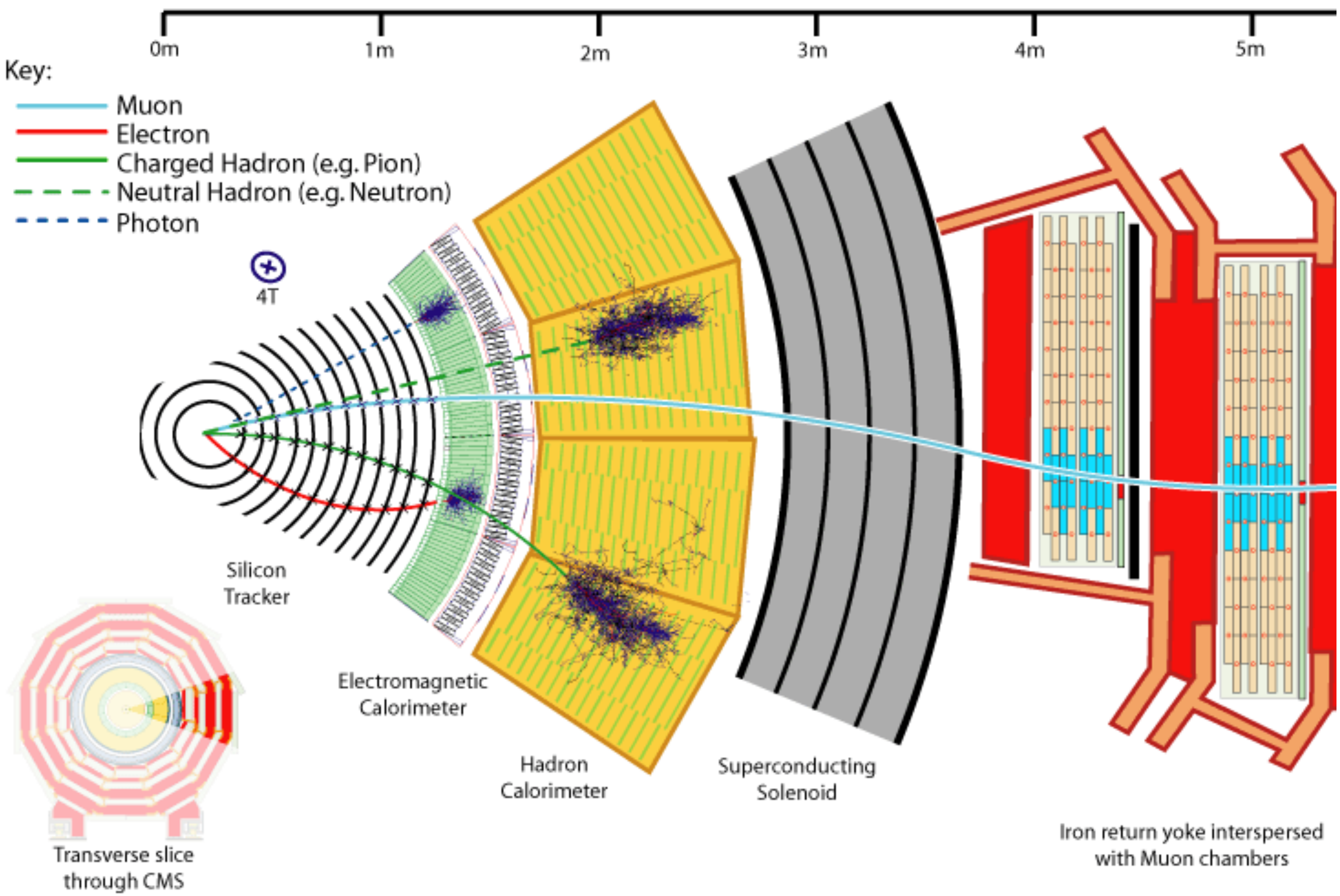






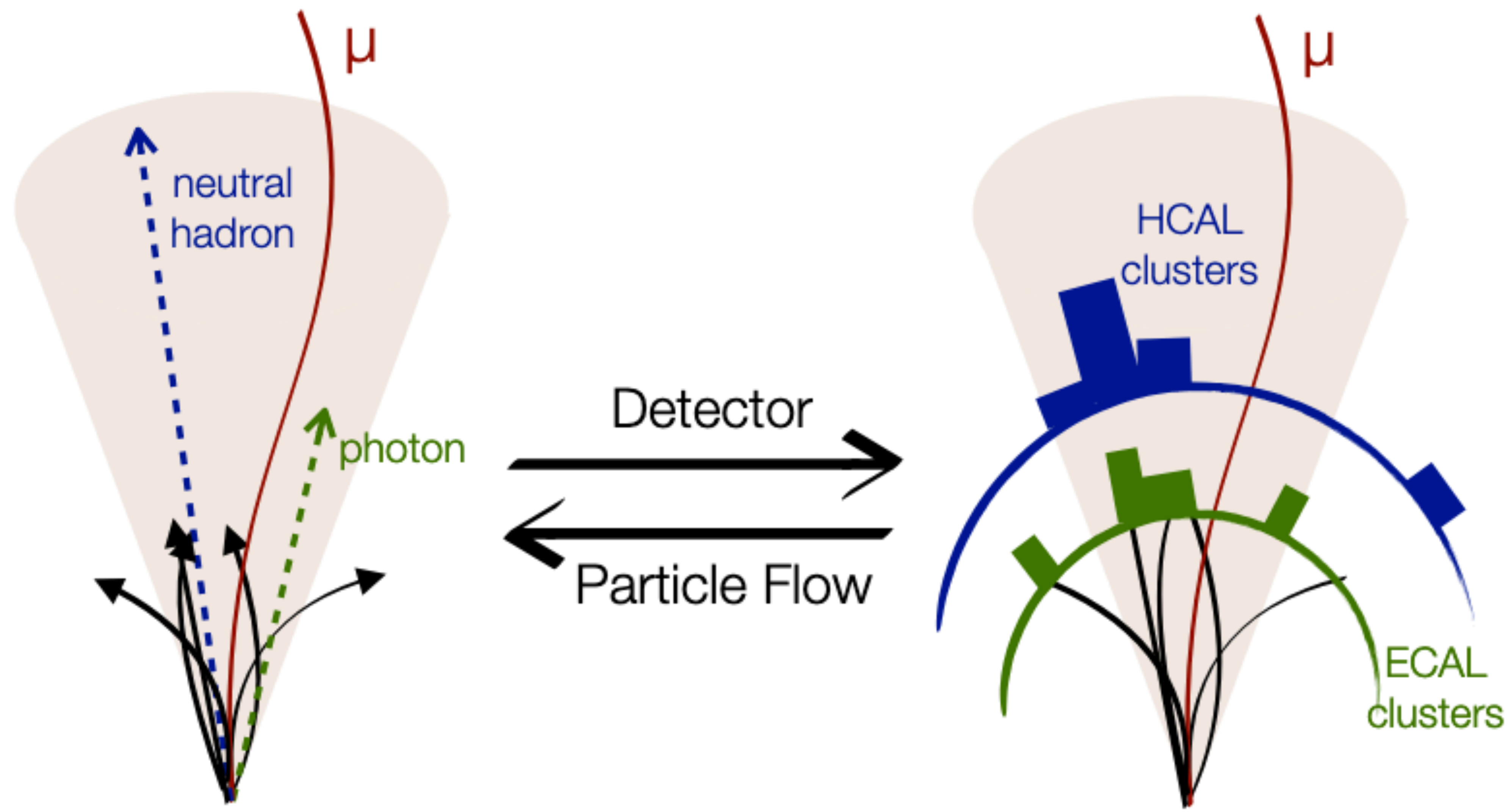
Efficient, realistic sim and reco needed before startup!

Current and future multilayered detectors…

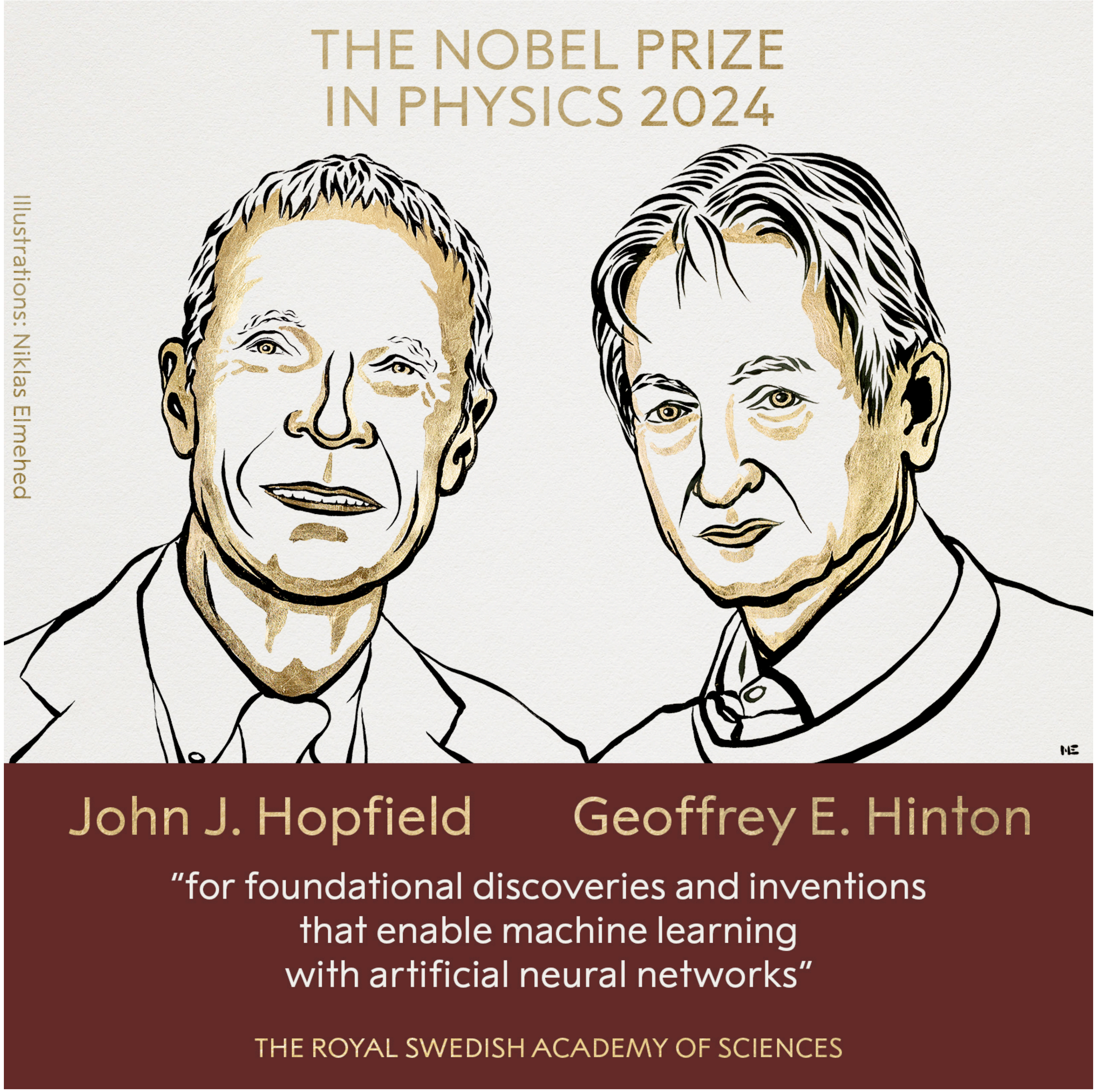


…need complex data reconstruction → particle flow algorithm

- Particles interact with detector, leave hits, tracks, clusters
- Complex, hand-tuned algorithms based on heuristics to reconstruct particles from multimodal detectors

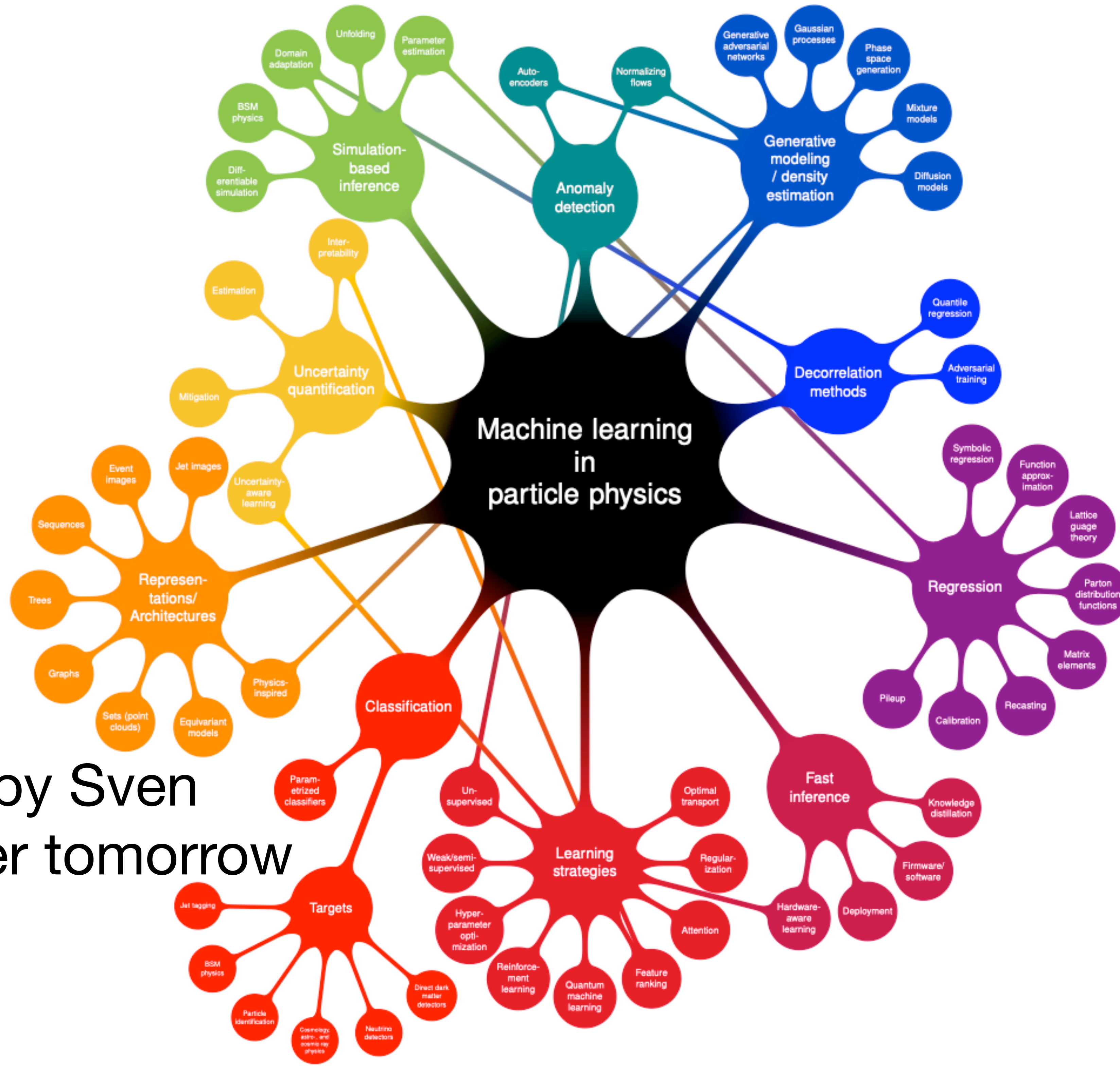


Fundamental physics contributed to modern machine learning…



… and modern machine learning is informing fundamental physics

Machine learning has a wide variety of applications in science

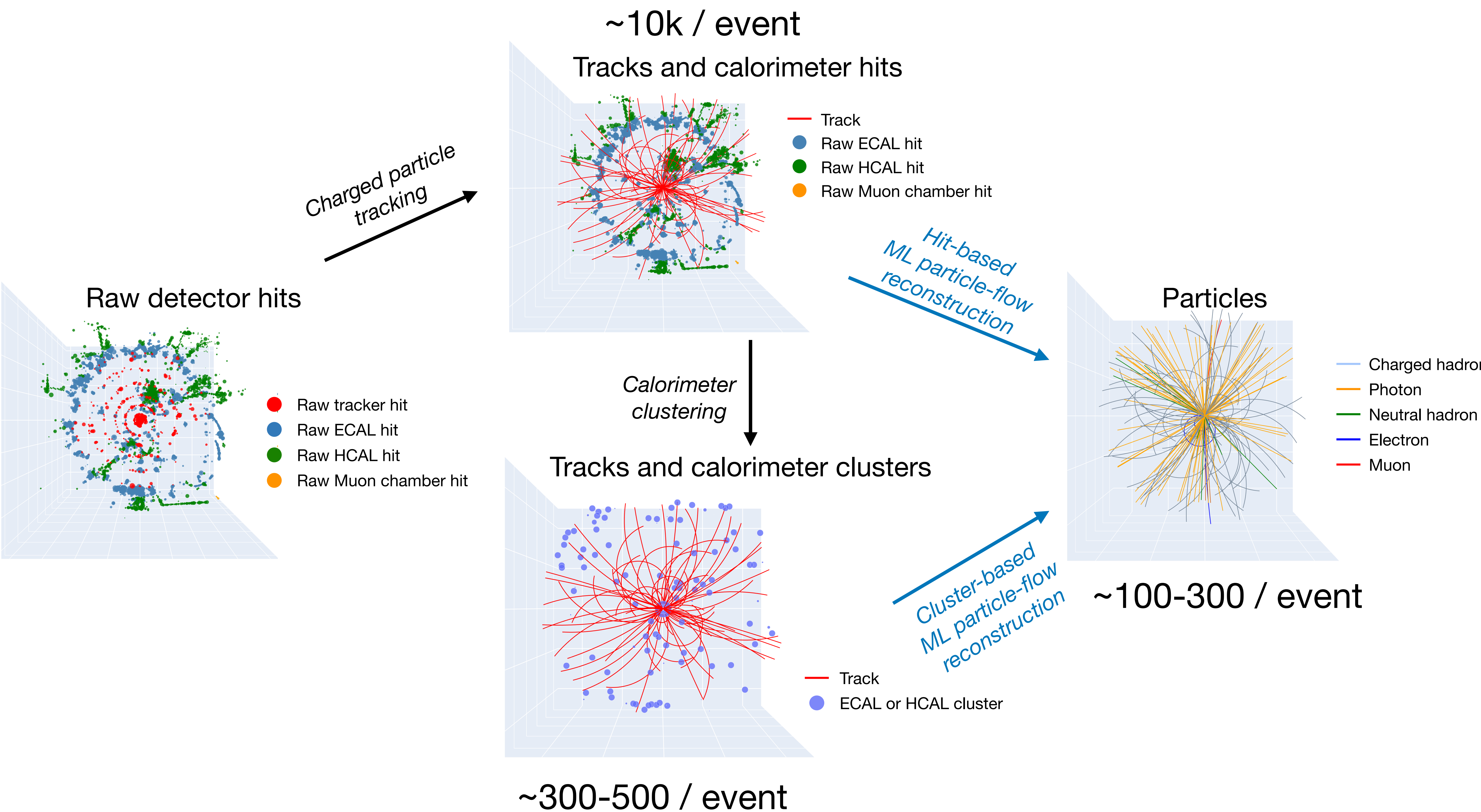


Talk by Laurits Tani today.

Talk by Sven Põder tomorrow

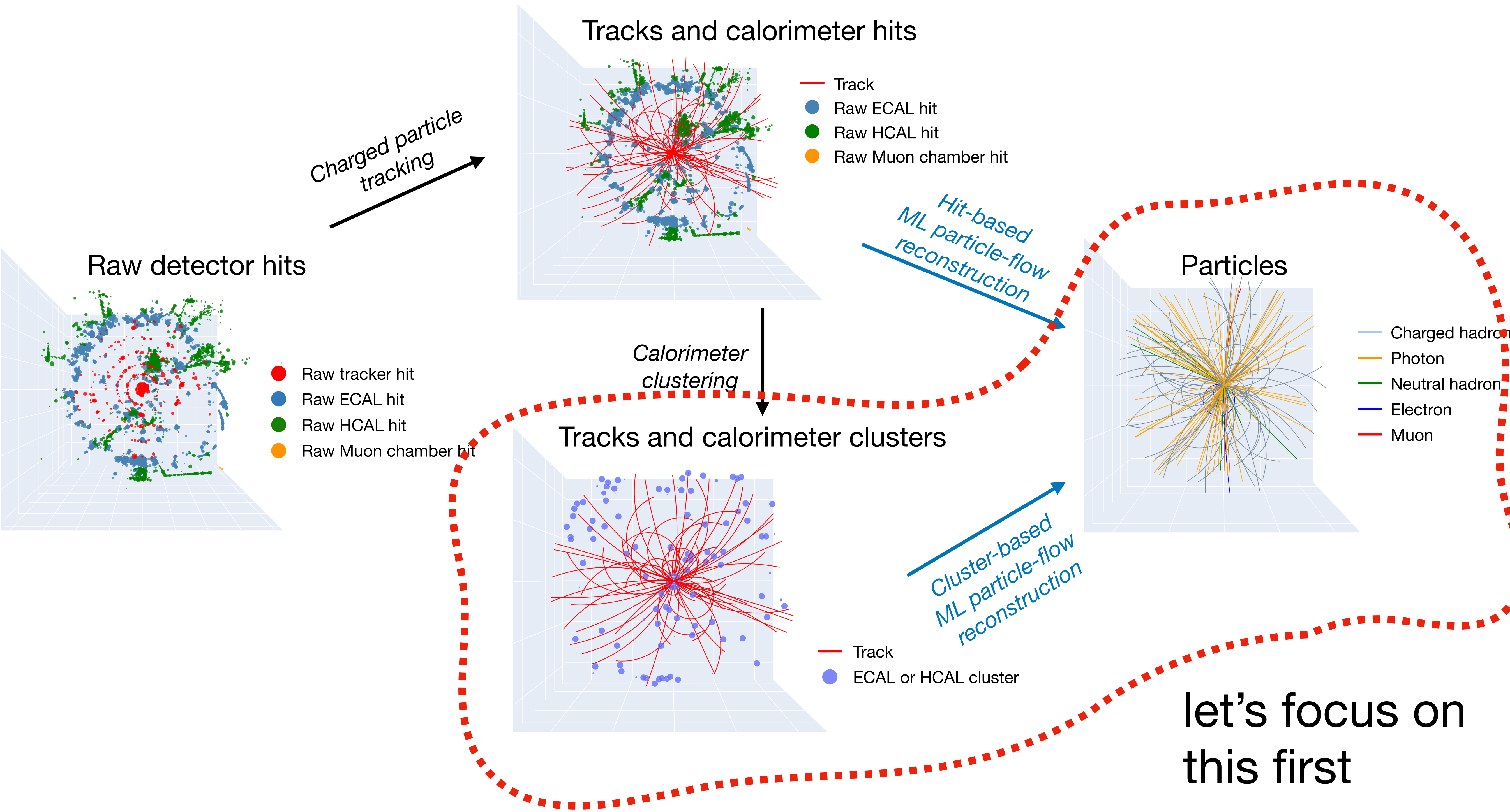
https://doi.org/10.5281/zenodo.7335953, credit J. Duarte

We have created a new **open dataset** with full simulation.

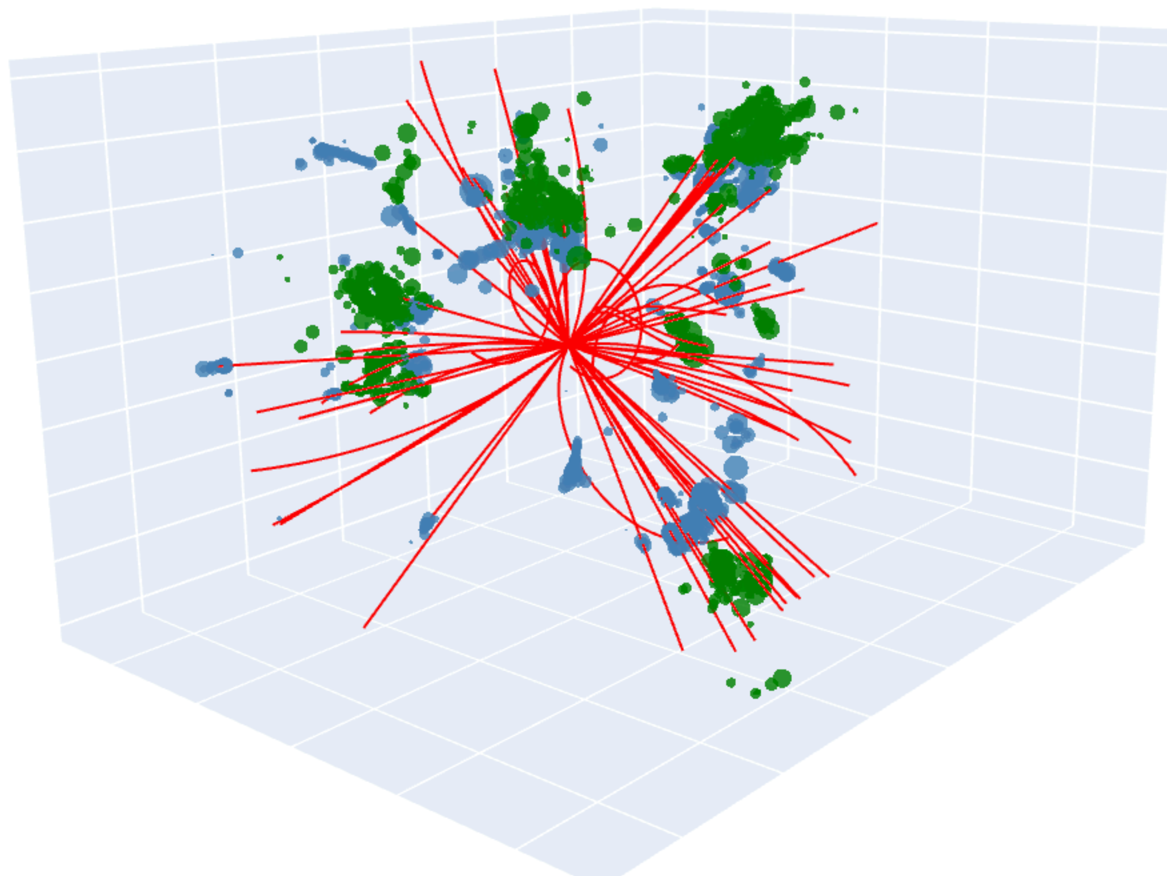


calo/tracker hits, tracks and calo clusters, baseline & target particles

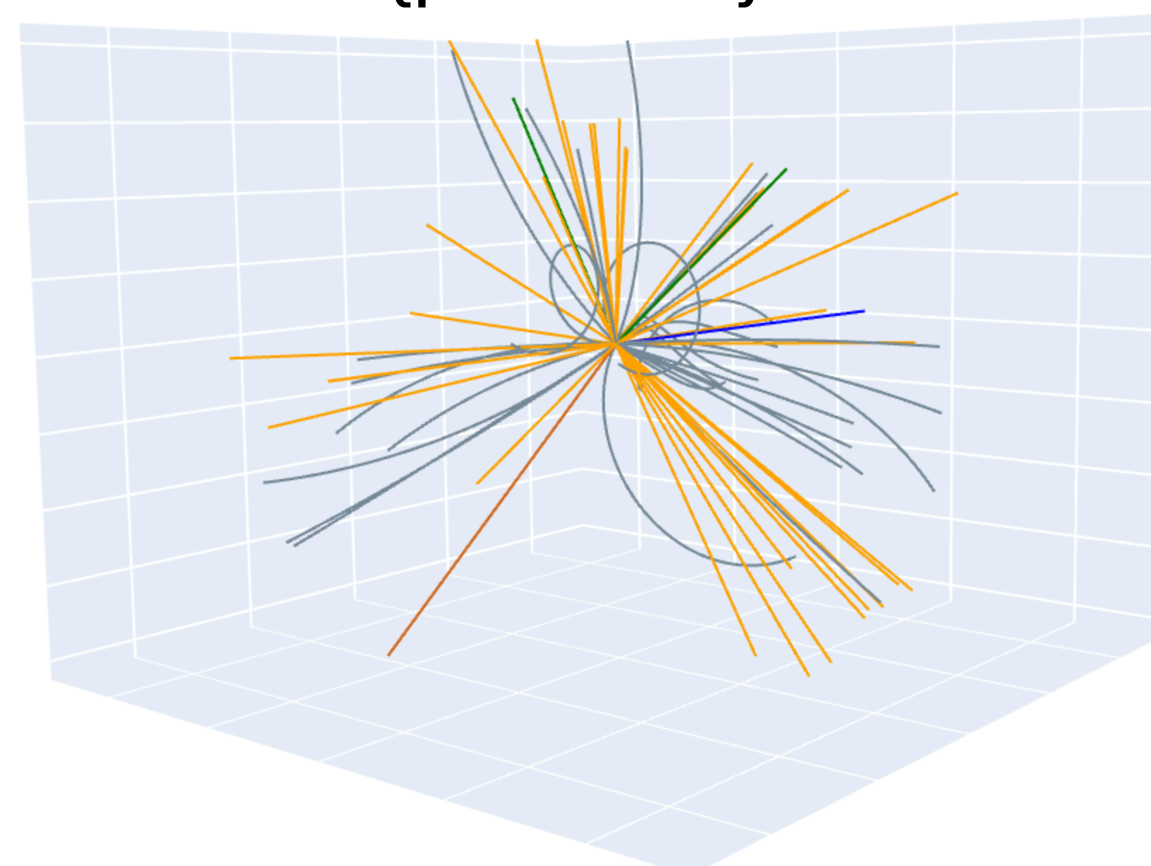
MLPF(tracks and clusters) → particles



{tracks, clusters} ∈ X



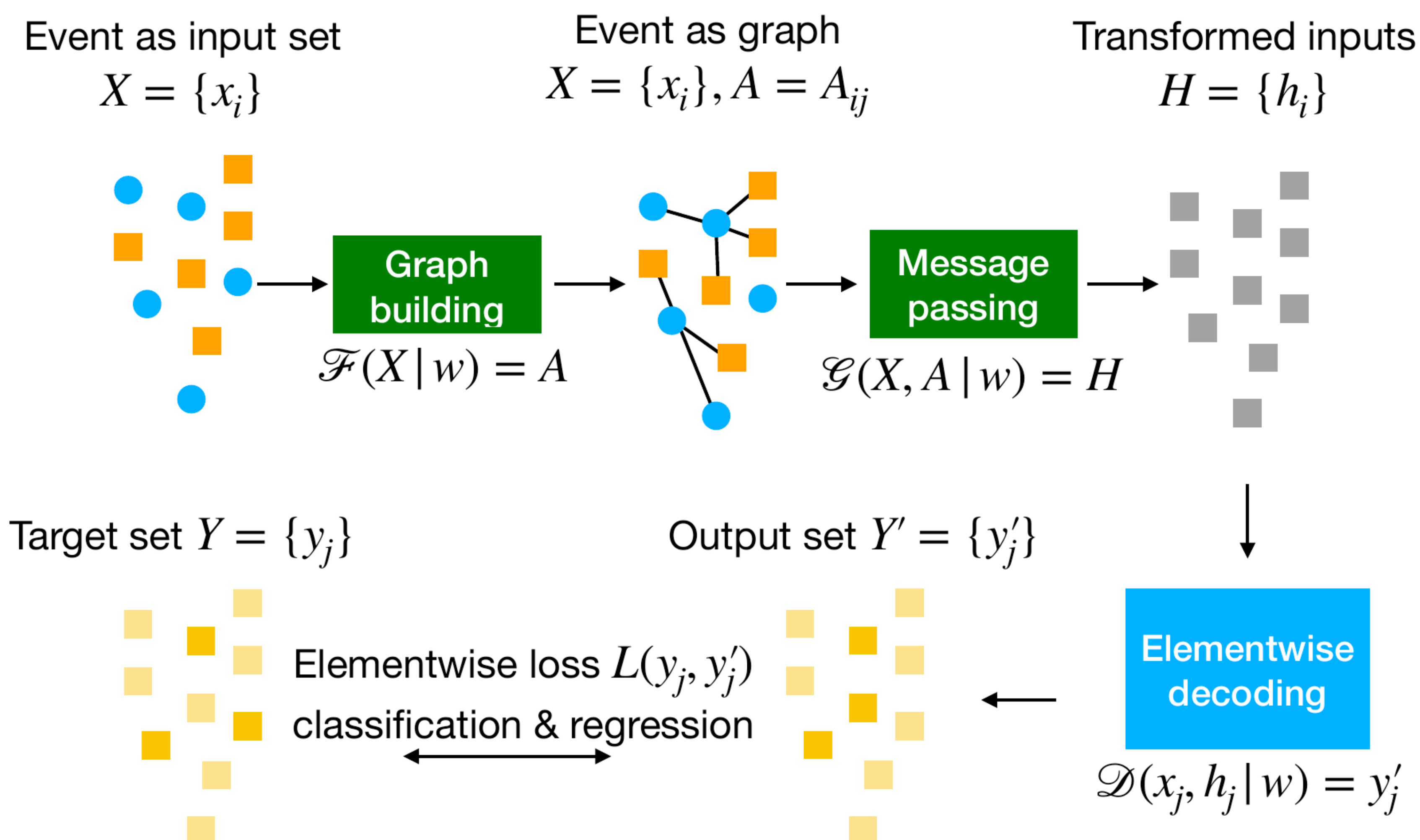
{particles} ∈ Y



neural_network(X) → Y

We can use ML for full event reconstruction!

graph structure learning + message passing
supervised multi-task output



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

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

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

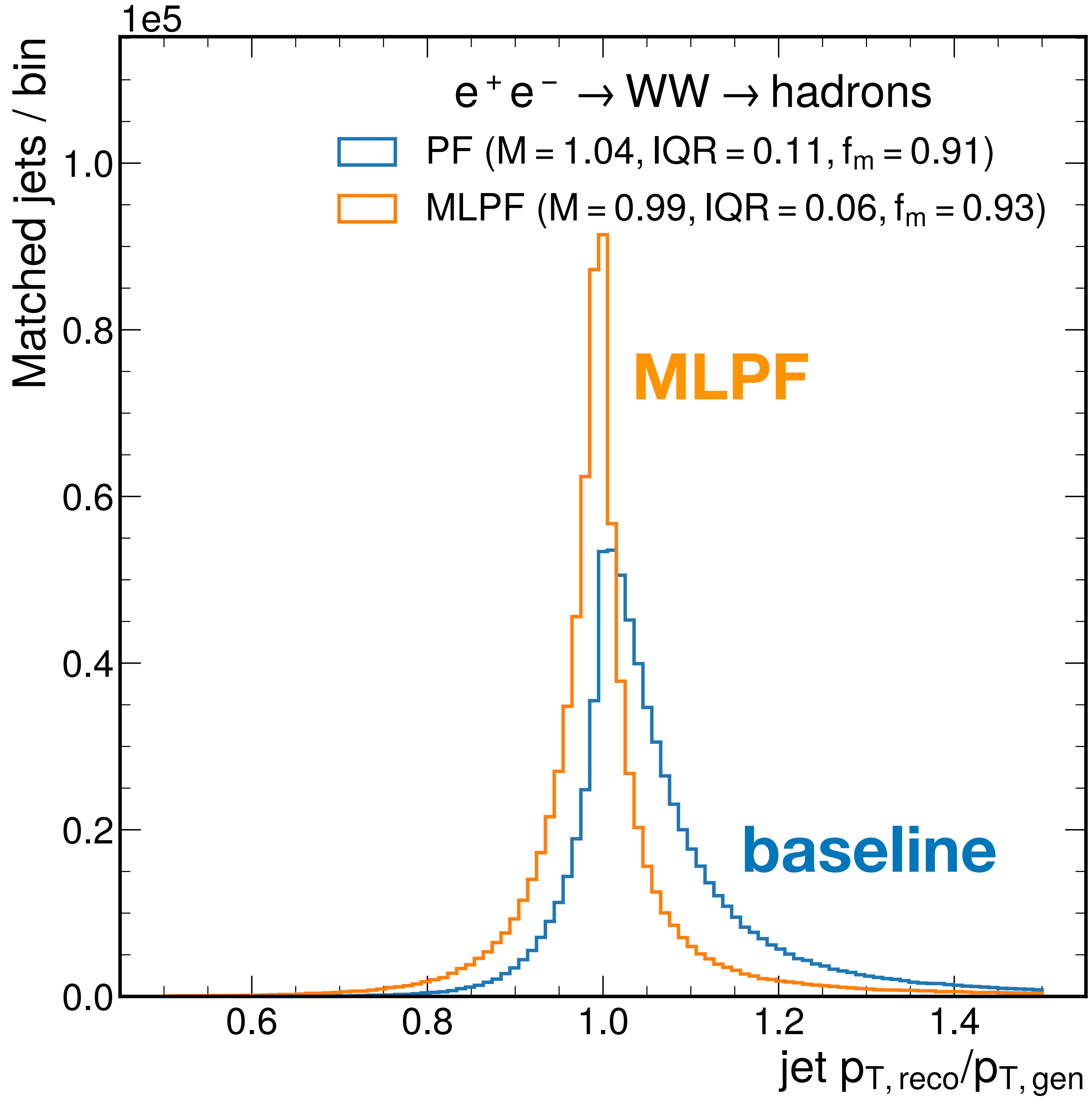
Trainable neural networks: $\mathscr{F}, \mathscr{G}, \mathscr{D}$

- 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

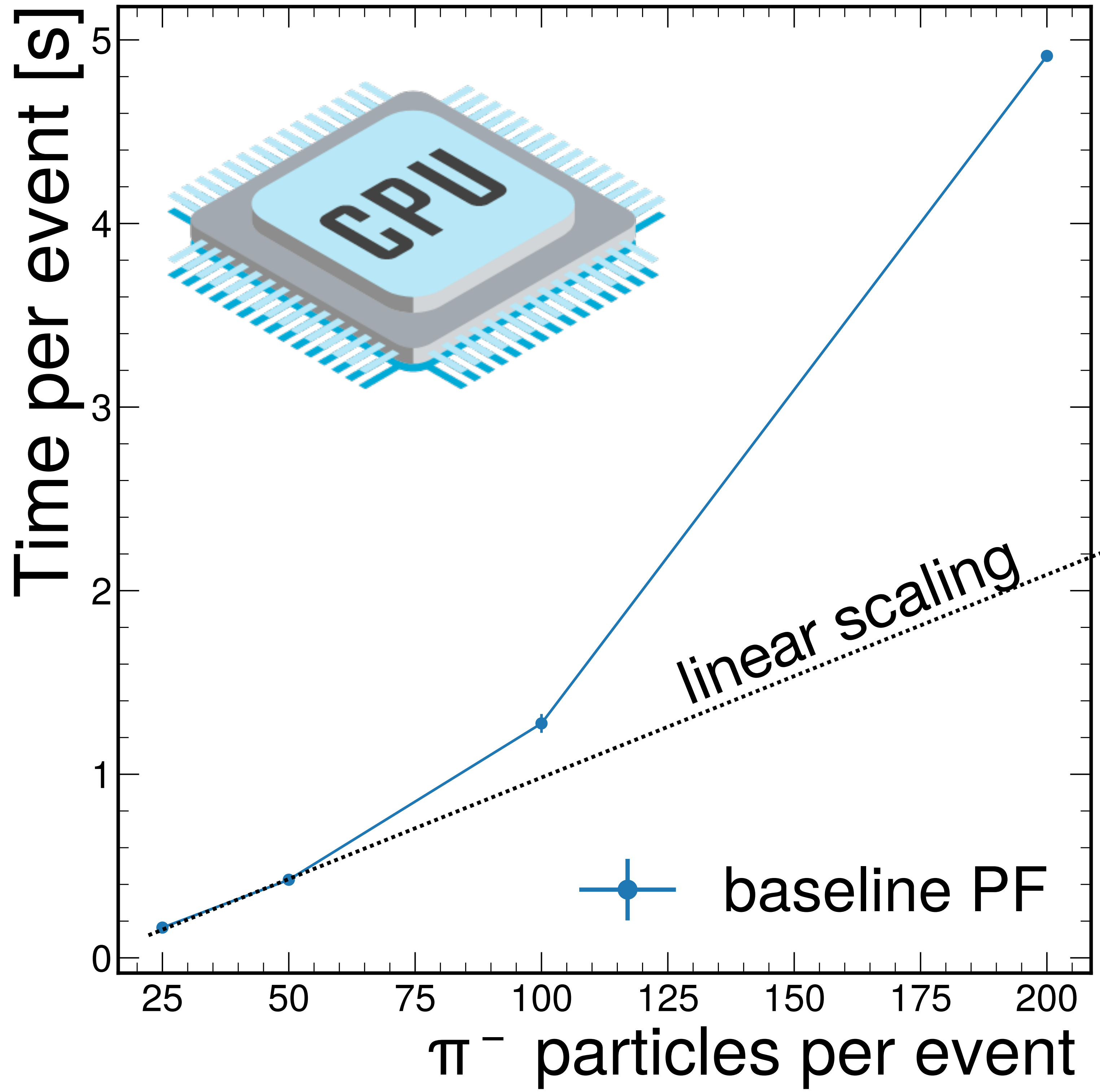
reconstructed particles → jets → match to true jets



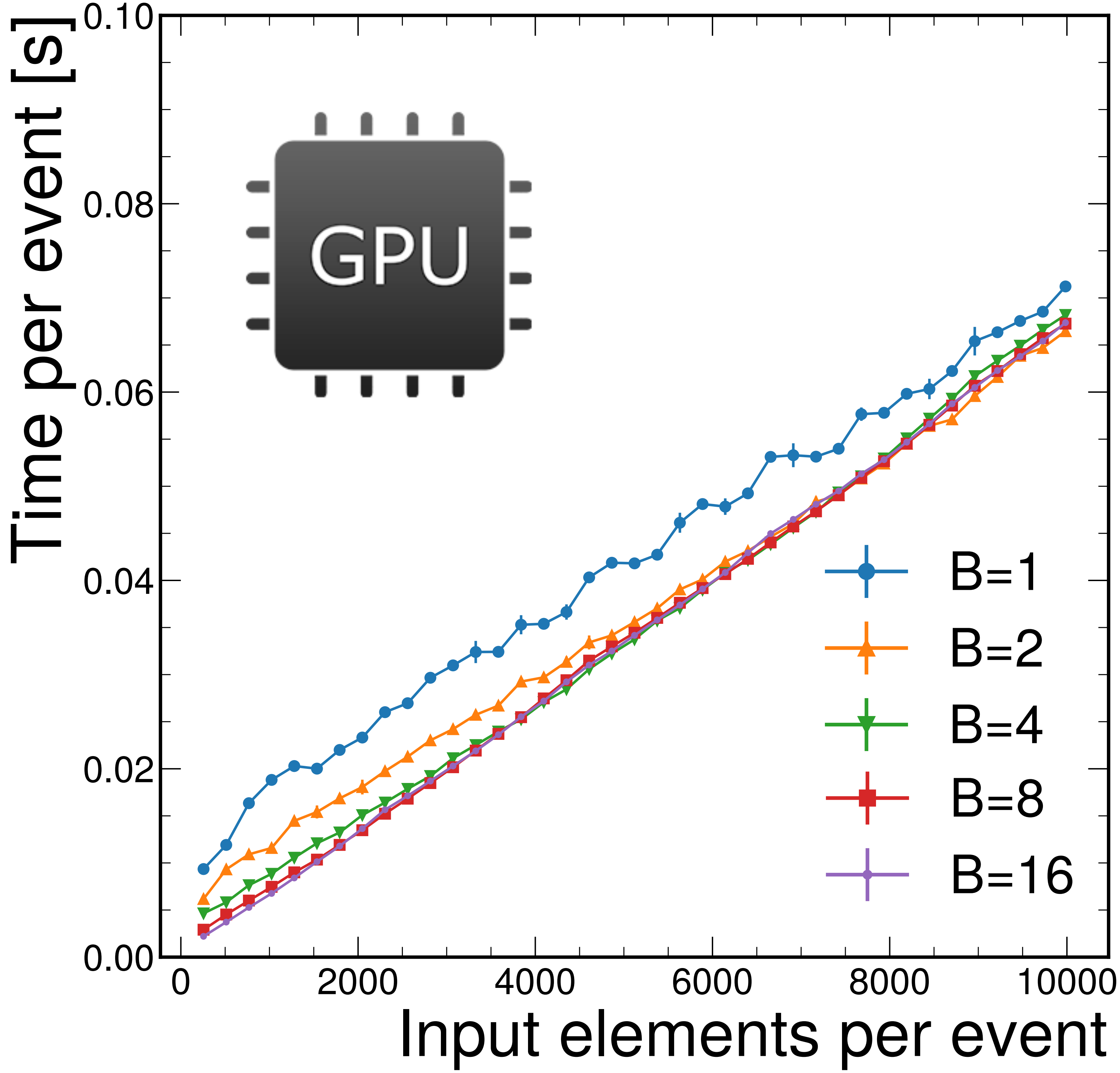
The resolution of jets is significantly better than the baseline*

*measured against jets from detectable Pythia particles

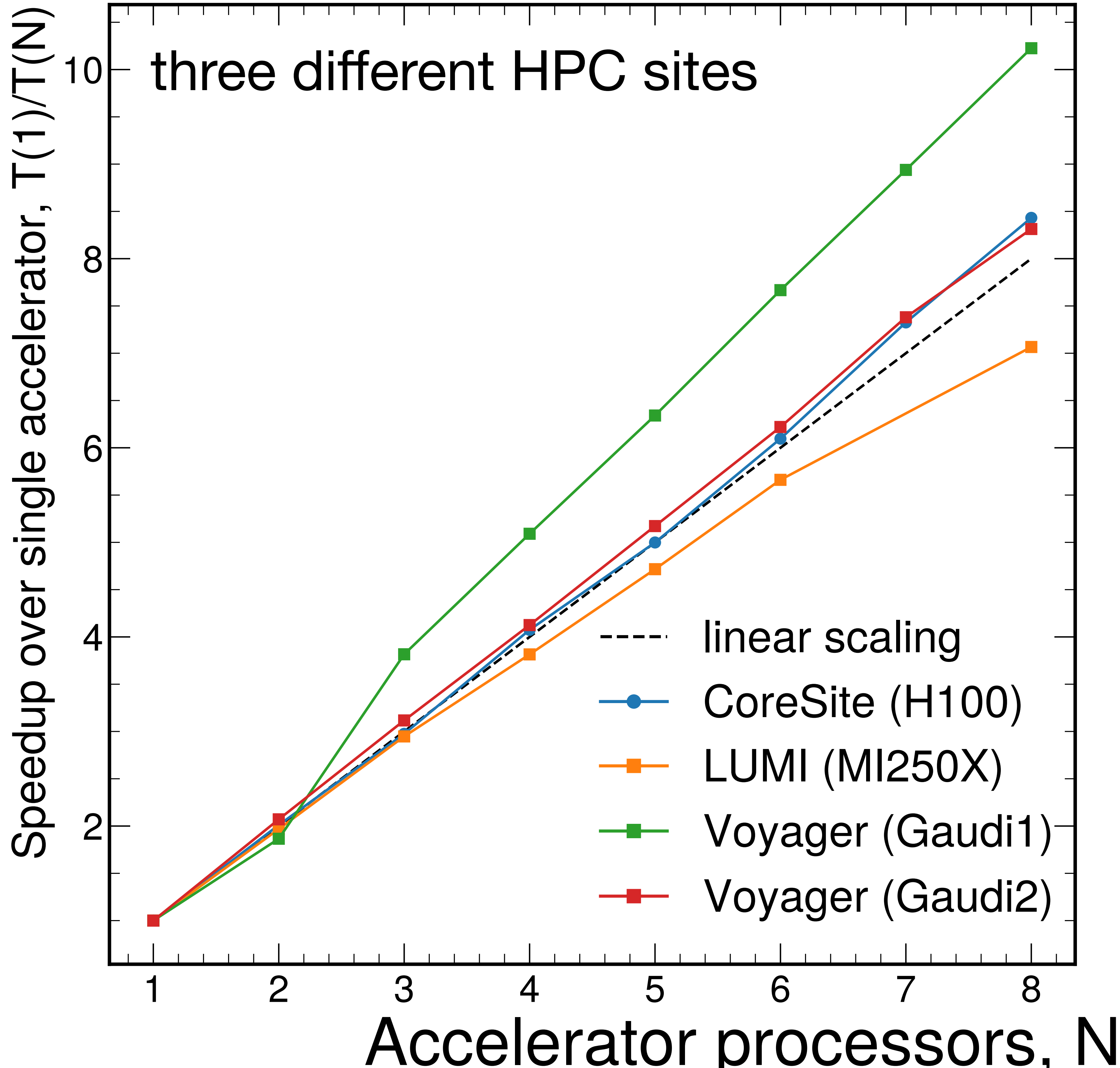
Baseline (untuned) algo runs **only on CPU**, scales ~quadratically, **runtime per event is in seconds**.



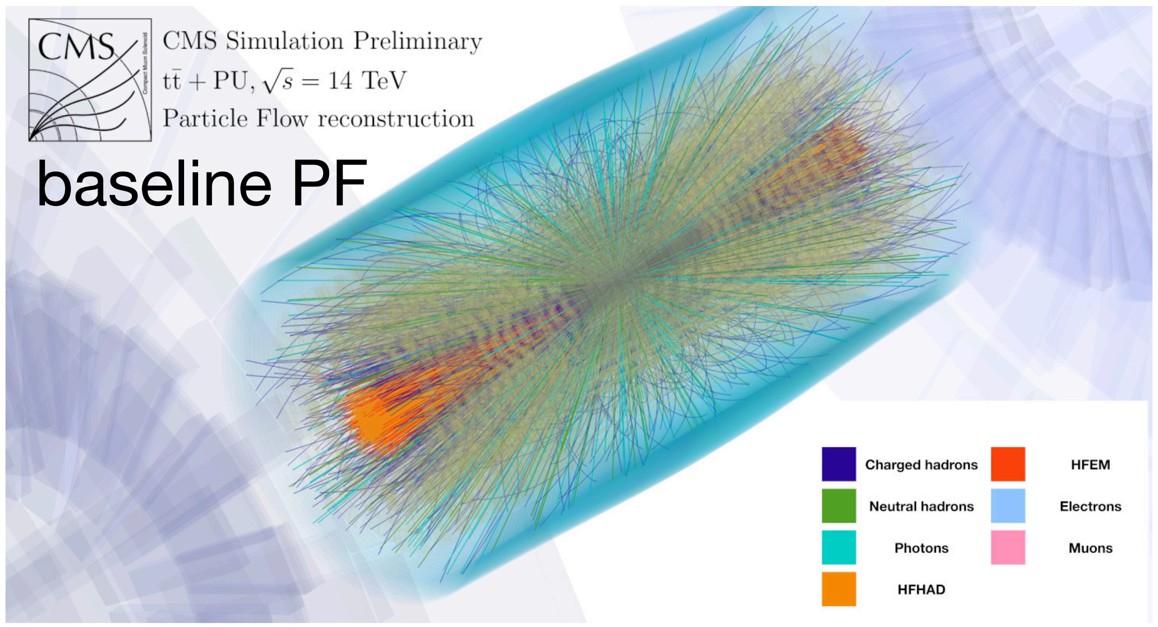
machine learning **scales linearly**, runs in **milliseconds per event.**

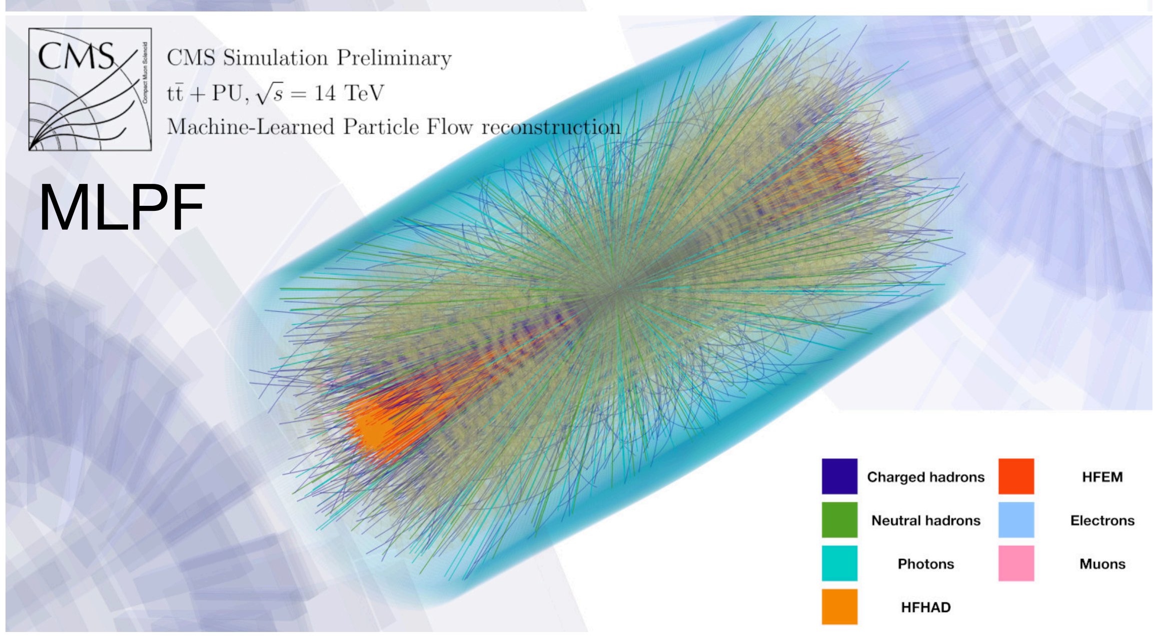


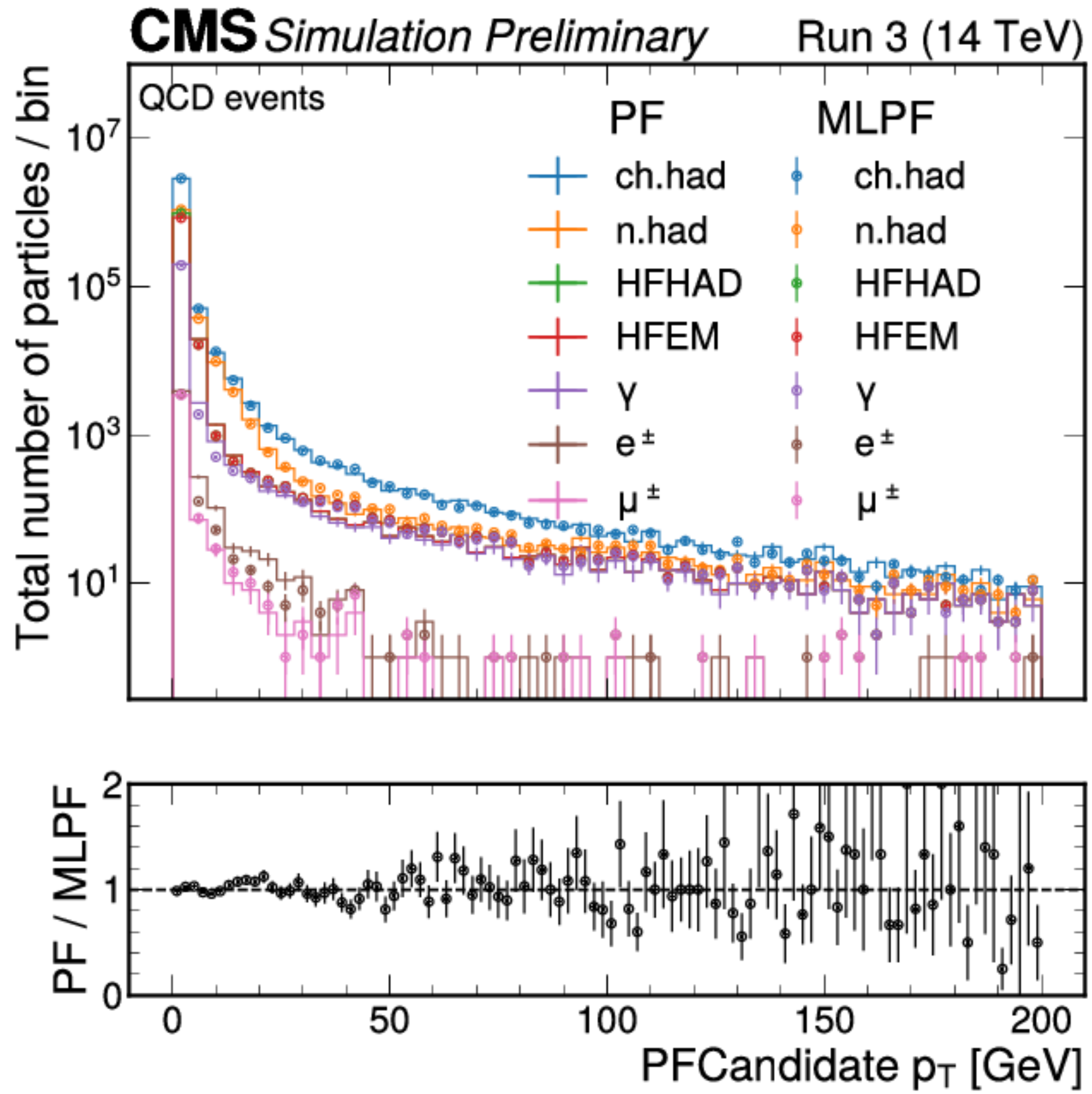
Portable on CPU, nVidia & AMD GPU, Intel Habana Gaudi chips



Applicable to CMS







Joosep Pata, Javier Duarte, FM, 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

Publications

Home > The European Physical Journal C > Article

MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks

Regular Article - Experimental Physics | Open access | Published: 02 May 2021

Volume 81, article number 381, (2021) Cite this article

Download PDF

You have full access to this open access article

Joosep Pata, Javier Duarte, Jean-Roch Vlimant, Maurizio Pierini & Maria Spiropulu

3543 Accesses 49 Citations 13 Altmetric Explore all metrics →

Eur. Phys. J. C **81**, 381 (2021). https://doi.org/10.1140/epjc/s10052-021-09158-w

communications physics

Explore content About the journal Publish with us

nature > communications physics > articles > article

Article | Open access | Published: 10 April 2024

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

Joosep Pata, Eric Wulff, Farouk Mokhtar, David Southwick, Mengke Zhang, Maria Girone & Javier Duarte

Communications Physics **7**, Article number: 124 (2024) | Cite this article

1342 Accesses | 1 Citations | 5 Altmetric | Metrics

Commun Phys **7**, 124 (2024). https://doi.org/10.1038/s42005-024-01599-5

Open to collaboration!

Testing ongoing in CMS

ACAT-2021 IOP Publishing
Journal of Physics: Conference Series **2438** (2023) 012100 doi:10.1088/1742-6596/2438/1/012100

Machine Learning for Particle Flow Reconstruction at CMS

Joosep Pata[1,*], Javier Duarte[2], Farouk Mokhtar[2], Eric Wulff[3], Jieun Yoo[4], Jean-Roch Vlimant[5], Maurizio Pierini[3], Maria Girone[3]
(on behalf of the CMS Collaboration)
[1]NICPB, Rävala pst 10, 10143 Tallinn, Estonia
[2]University of California San Diego, La Jolla, CA 92093, USA
[3]European Center for Nuclear Research (CERN), CH 1211, Geneva 23, Switzerland
[4]University of Illinois Chicago, Chicago, IL 60605, USA
[5]California Institute of Technology, Pasadena, CA 91125, USA

E-mail: *joosep.pata@cern.ch

Abstract. We provide details on the implementation of a machine-learning based particle flow algorithm for CMS. The standard particle flow algorithm reconstructs stable particles based on calorimeter clusters and tracks to provide a global event reconstruction that exploits the combined information of multiple detector subsystems, leading to strong improvements for quantities such as jets and missing transverse energy. We have studied a possible evolution of

Available on CMS information server **CMS DP -2022/061**

The Compact Muon Solenoid Experiment

CMS Performance Note

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

16 November 2022

Progress towards an improved particle flow algorithm at CMS with machine learning

Joosep Pata, Farouk Mokhtar, Javier Duarte, Eric Wulff, Dylan Rankin, Maurizio Pierini, Jean-Roch Vlimant

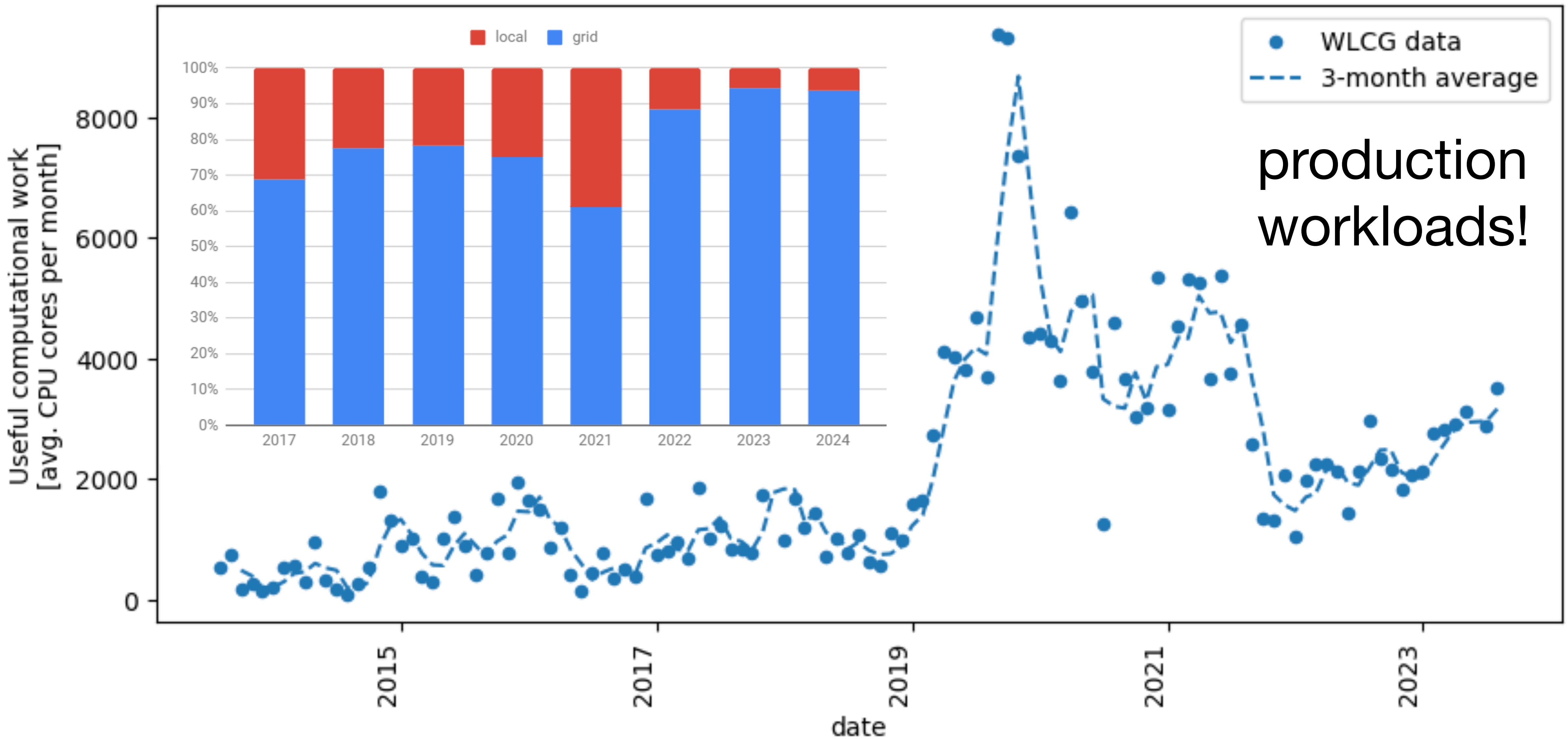
Abstract

ACAT 2022, http://cds.cern.ch/record/2842375

J. Phys. Conf. Ser. 2438 012100, 10.1088/1742-6596/2438/1/012100

Estonian Tier2

~8k CPU cores, both grid and local scientific usage.

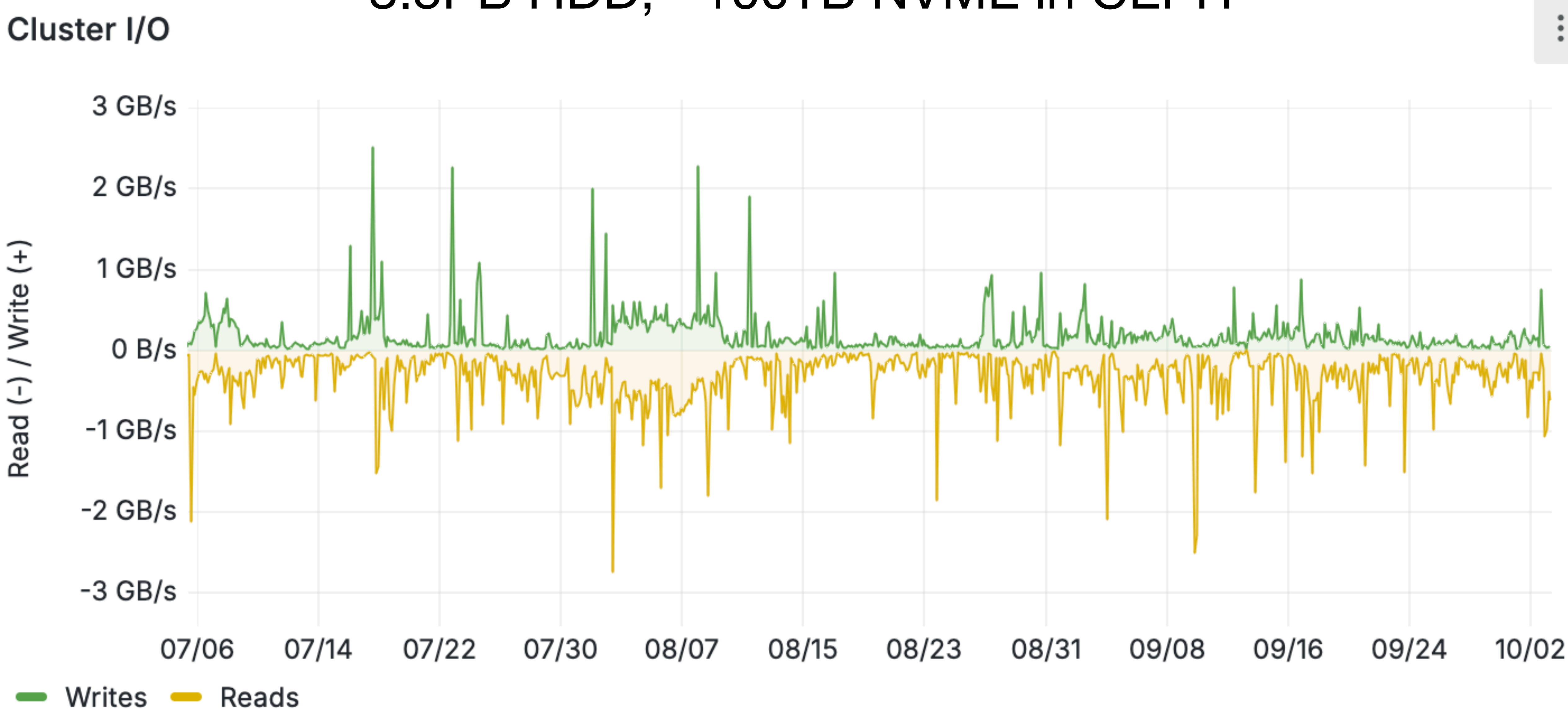


production workloads!

>10 years of contribution to CERN CMS computing through WLCG! But most recent CPUs from 2019…

Storage

~3.5PB HDD, ~100TB NVME in CEPH

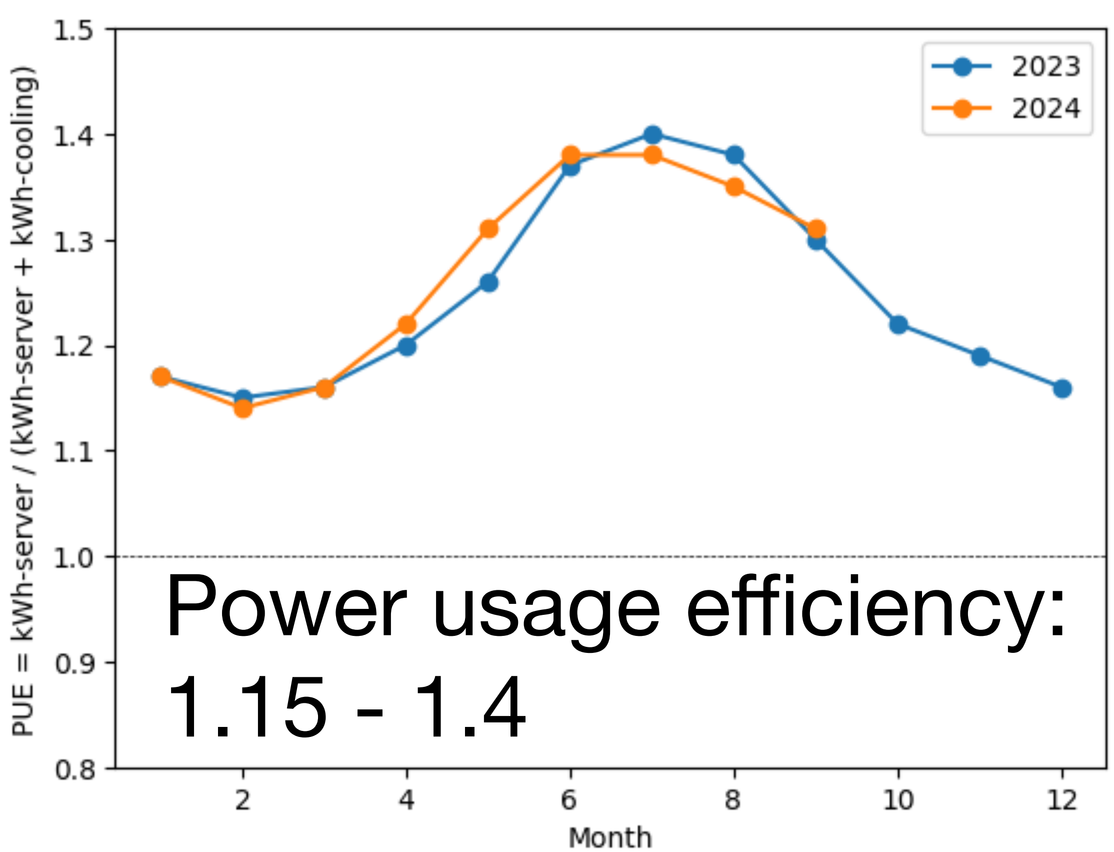


Sustained, stable and efficient production workloads (but HDD out of warranty and >7 years old)

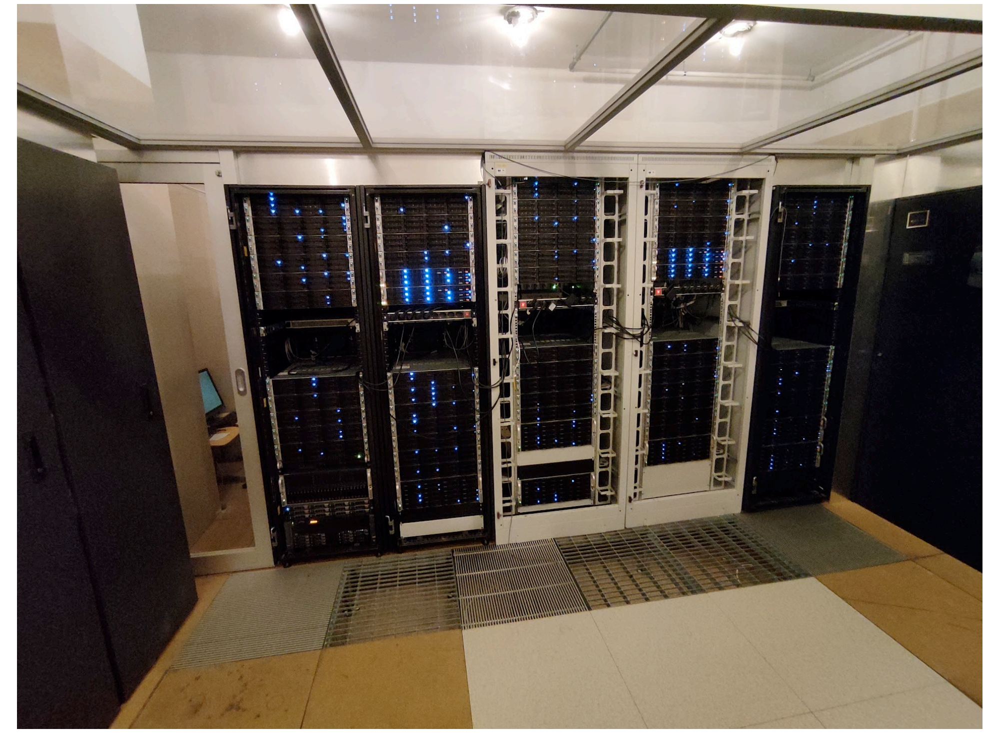
Current infrastructure

- room-based CRAC[1], rooftop dry-cooler[2], cold air supply via raised floor

- Limited resilience, out of warranty



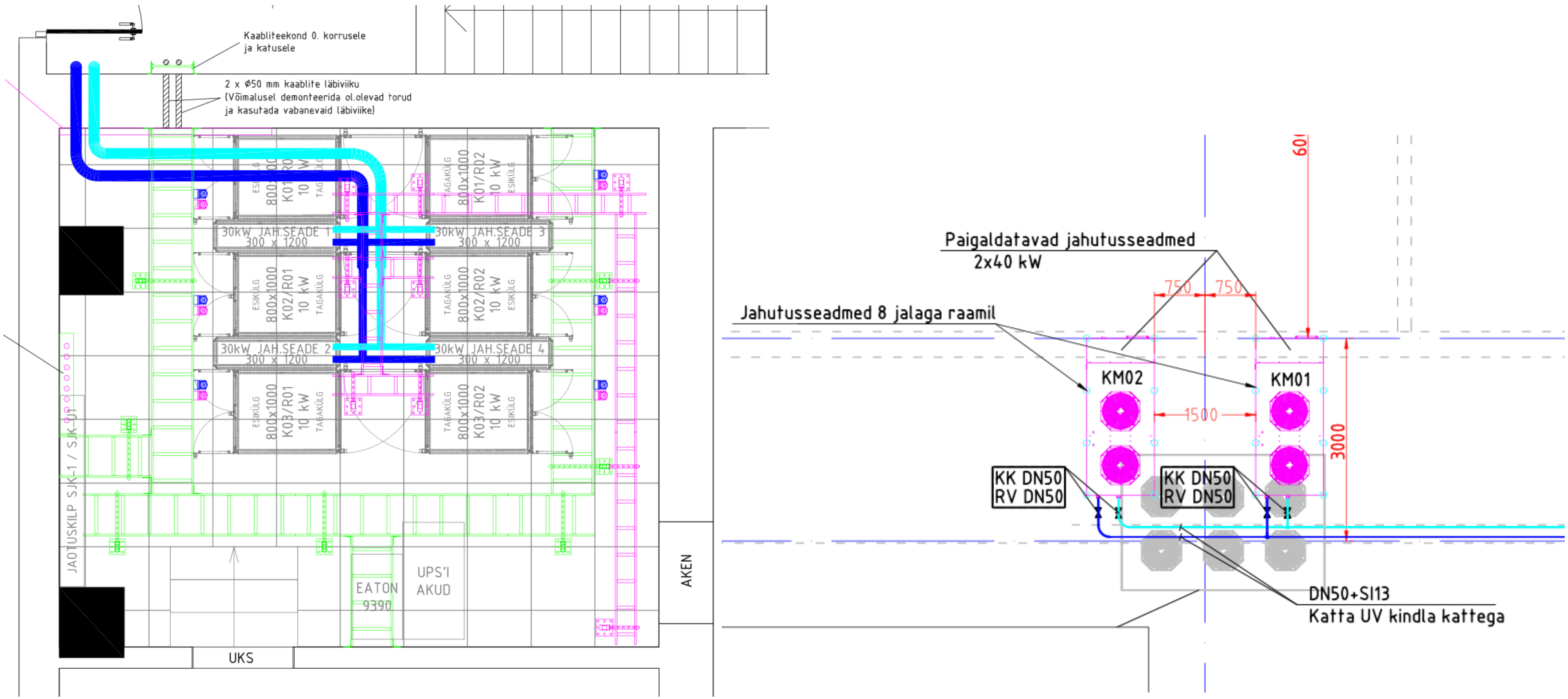
Power usage efficiency:
1.15 - 1.4



1) Rittal UPA 491B, 50-70 kW cooling power
2) SAL6N 6466, 60-90 kW cooling power

Planned infrastructure

Improve sustainability and efficiency



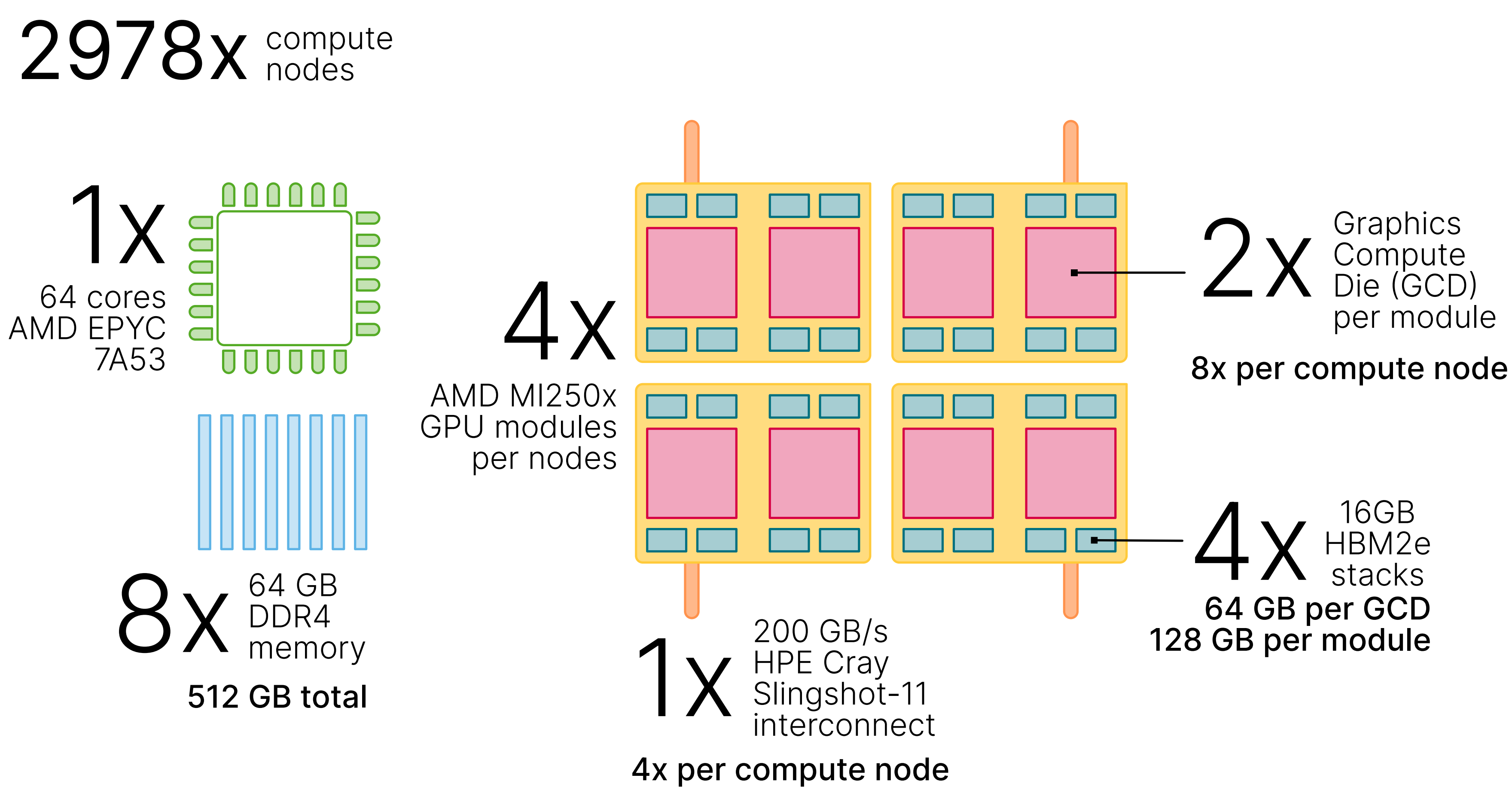
4x in-row coolers, 2x chillers on the rooftop, direct hot air extraction from servers, better heat containment.

LUMI for HEP

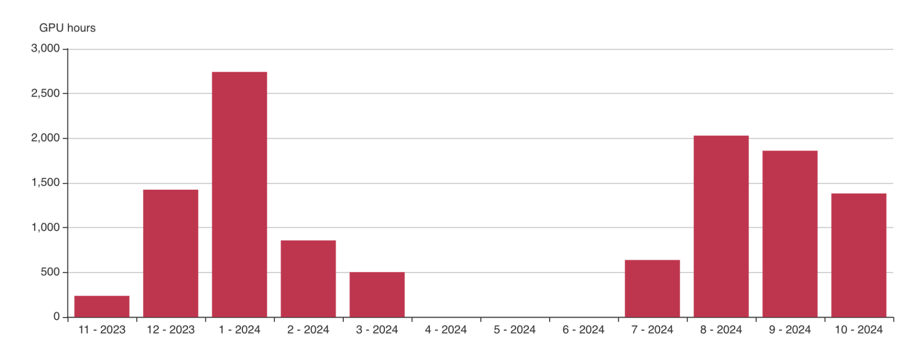
LUMI is a pre-exascale EuroHPC supercomputer in Finland

Estonia has ~1% share through Estonian Scientific Computing Infrastructure ETAIS, HEP users are already starting to use it for publications.

AMD GPUs for ML

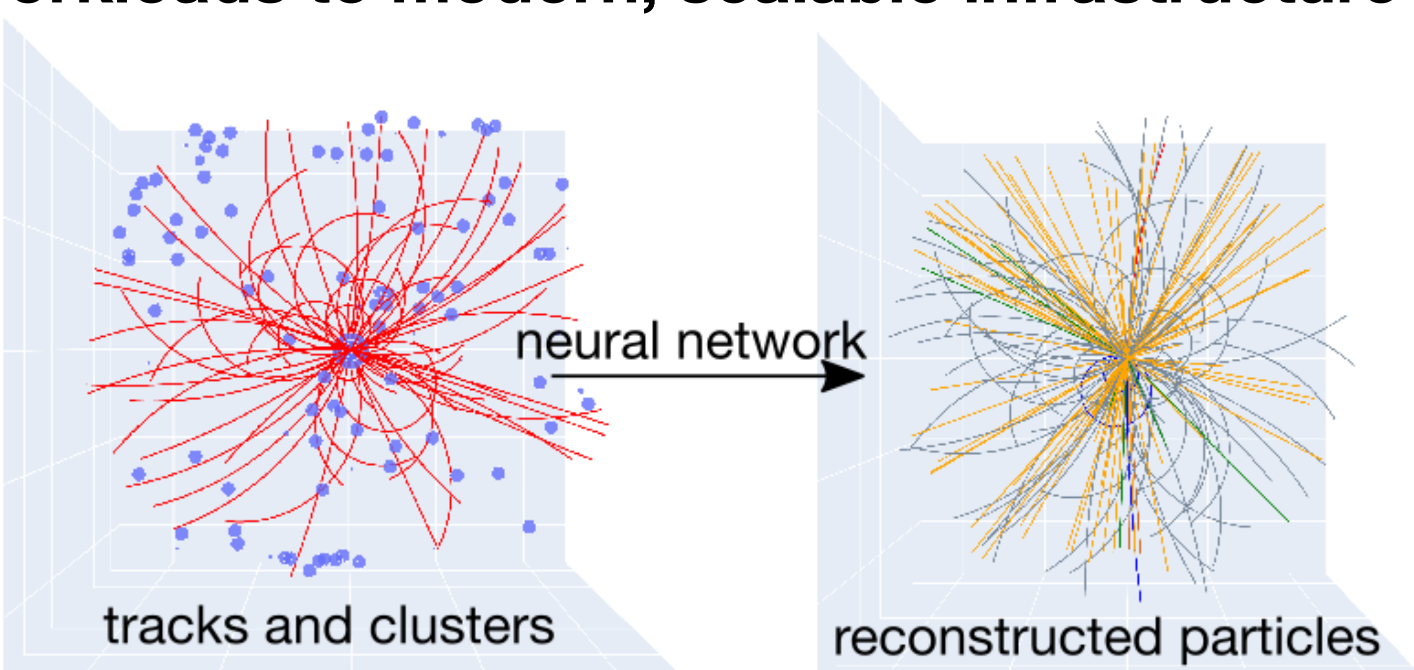


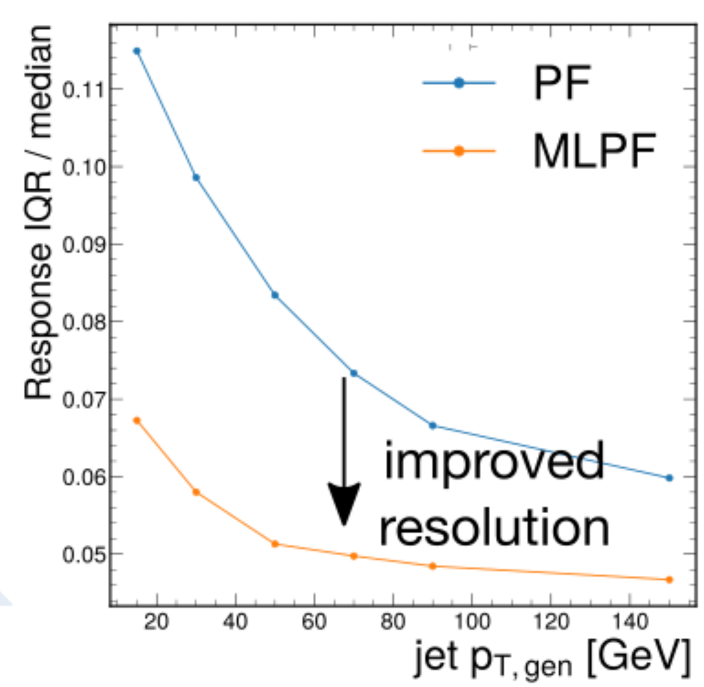
AMD ROCm + PyTorch =



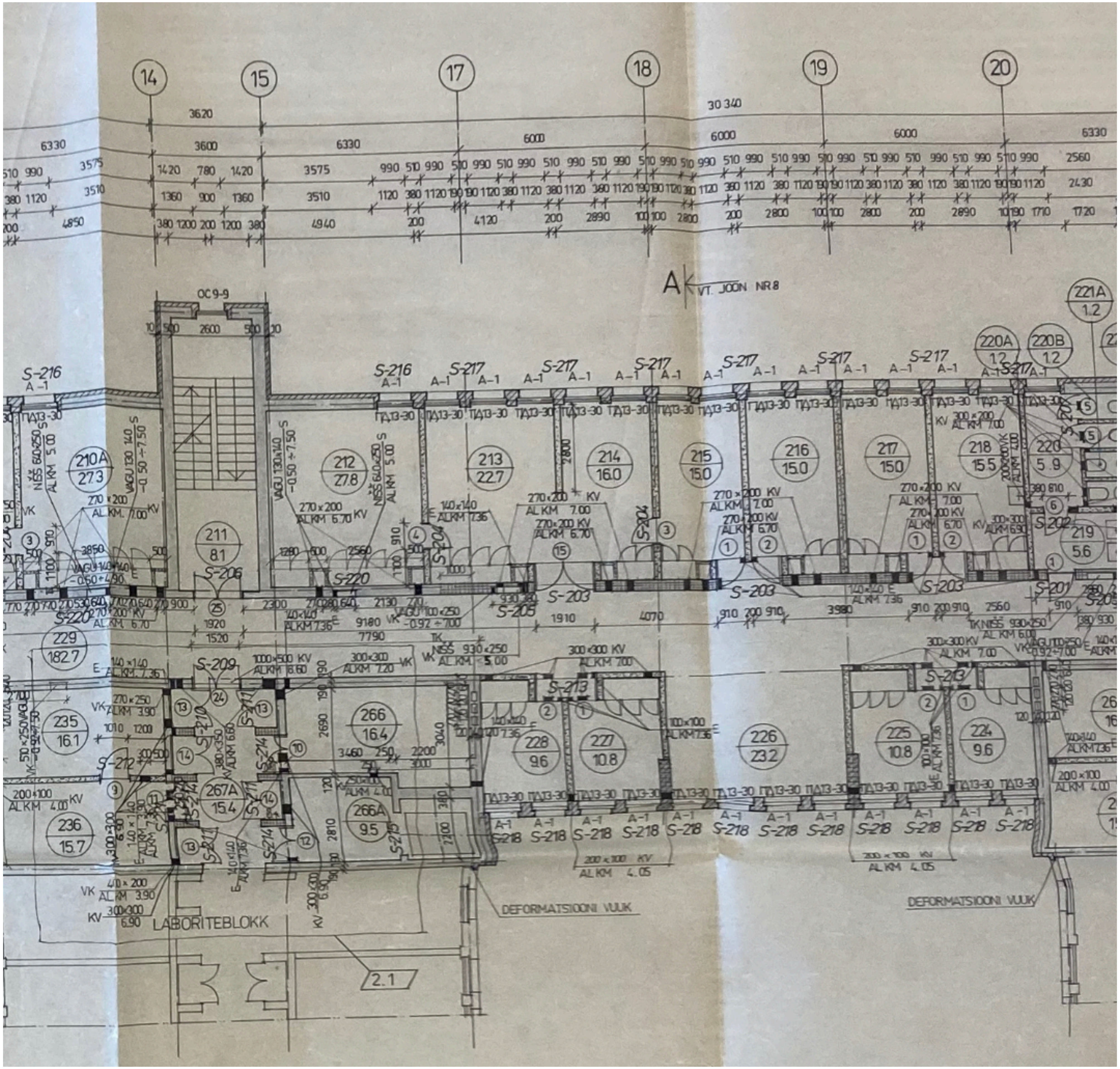
Summary

- AI/ML is an increasingly **important tool for science and society,** but dedicated and sustained resources are needed
- Together with UCSD, CERN and others, we are developing a common **ML-based reconstruction stack for current and future detectors**
- Local HPC / Tier2 is a force multiplier for scientists & national contribution to CERN science; allows efficient use of infrastructure, flexible for R&D
- Need to collaborate with regional HPC (e.g. LUMI) to efficiently **scale out workloads to modern, scalable infrastructure**





Backup


14
15
17
18
19
20
30 340
A
VT. JOON NR 8
OC 9-9
210A
27.3
211
8.1
212
27.8
213
22.7
214
16.0
215
15.0
216
15.0
217
15.0
218
15.5
220
5.9
219
5.6
220A
1.2
220B
1.2
221A
1.2
229
182.7
235
16.1
236
15.7
266
16.4
266A
9.5
267A
15.4
228
9.6
227
10.8
226
23.2
225
10.8
224
9.6
LABORITEBLOKK
DEFORMATSIOONI VUUK
DEFORMATSIOONI VUUK
2.1

Open FullSim data

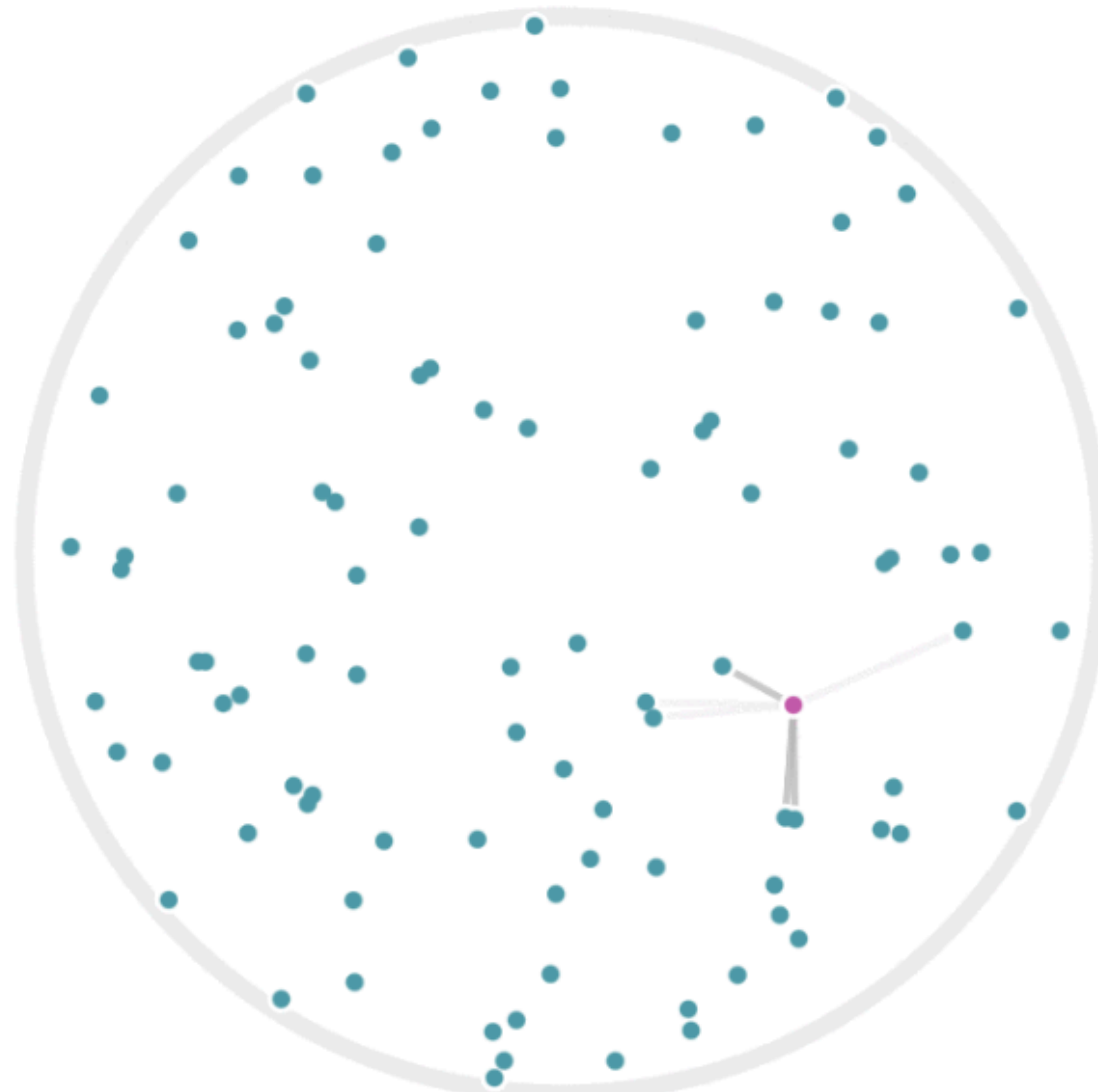
Particle Flow Reconstruction

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.

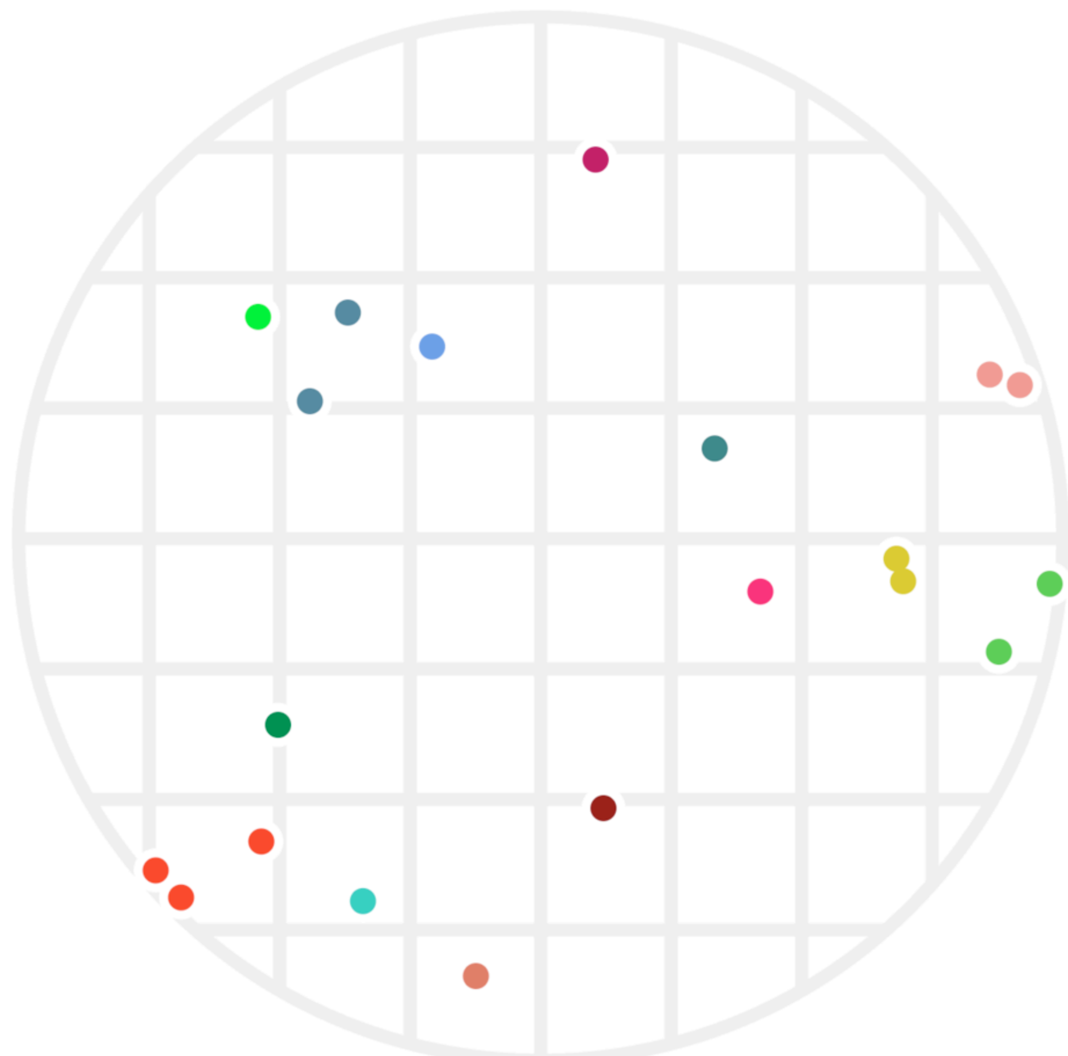
https://www.coe-raise.eu/od-pfr
https://doi.org/10.5281/zenodo.8260741

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.

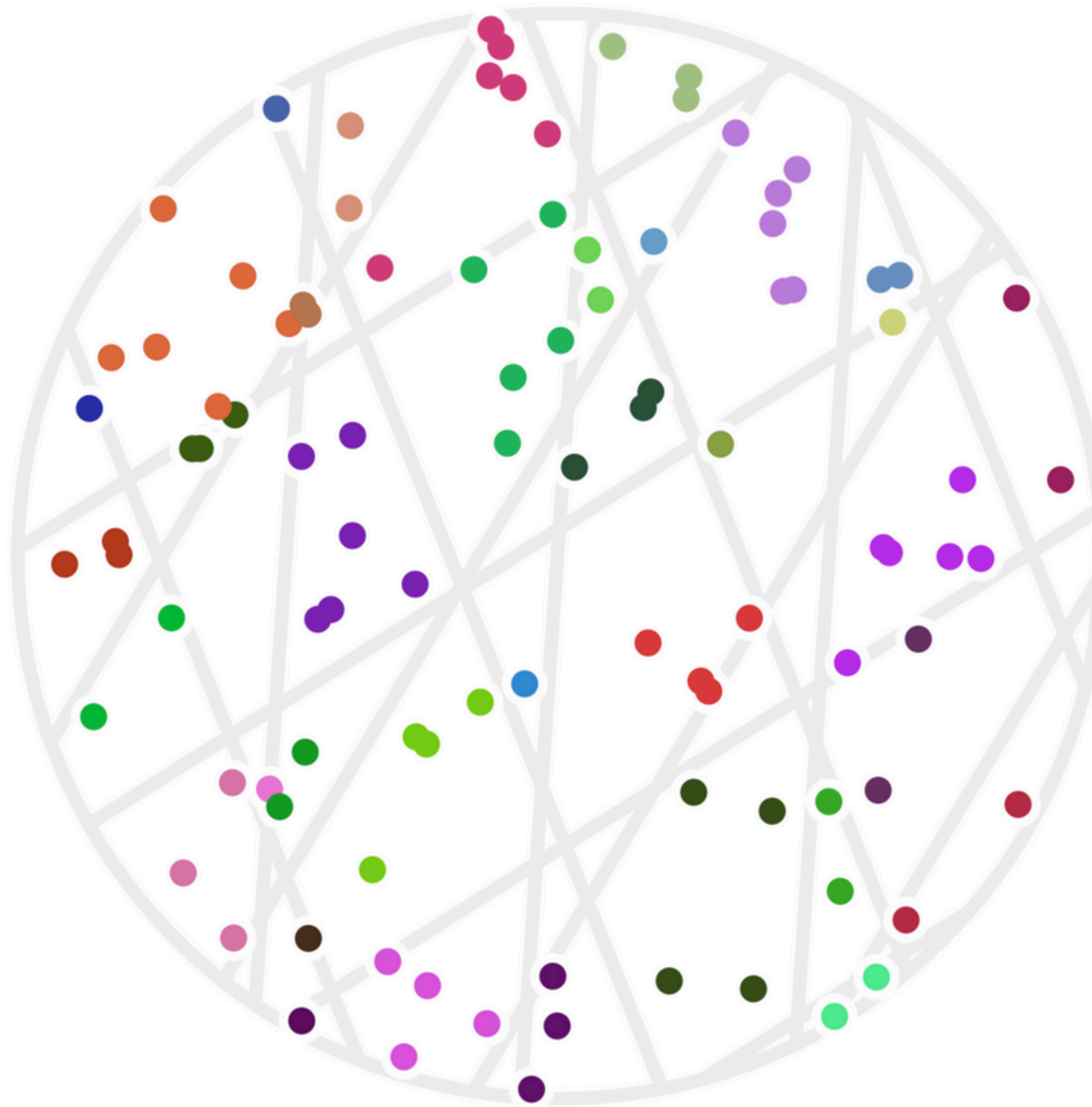


particle features to bin index: hash function

Locality sensitive hashing

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

Randomized bins (hash functions) work even better!

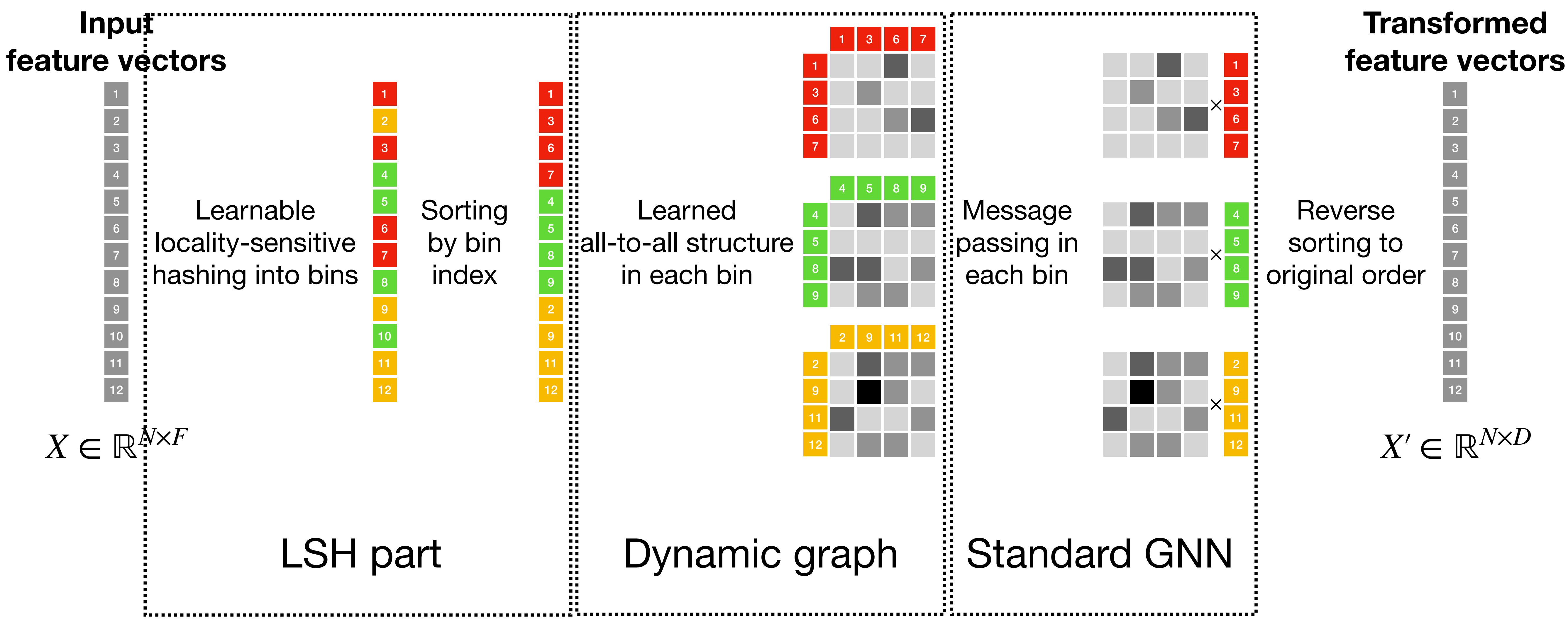


Simple to implement in TensorFlow, PyTorch, JAX using native operations: high portability to Nvidia, AMD, Intel Gaudi etc. **today**

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

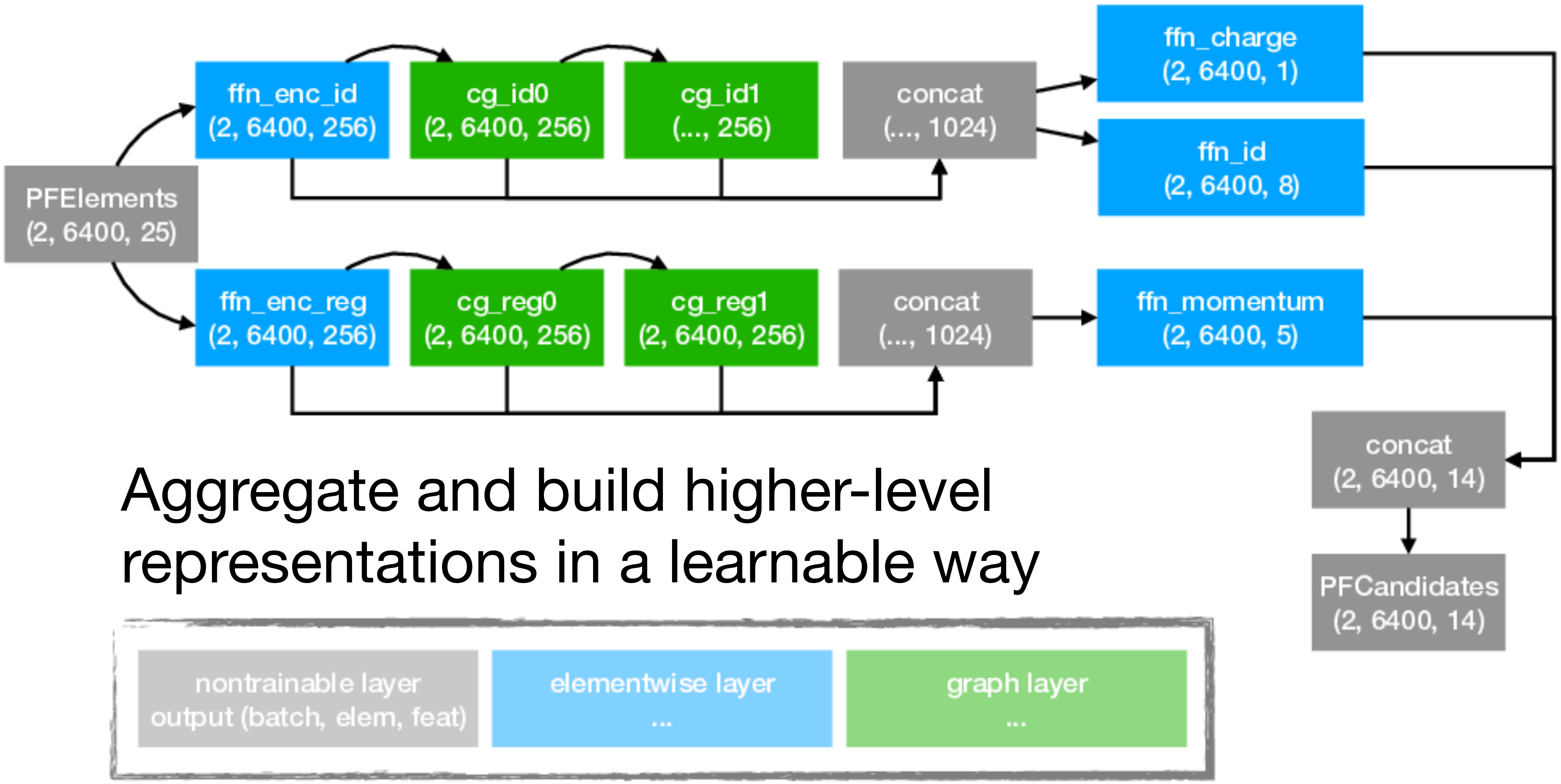
Scalable GNN based on the Reformer architecture

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

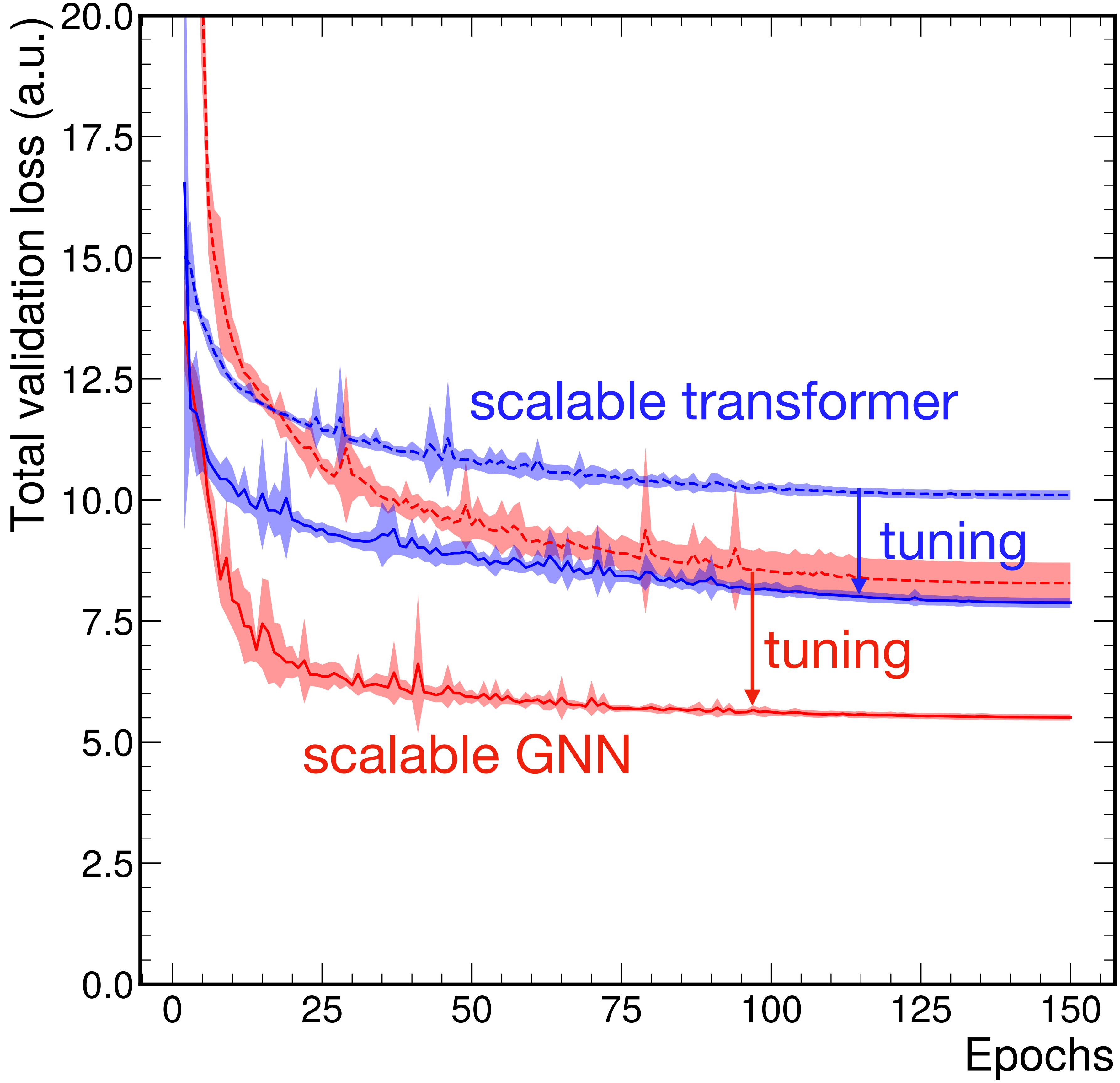


https://blog.research.google/2020/01/reformer-efficient-transformer.html

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

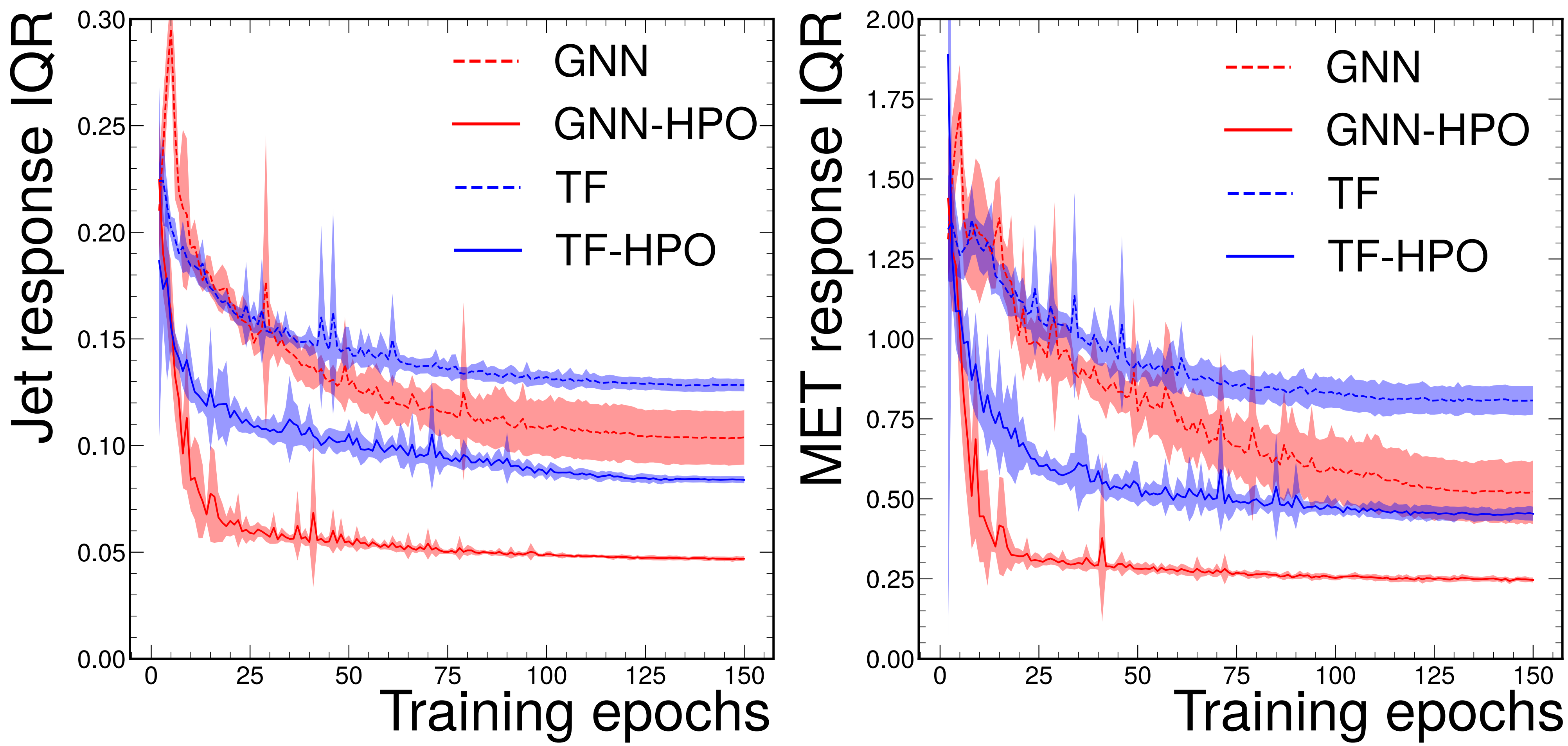


Aggregate and build higher-level representations in a learnable way

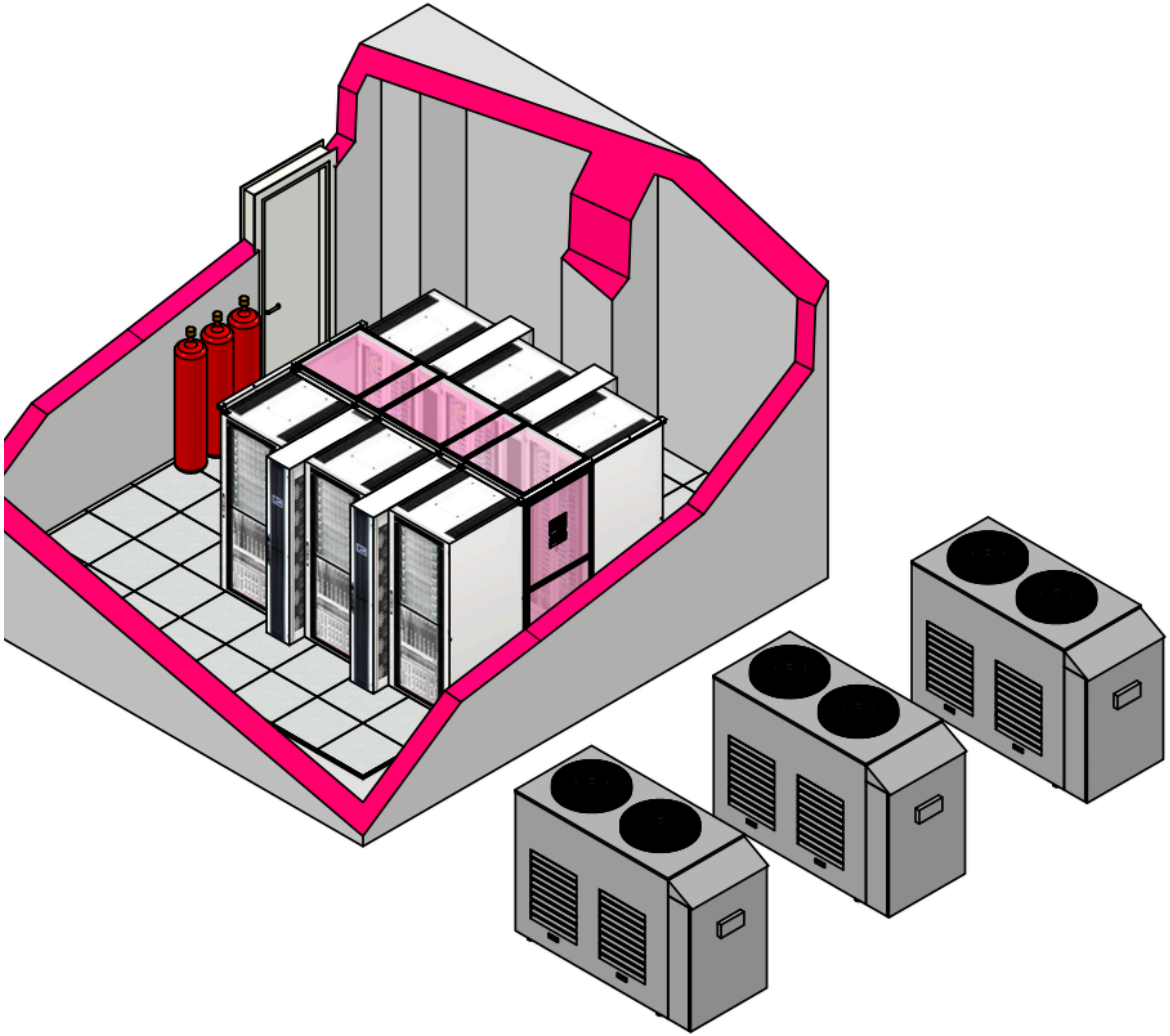


Better-than-SOTA result!

Even though we optimise a particle-based loss…



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



Rittal – The System.
Faster – better – everywhere.

RITTAL

Energy consumption Calculation

Model : SK3232.721

Type : FreeCooling

CITY : TALLIN

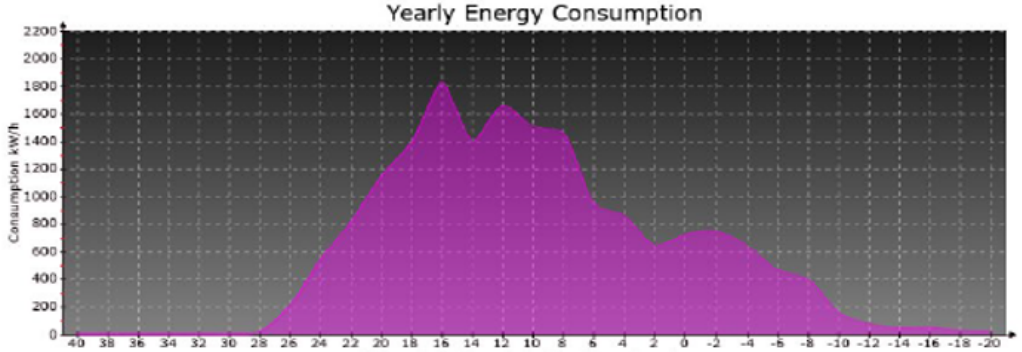
Yearly Energy Consumption		
FreeCooling Chiller :	25.191,64	kWh
Additional :	10.512,00	kWh
Total :	35.703,64	kWh

Yearly Operating Cost		
FreeCooling Chiller :	€	2.519,16
Additional :	€	1.051,20
Total :	€	3.570,36

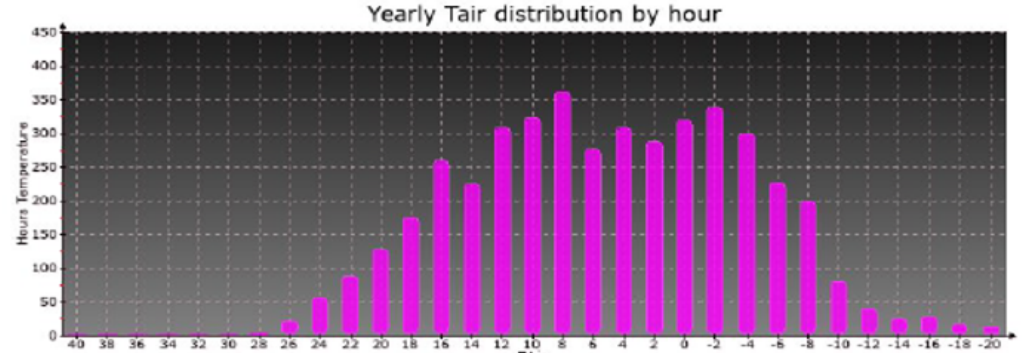
INPUT DATA

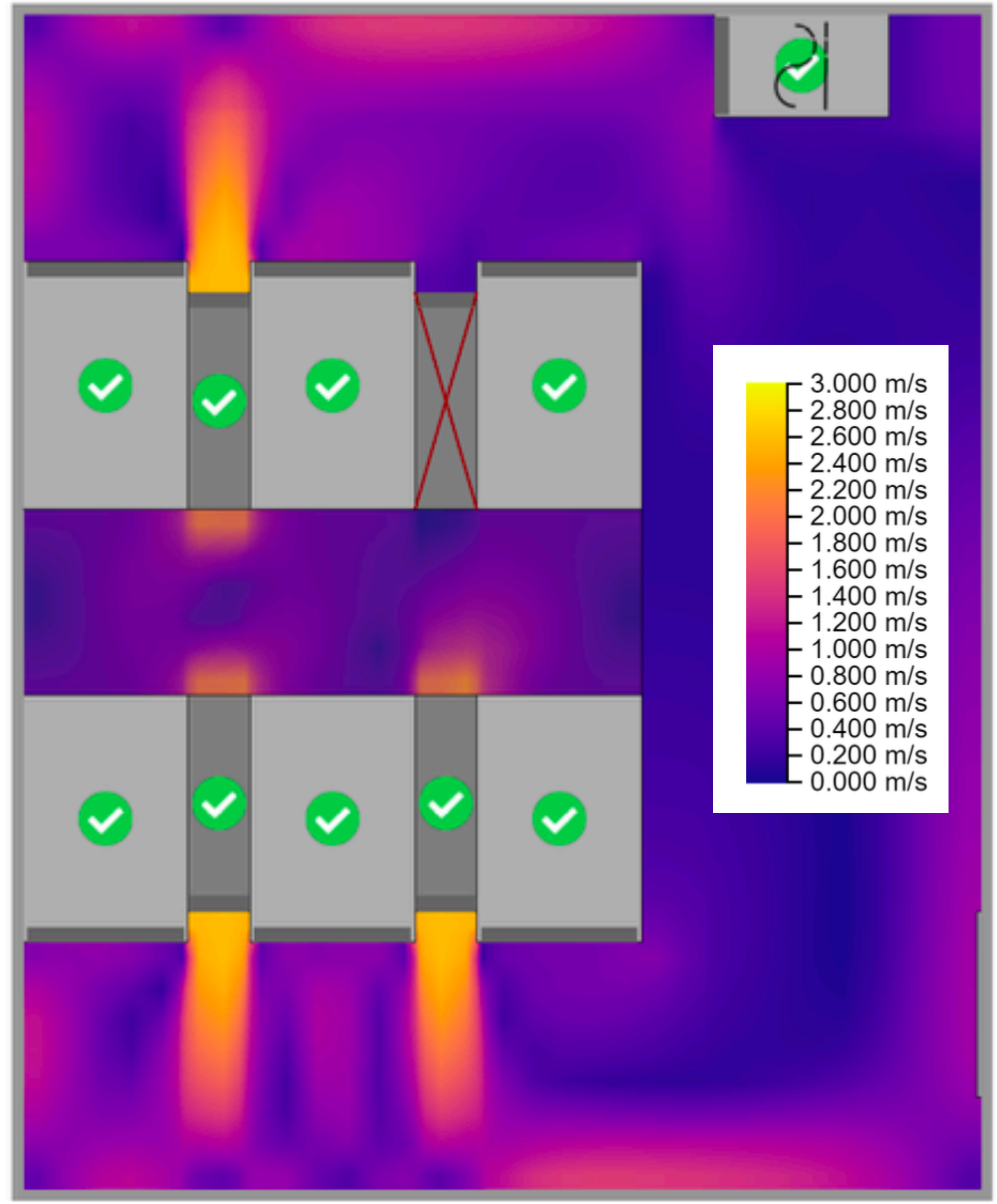
Inlet fluid temperature : 20 °C
Outlet fluid temperature : 15 °C
Ethylene Glycol : 40 %
Thermal load : 32 kW

Yearly Energy Consumption



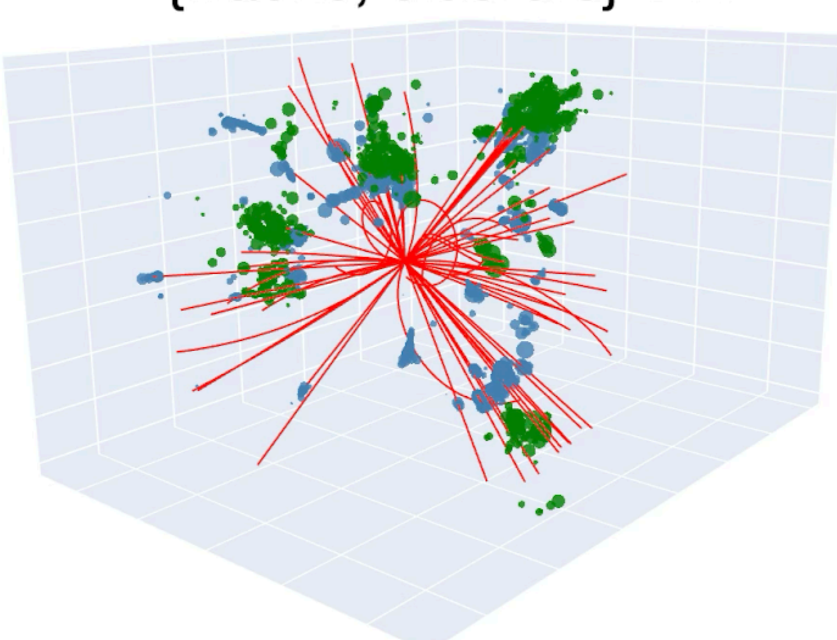
Yearly Tair distribution by hour



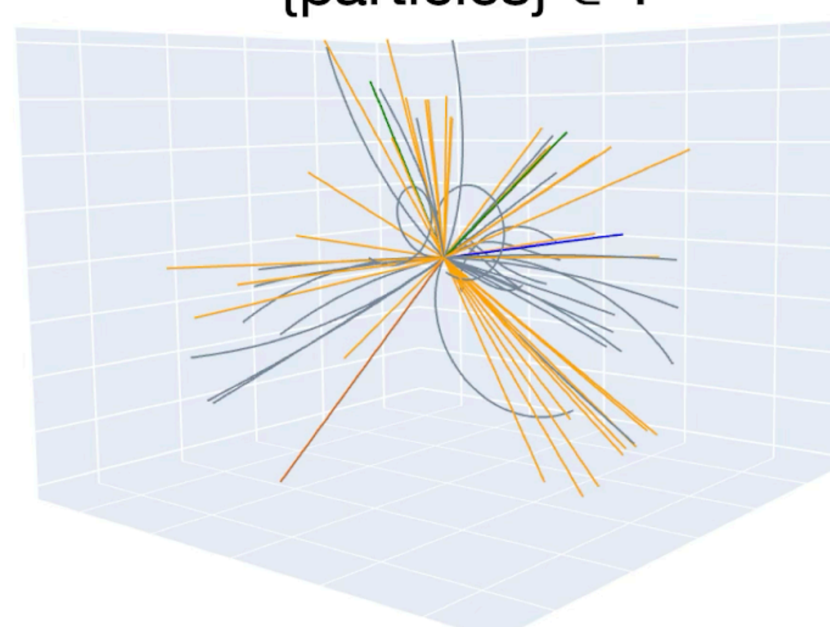


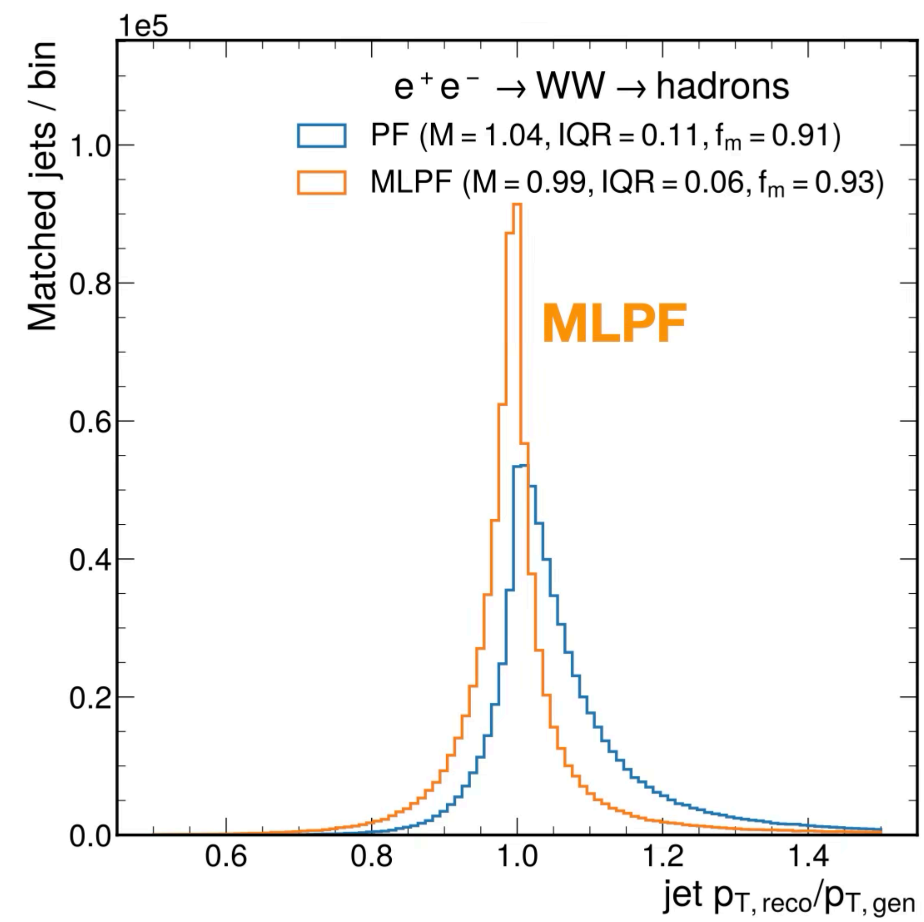
	IT load, kW	No	Power consumption, kW	Total annual energy consumption, kWh	Partial PUE (only cooling)
LCP CW	60	4	0,8063	28253	1,054
Chiller	60	2	-	35704	1,068
Total cooling system				63957	**1,122**

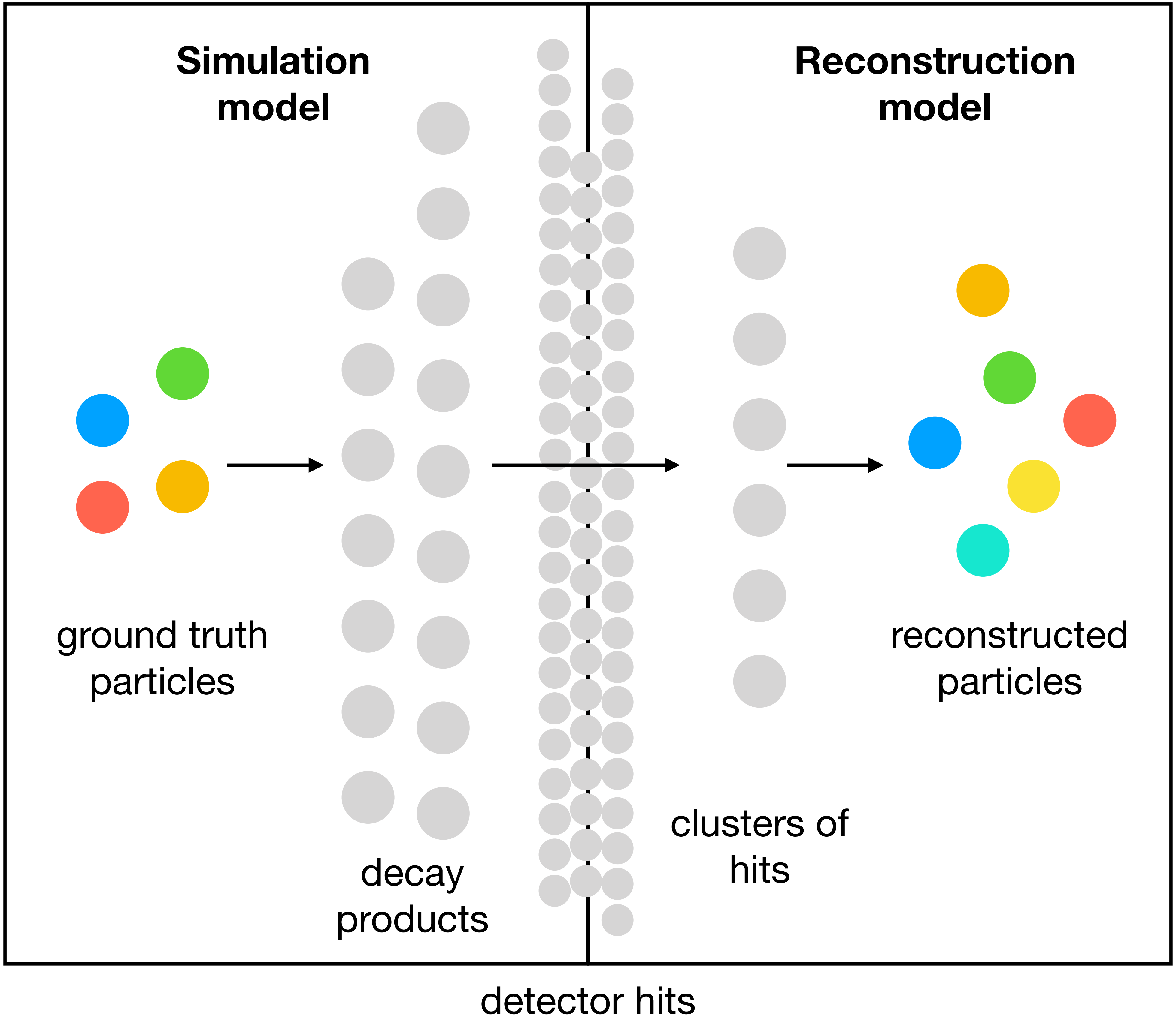
{tracks, clusters} ∈ X



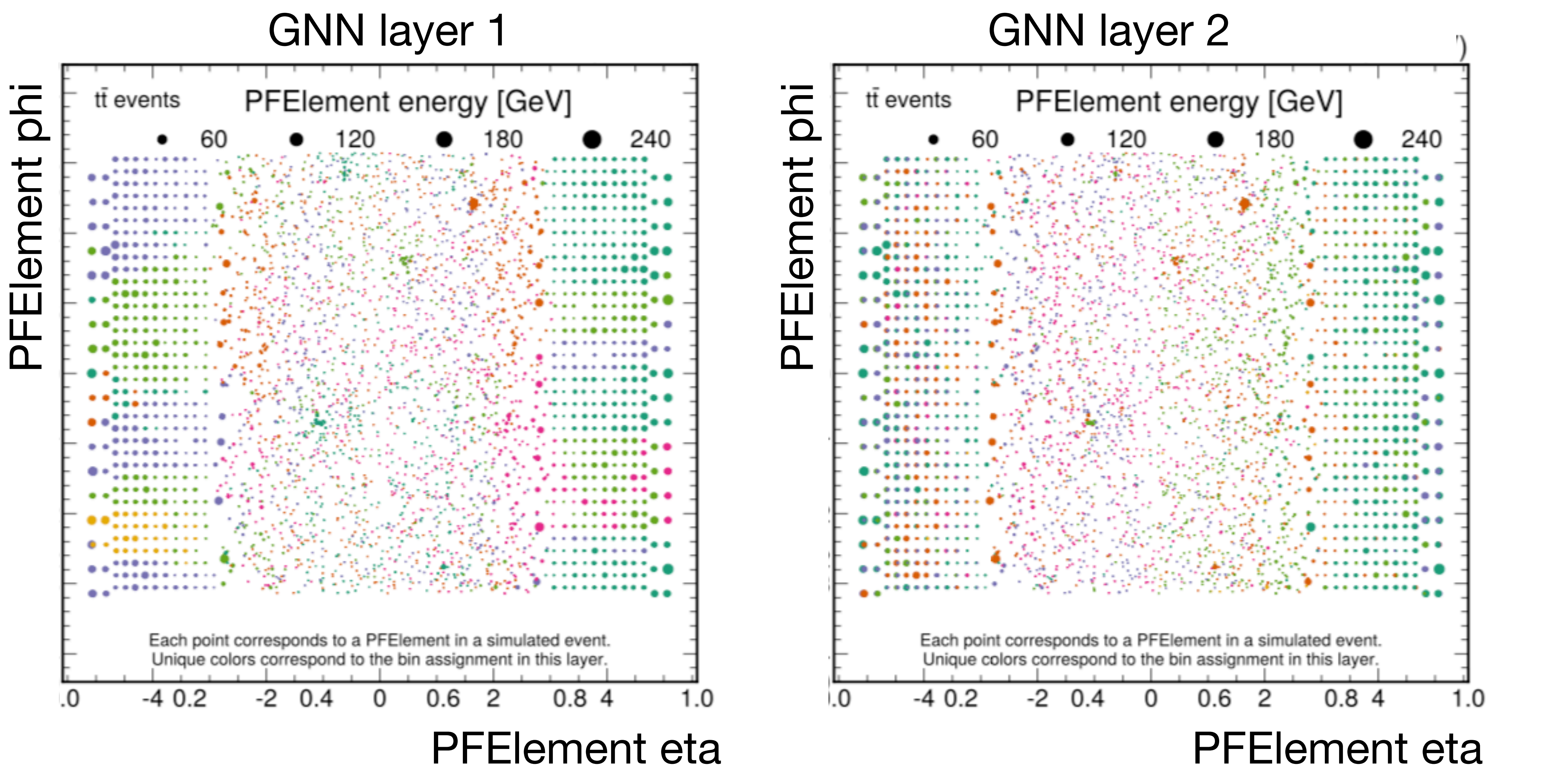
{particles} ∈ Y






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!