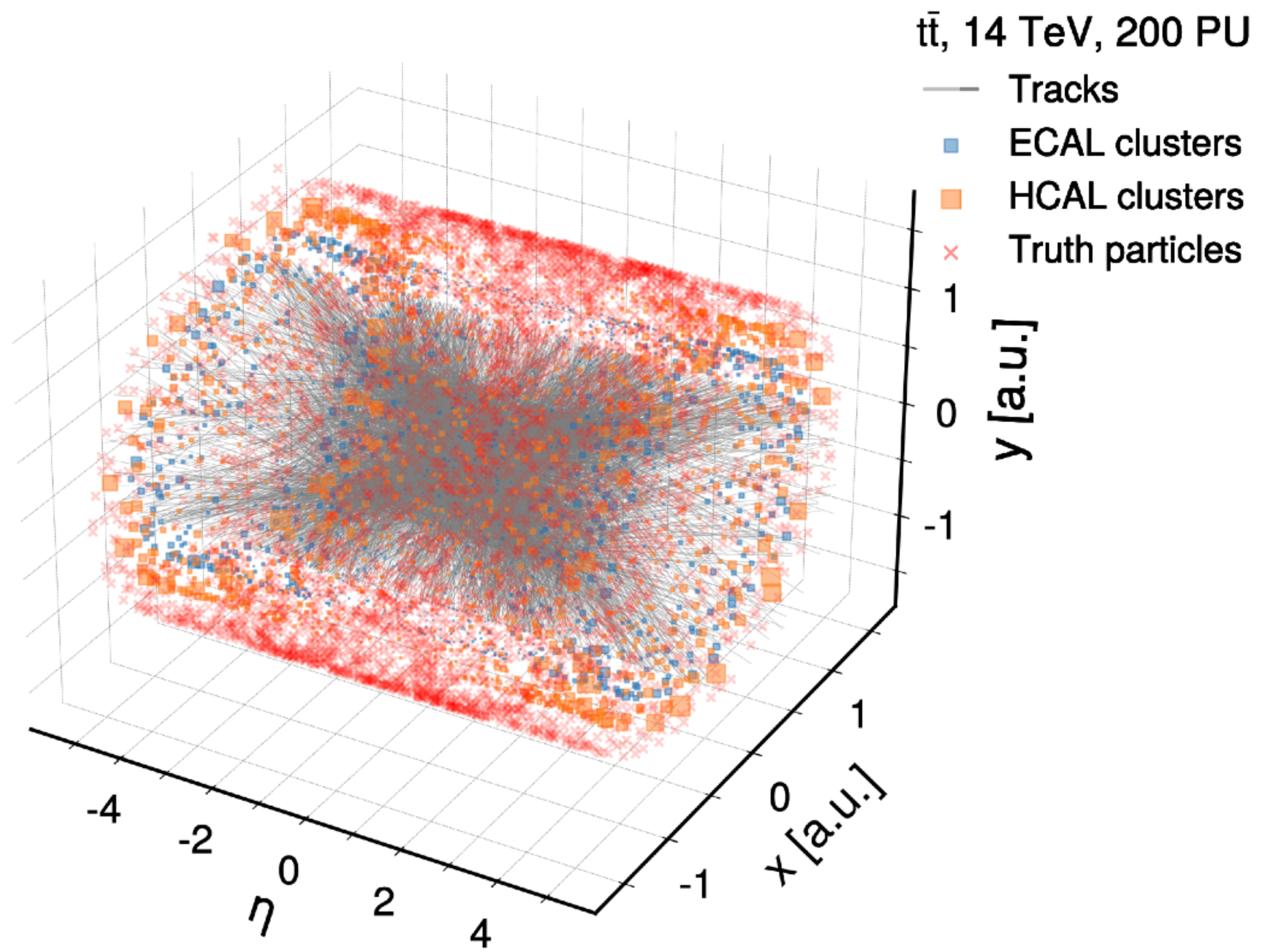
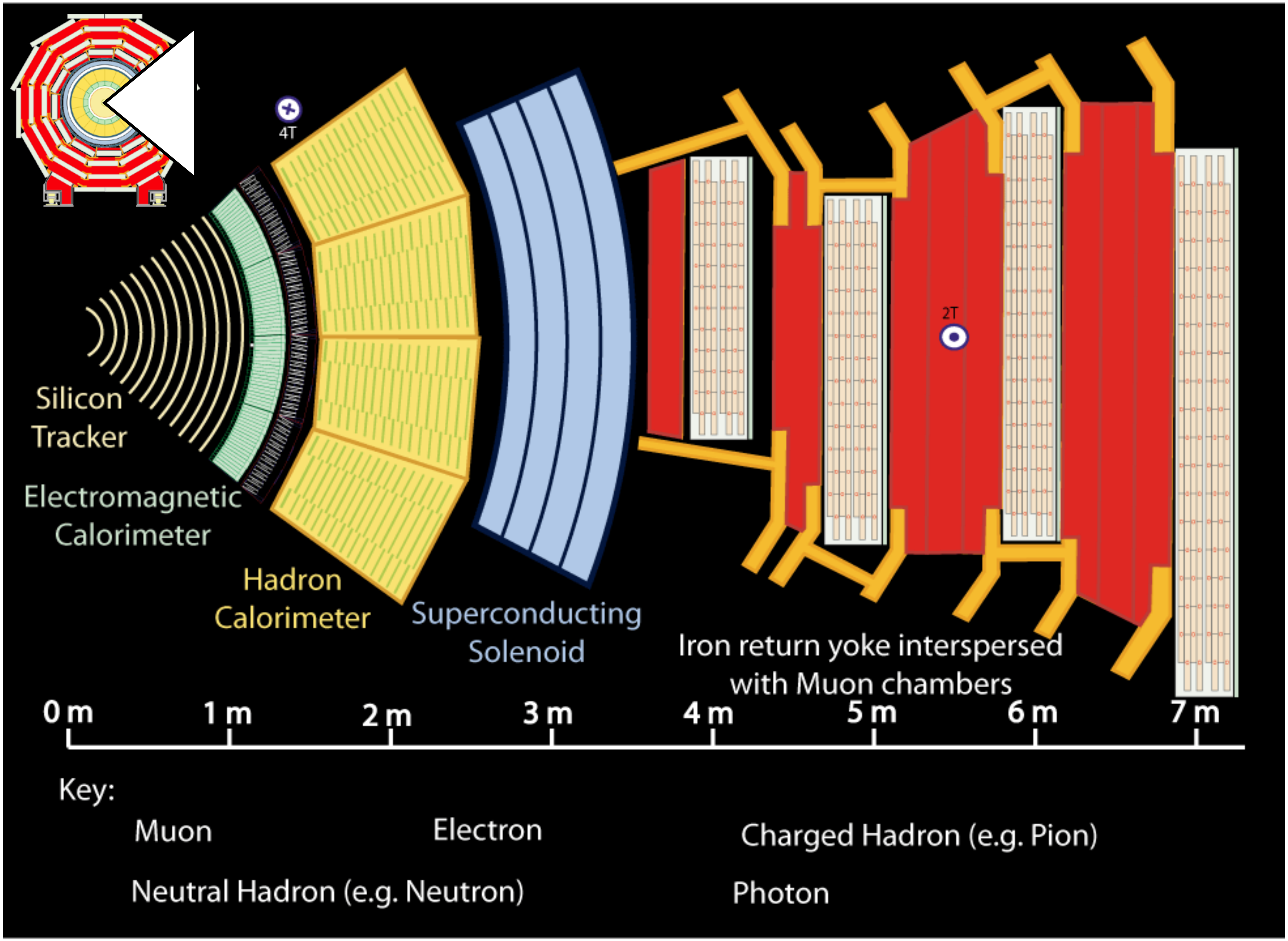
Machine learning for particle flow at CMS



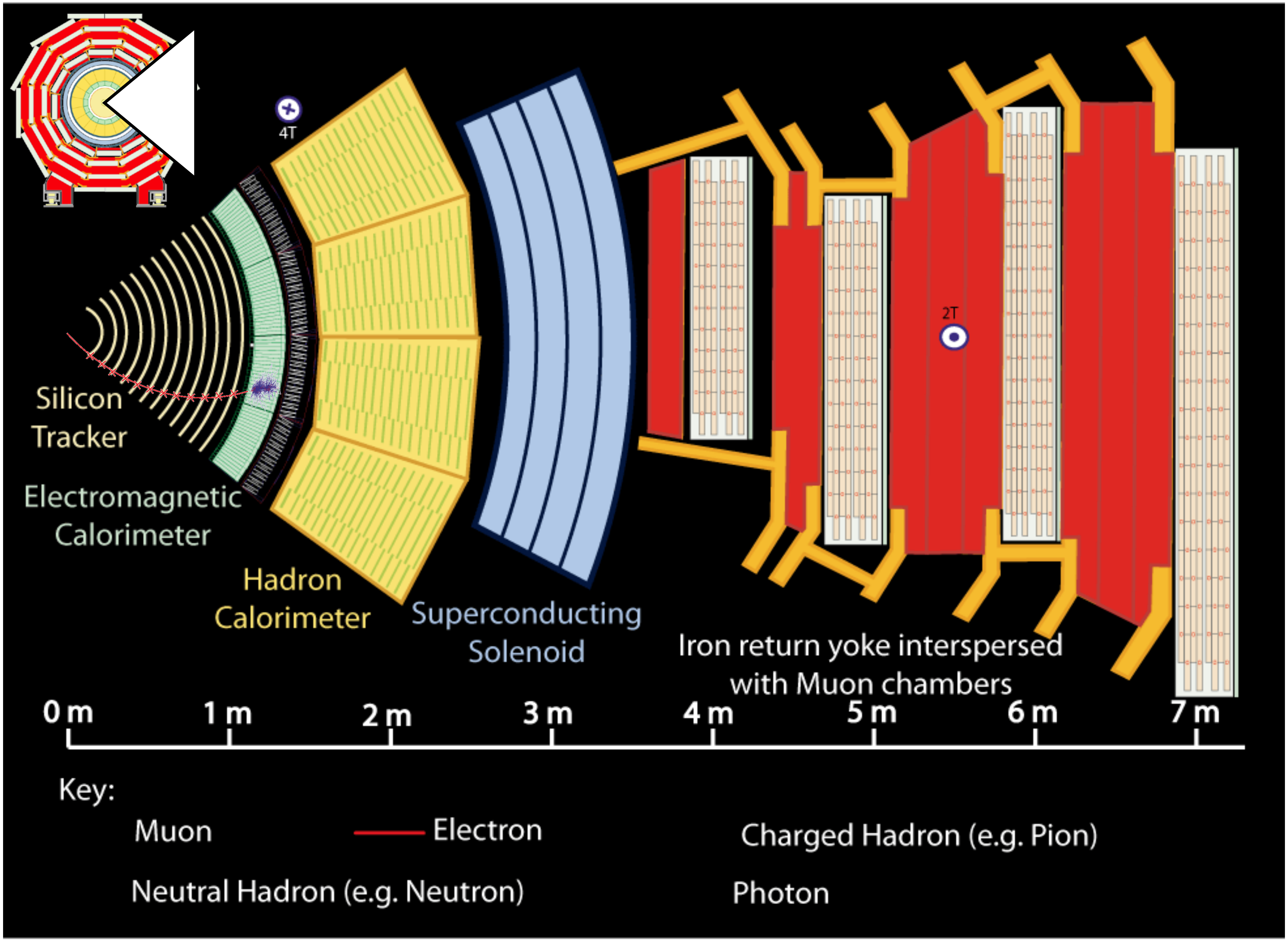
Dylan Rankin [MIT] *on behalf of the CMS Collaboration*

November 4th, 2022

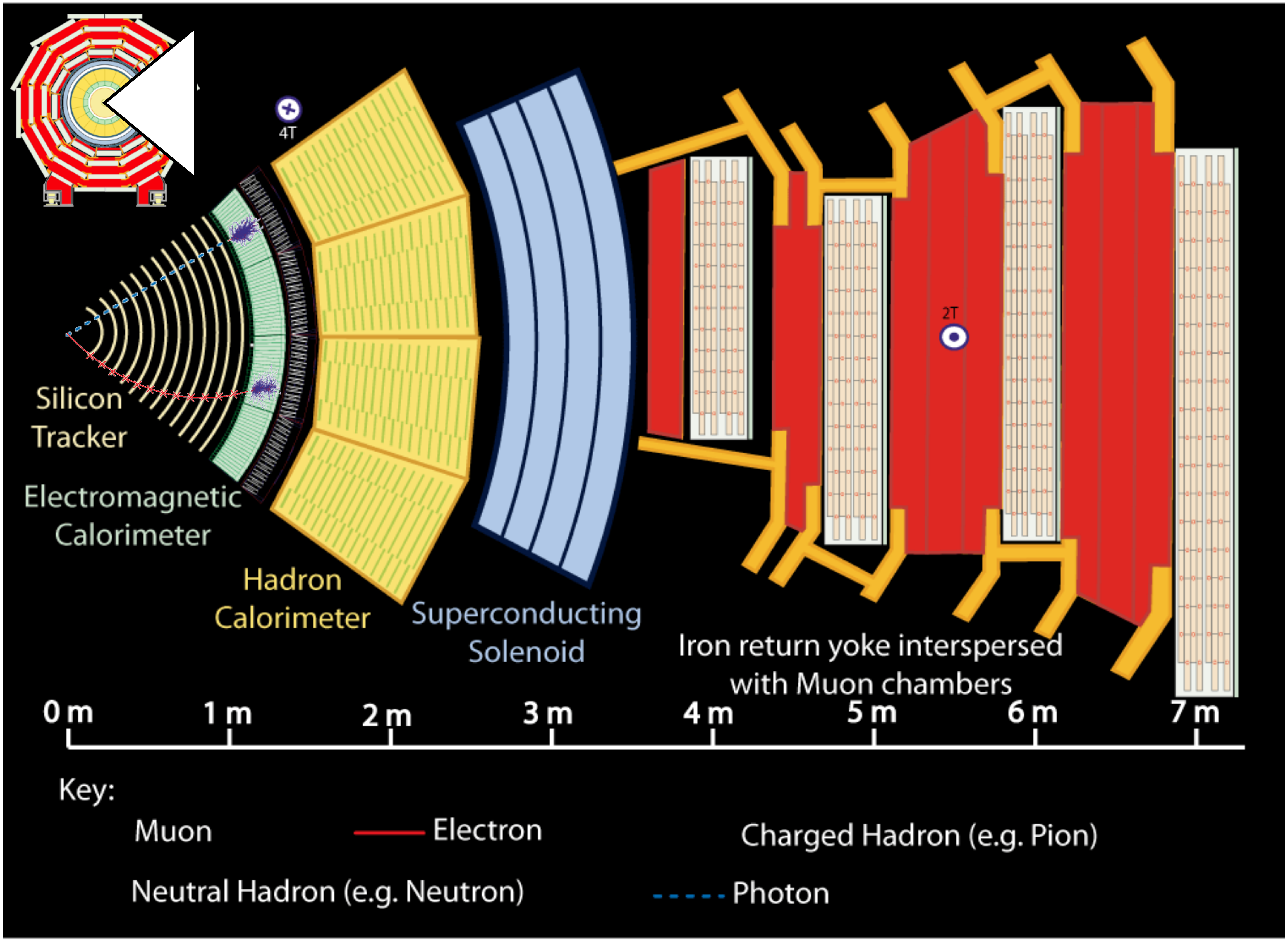
CMS


4T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
2T
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

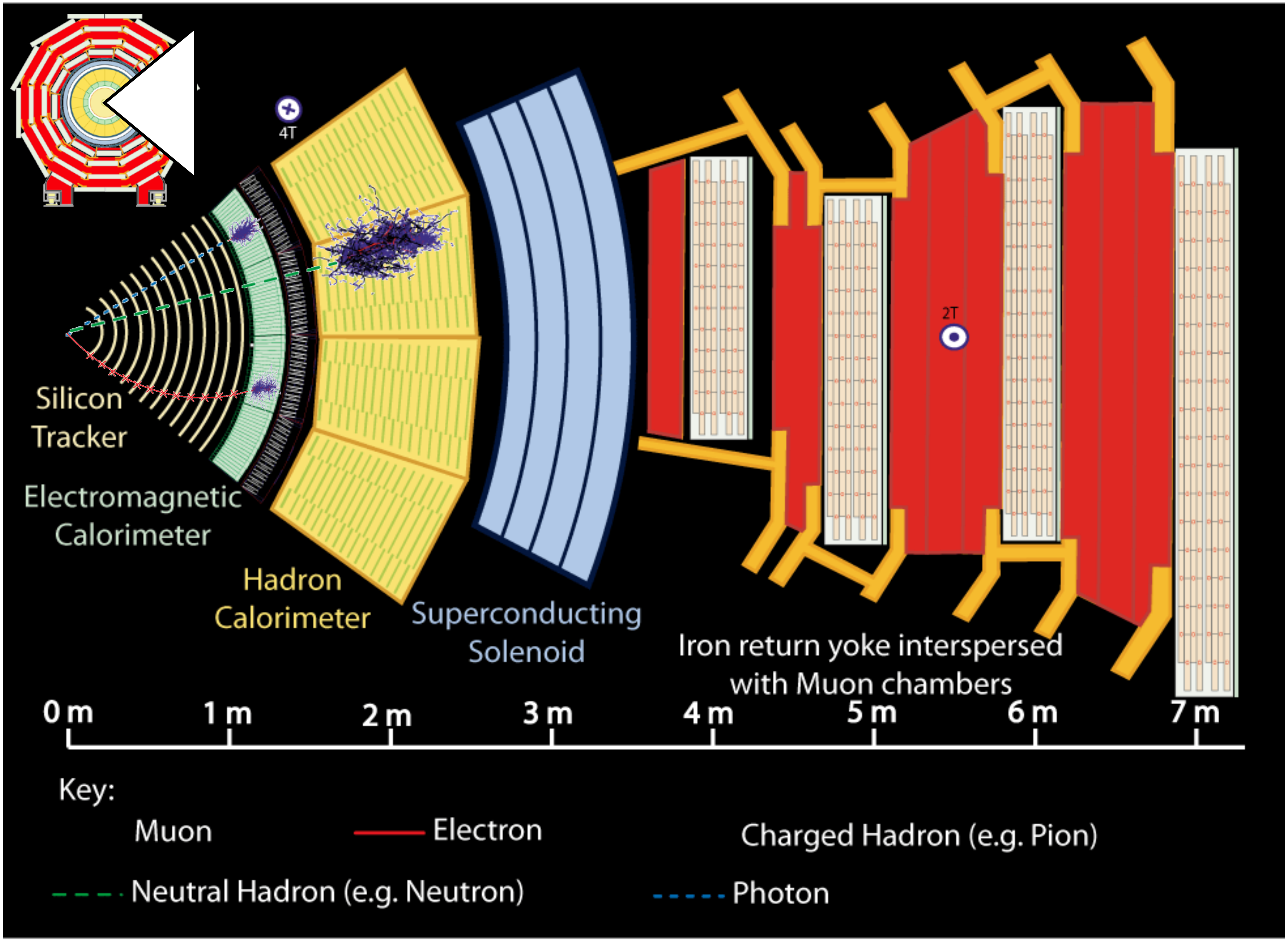
CMS


4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

CMS

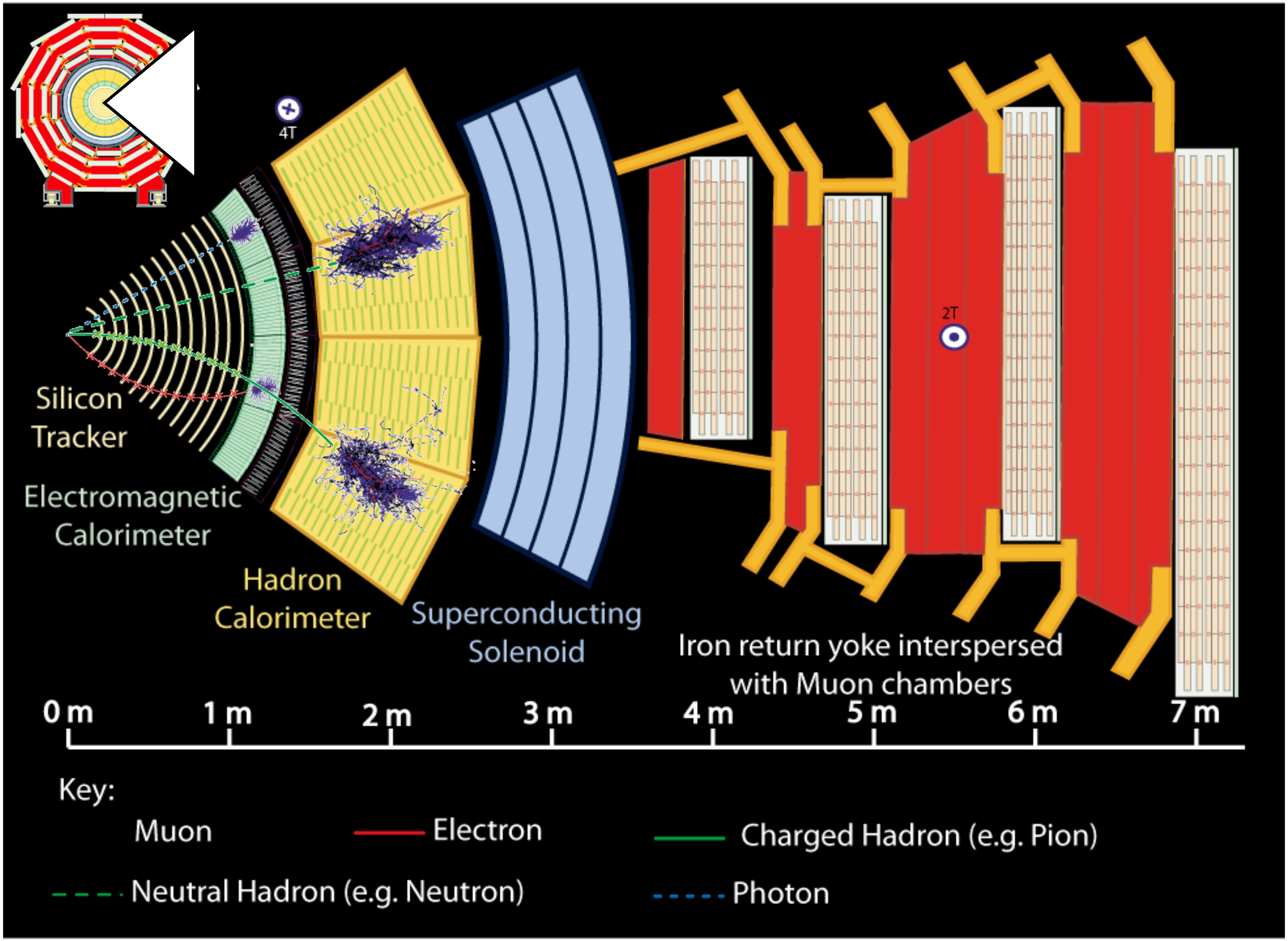

4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

CMS

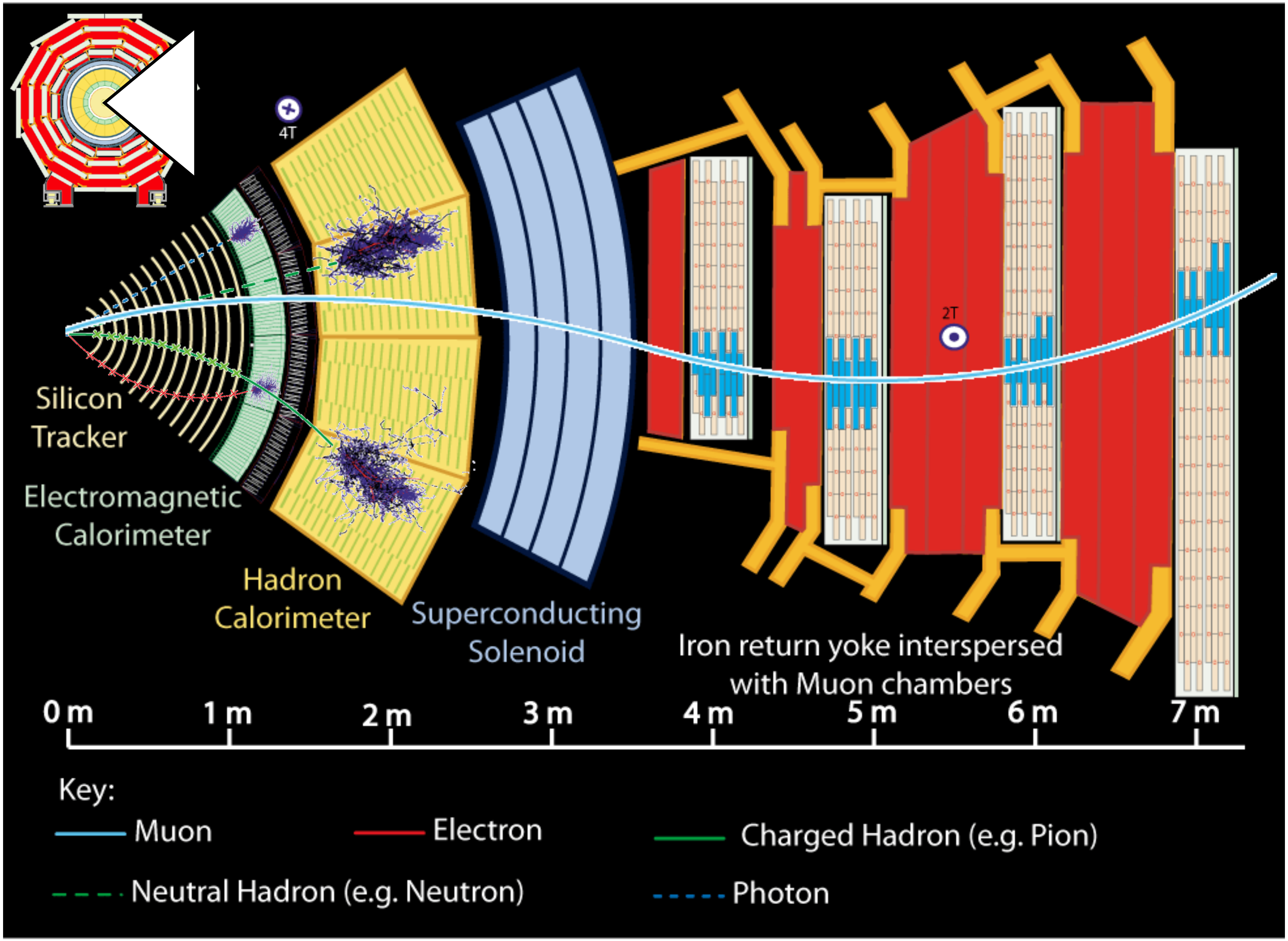


4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

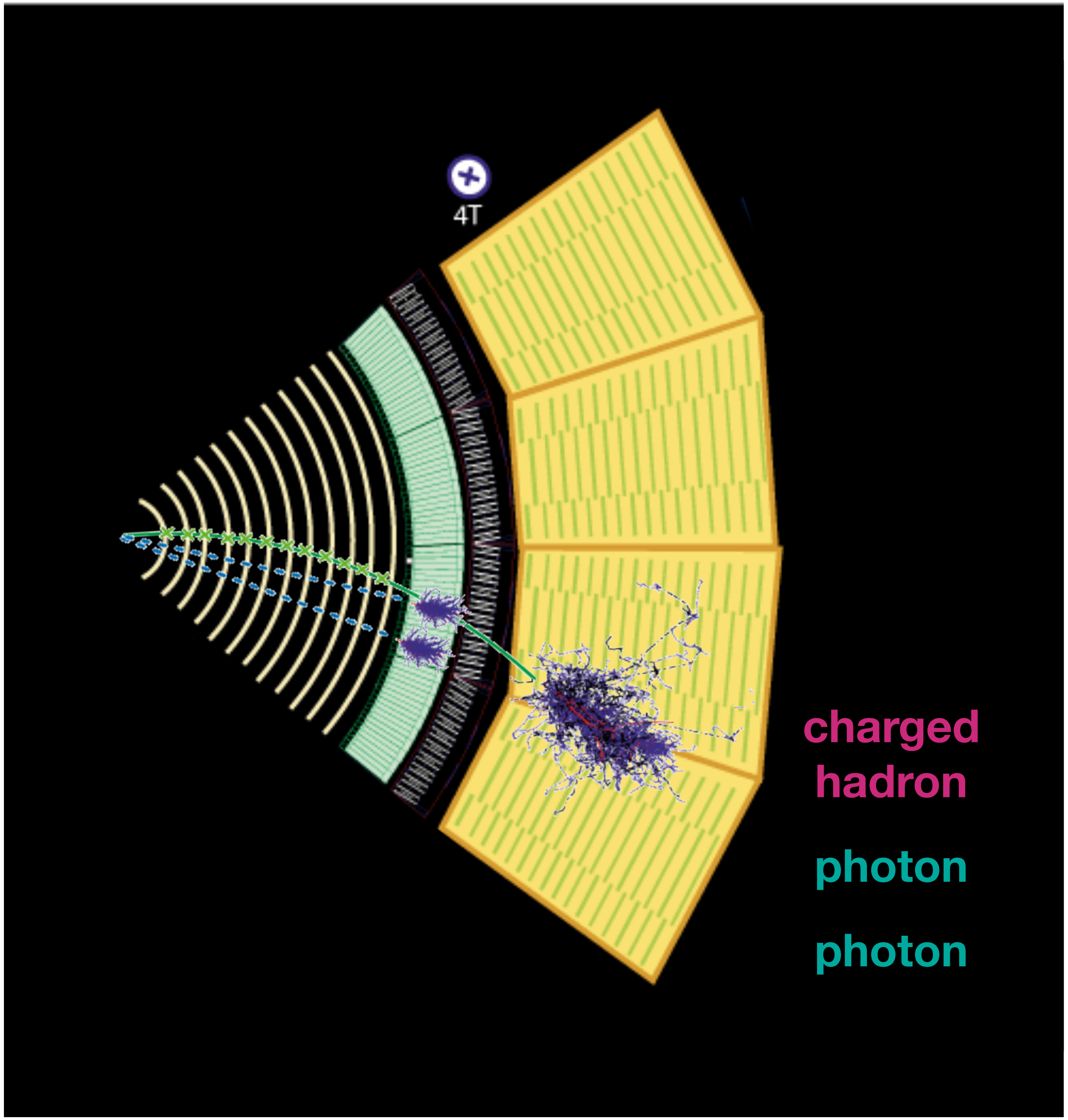
CMS


4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

CMS

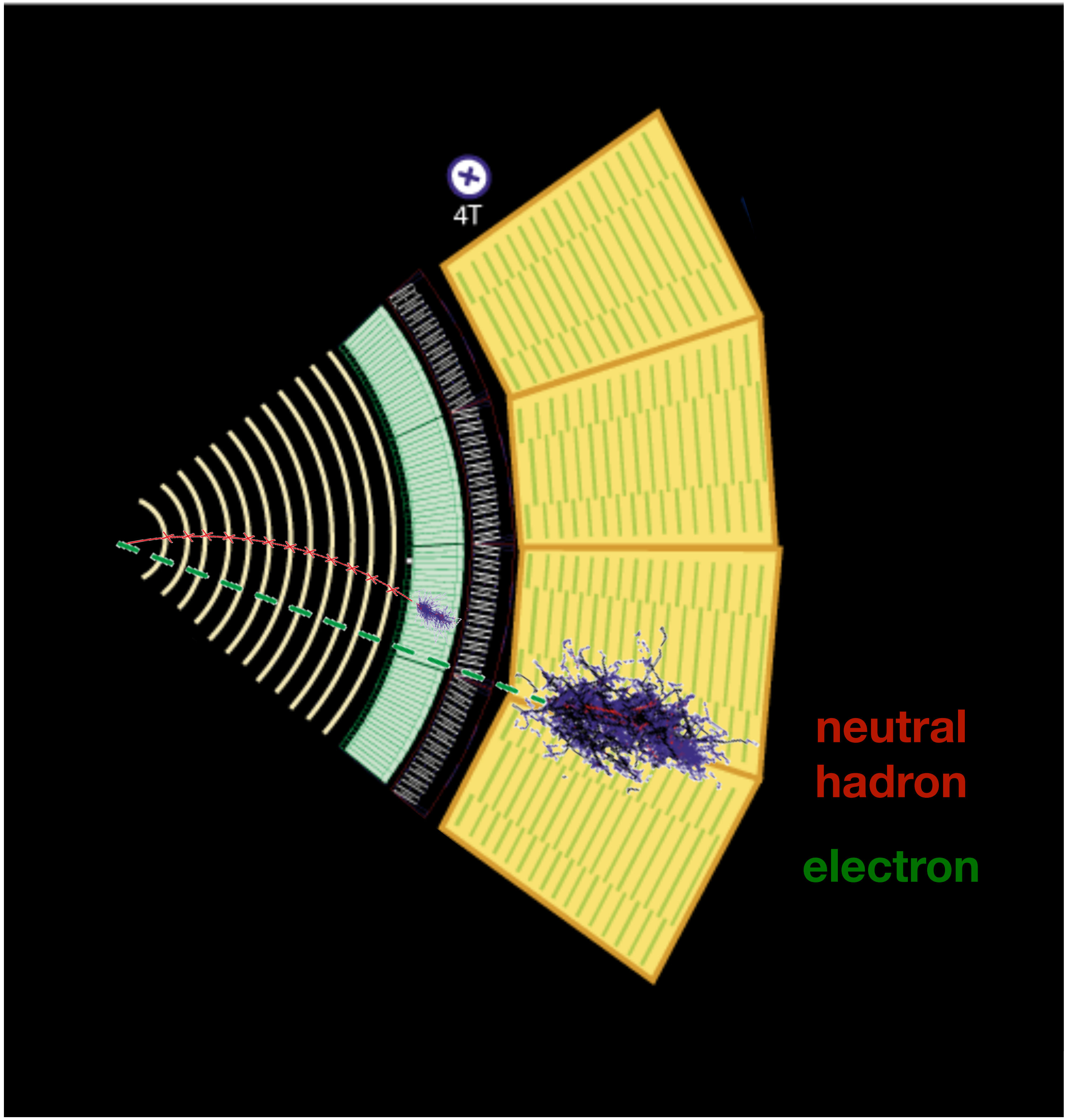

4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0 m
1 m
2 m
3 m
4 m
5 m
6 m
7 m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon

Particle Flow (PF)



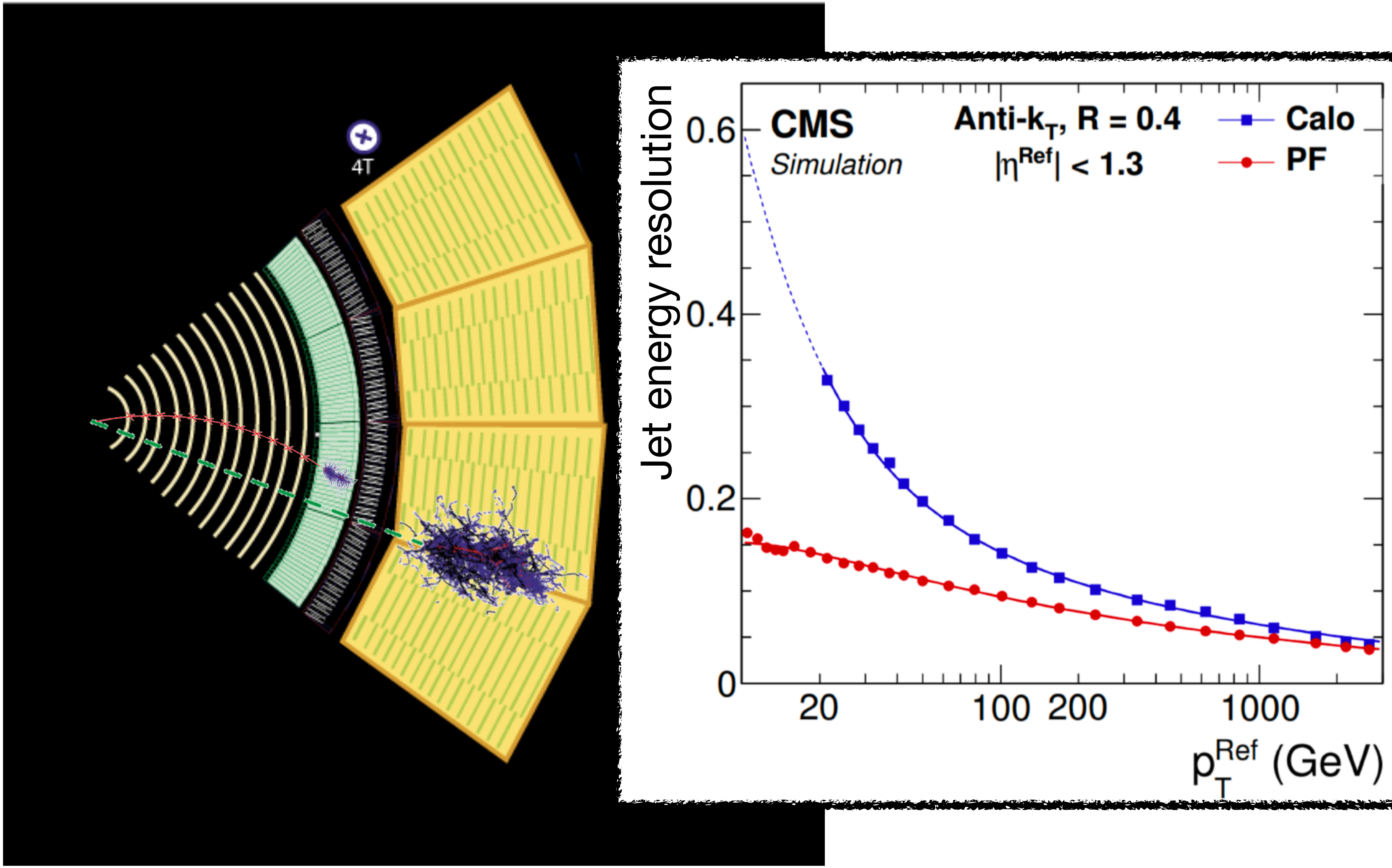
- Particle flow (PF) algorithm combines information from all subdetectors to reconstruct particles
  - ex. track + electromagnetic cluster + hadronic cluster = charged hadron (π^+) + photon (+ photon ?)
- Improved energy resolution

Particle Flow (PF)



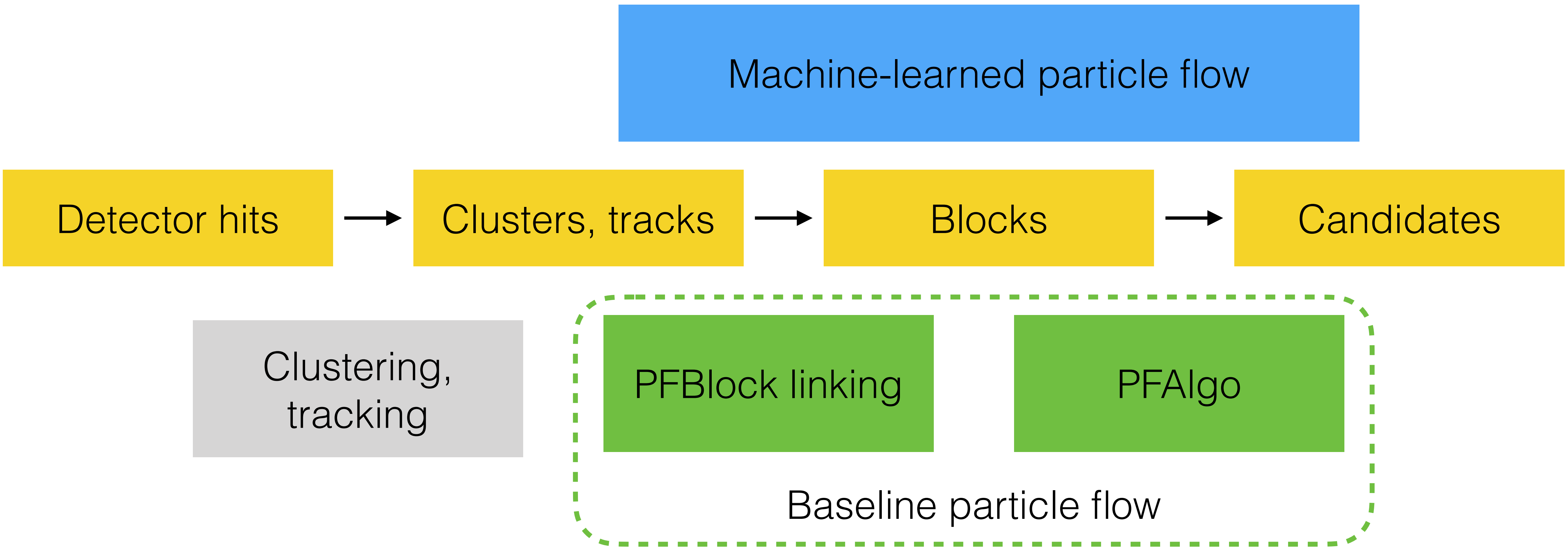
- Particle flow (PF) algorithm combines information from all subdetectors to reconstruct particles
    - ex. track + electromagnetic cluster + hadronic cluster = neutral hadron (K_L) + electron
- Improved energy resolution

Particle Flow (PF)



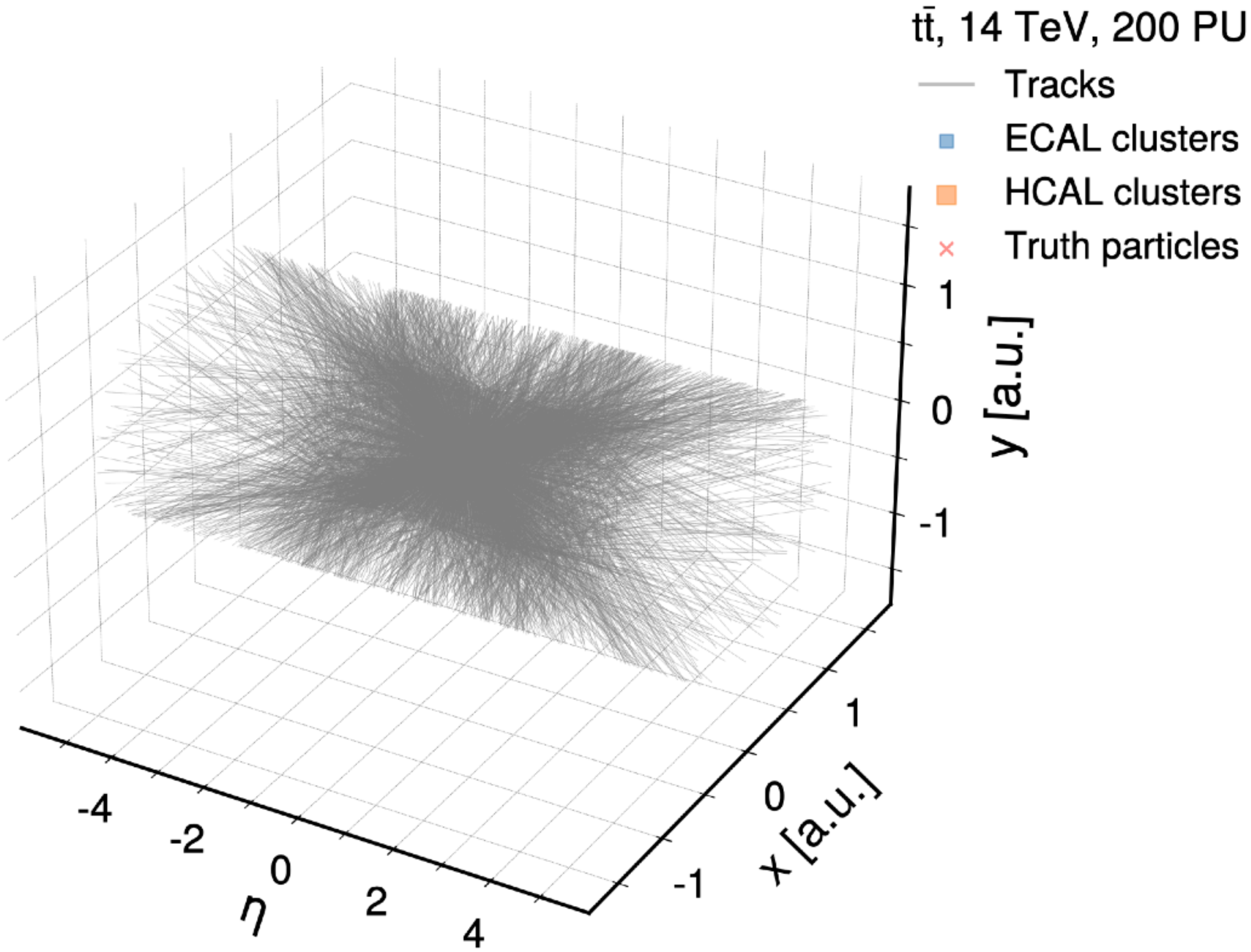
Reconstruction

- Particle flow starts from clusters & tracks (not raw hits), outputs particle candidates
- Could we replace this with an end-to-end ML algorithm?



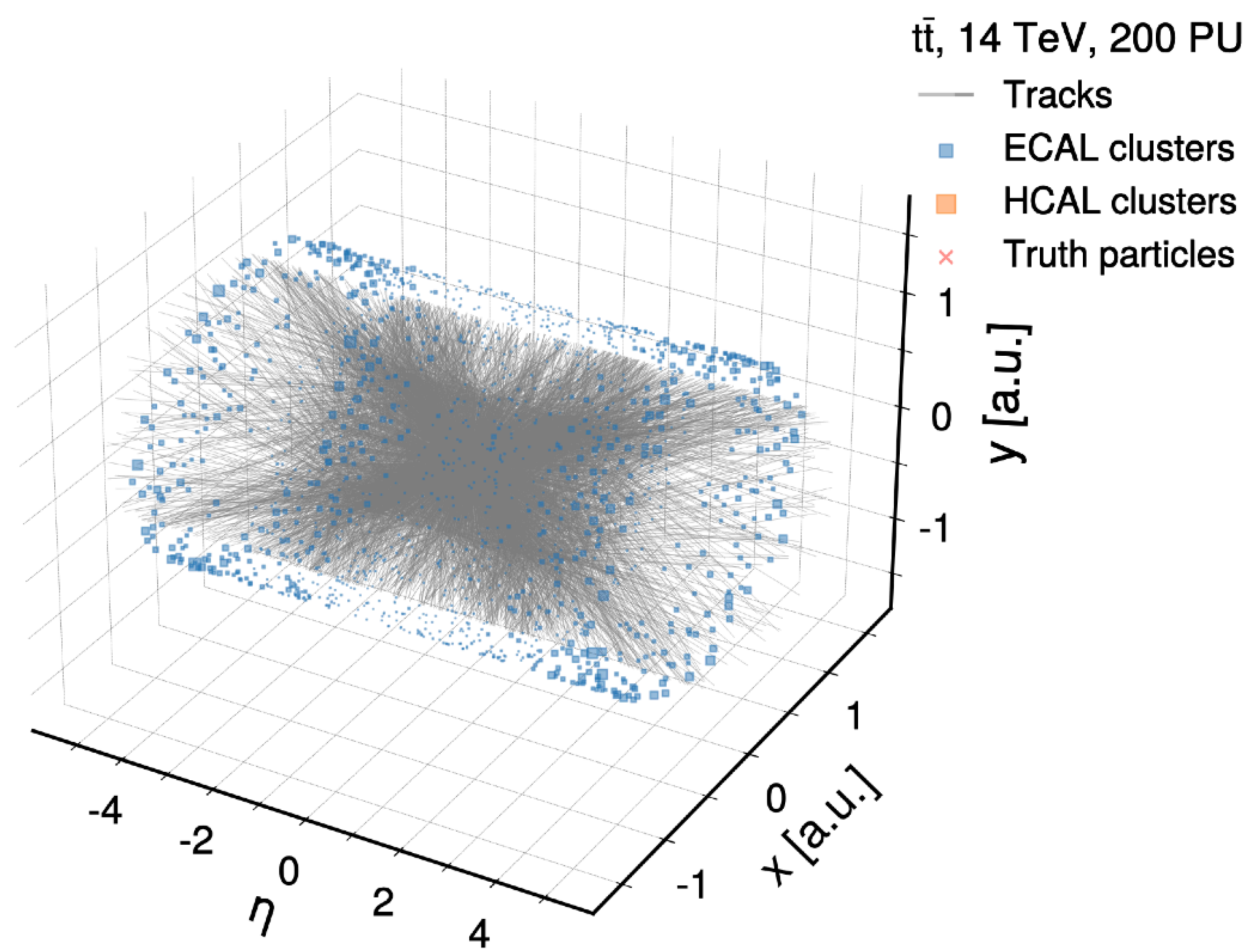
MLPF

- Inputs: **tracks (KF & GSF)**, **ECAL clusters (default & superclusters)**, **HCAL clusters**, **BREM points**
- Target set of **particles** $Y = \{y_i\}$
- Goal: construct a mapping $\mathcal{U}(X) = Y' \sim Y$ that minimizes some distance $||Y - Y'||$



MLPF

- Inputs: **tracks (KF & GSF)**, **ECAL clusters (default & superclusters)**, **HCAL clusters**, **BREM points**
- Target set of **particles** $Y = \{y_i\}$
- Goal: construct a mapping $\mathscr{U}(X) = Y' \sim Y$ that minimizes some distance $||Y - Y'||$

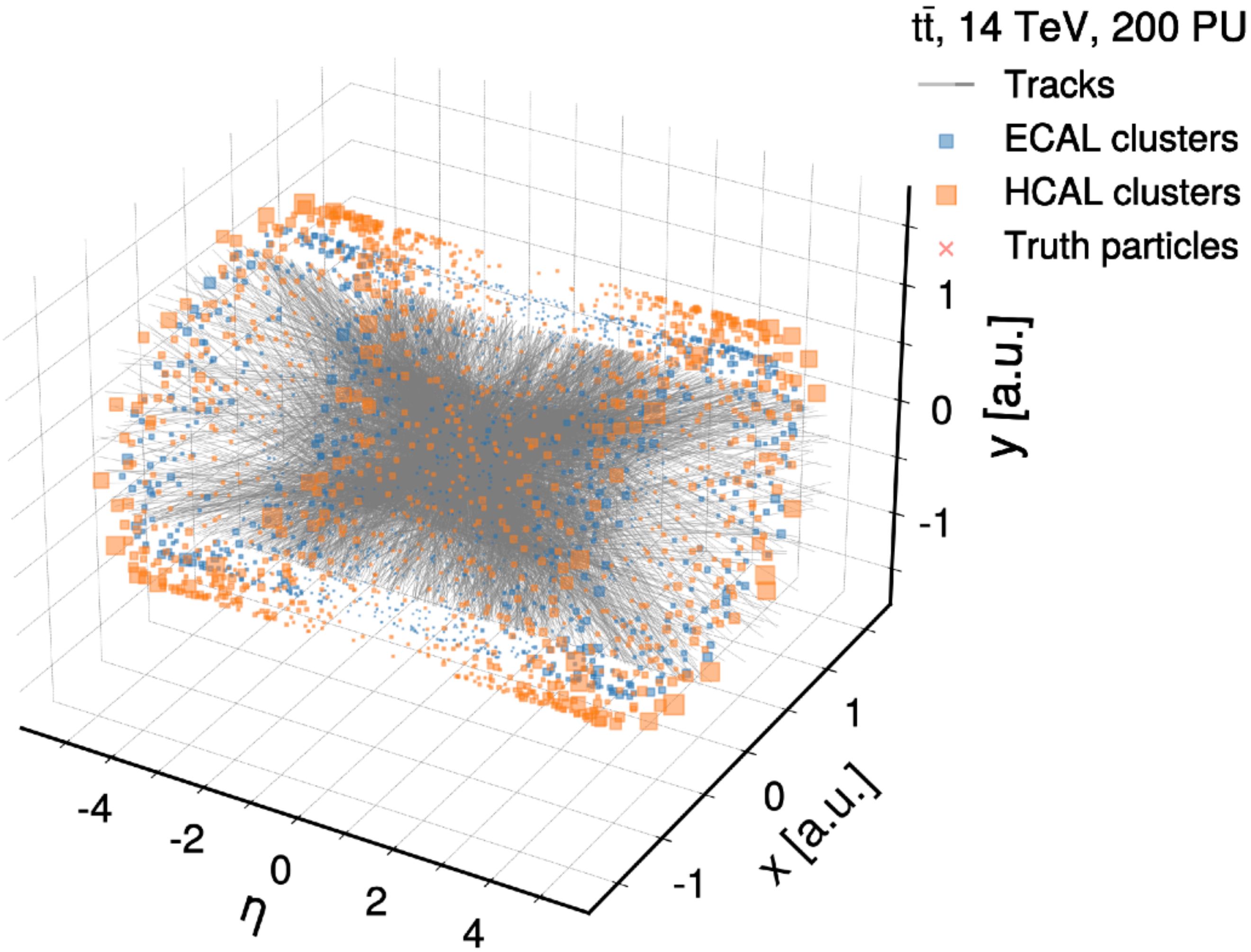


MLPF

- Inputs: **tracks (KF & GSF)**, **ECAL clusters (default & superclusters)**, **HCAL clusters**, **BREM points**

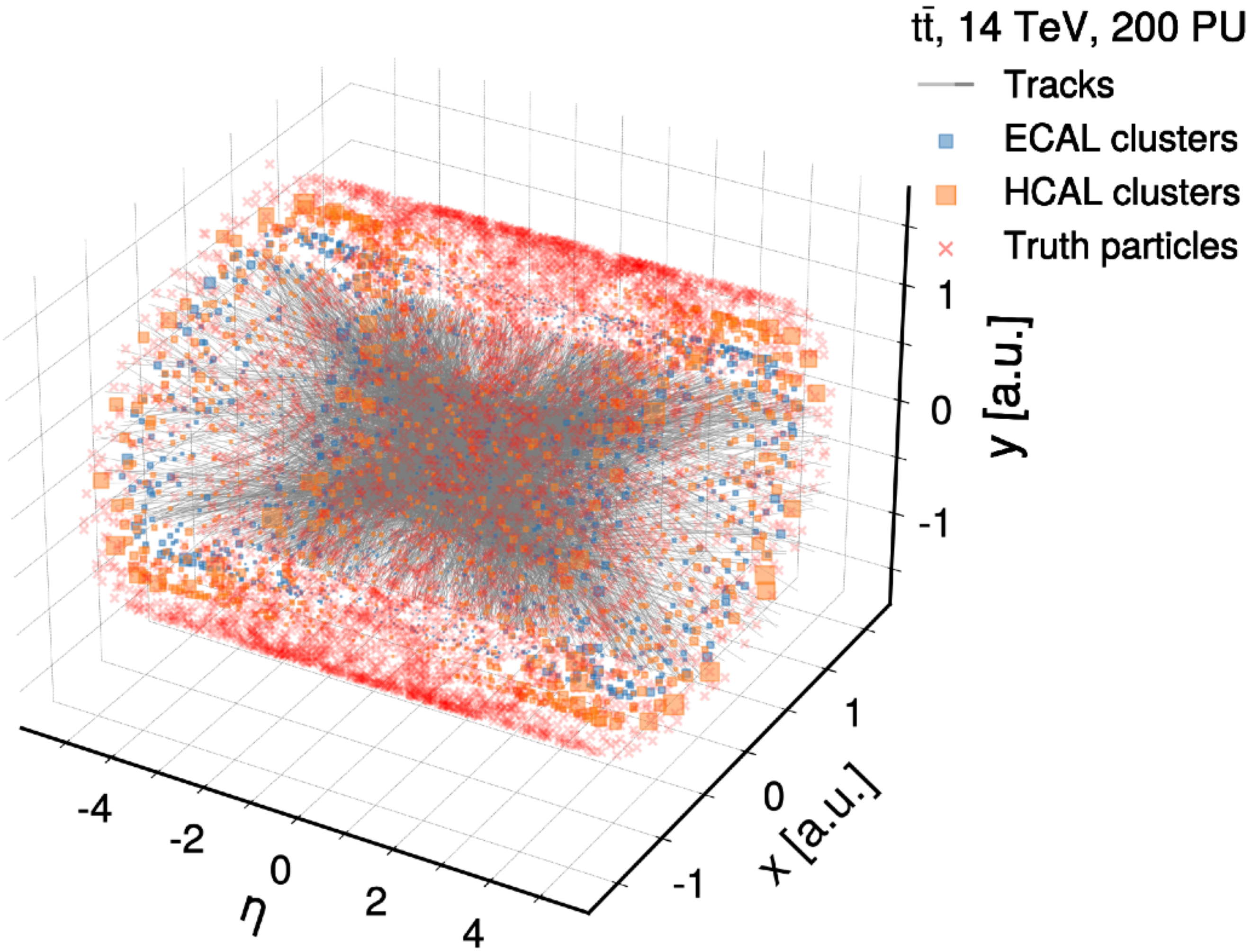
- Target set of **particles**
  $Y = \{y_i\}$

- Goal: construct a mapping $\mathscr{U}(X) = Y' \sim Y$ that minimizes some distance $||Y - Y'||$



MLPF

- Inputs: **tracks (KF & GSF)**, **ECAL clusters (default & superclusters)**, **HCAL clusters**, **BREM points**
- Target set of **particles** $Y = \{y_i\}$
- Goal: construct a mapping $\mathcal{U}(X) = Y' \sim Y$ that minimizes some distance $||Y - Y'||$



Training

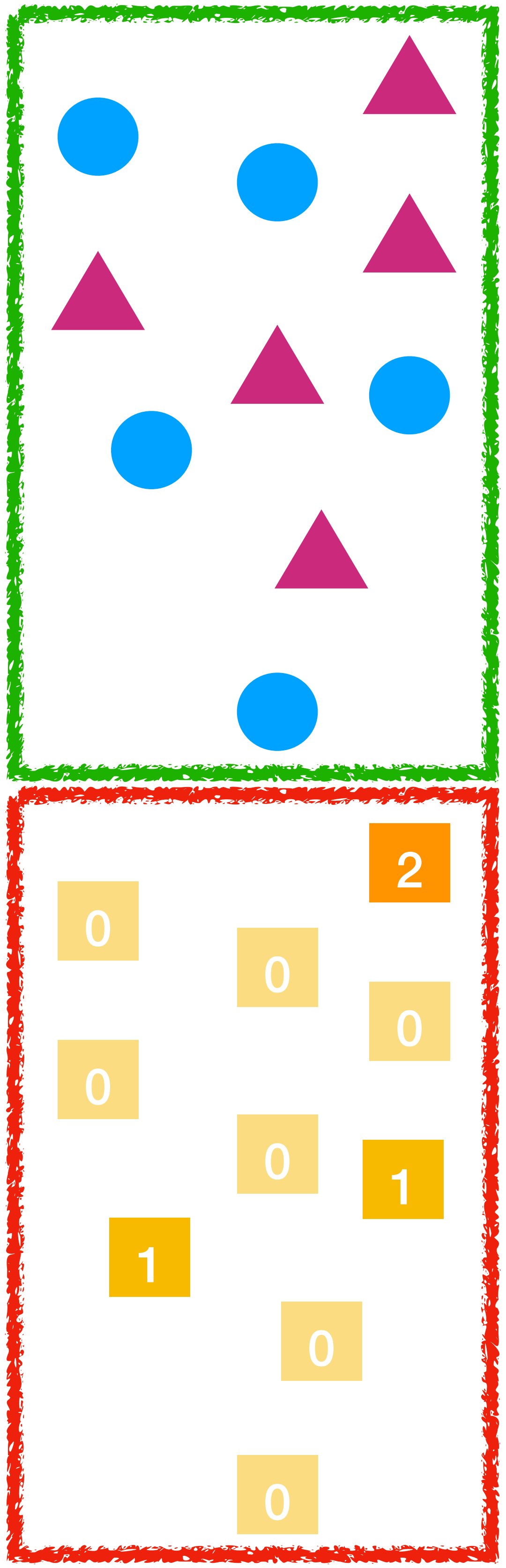
- $\mathscr{L} = \sum_{i \in \text{event}} L(y_i, y_i'),\ L(y_i, y_i') \equiv CLS(c_i, c_i') + \alpha\ REG(p_i, p_i')$
  - Separate terms for classification (CLS) and regression (REG)
- Focal loss used for classification
  - $\text{FL}(p_t) = -(1 - p_t)^\gamma\ \log(p_t)$
- Huber loss used for regression
  - $\text{HL}(y, y') = \begin{cases} \frac{1}{2}(y - y')^2, & \text{for } |y - y'| \le \delta \\ \delta \cdot (|y - y'| - \frac{1}{2}\delta), & \text{otherwise} \end{cases}$

Training

- Use object condensation [1] approach:
  - Zero-pad target set Y such that $|Y| = |X|$
- Allows loss to handle arbitrary event sizes
- In addition to particle classes also allow 0 class
  - Apply threshold on 0 class to remove extra particles

Input set
$X = \{x_i\}$

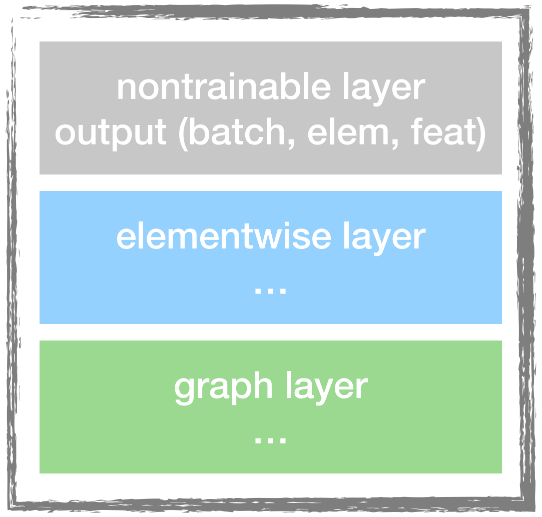
Target set
$Y = \{y_i\}$

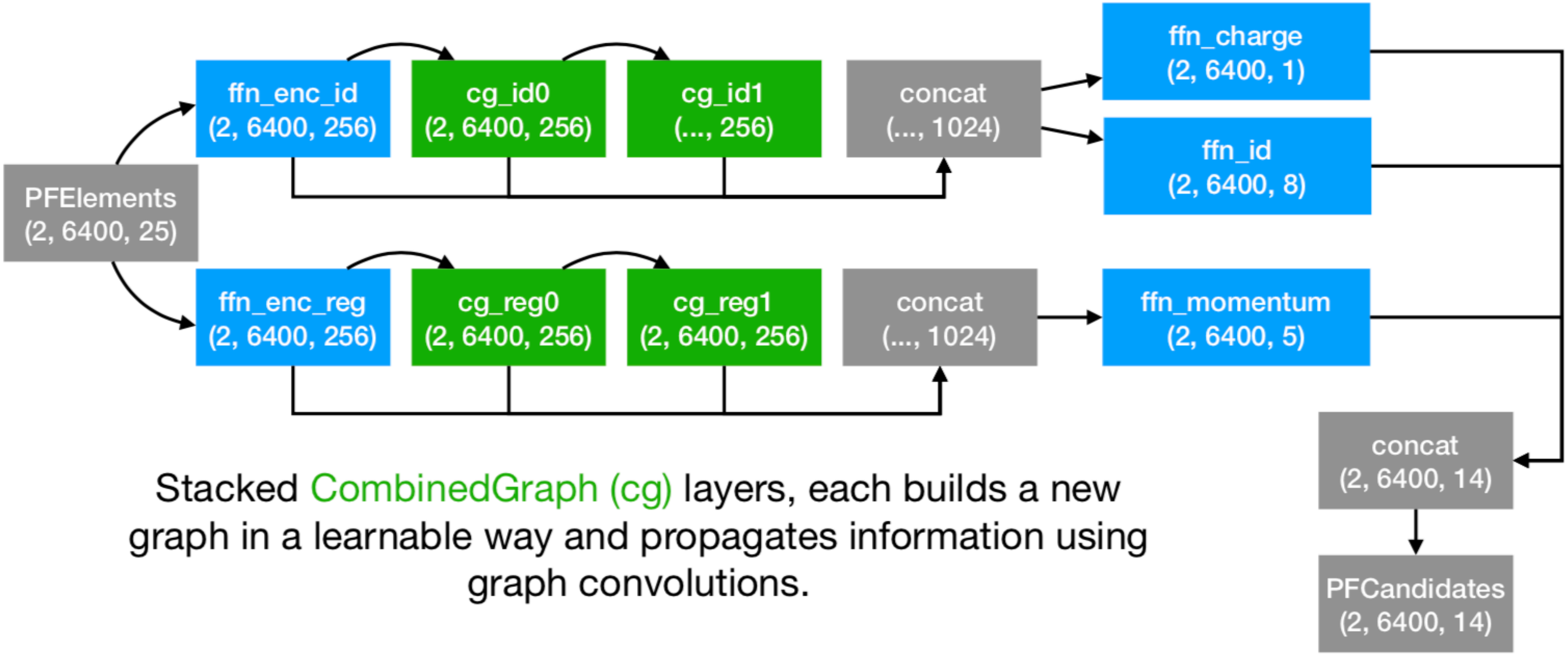


[1] 2002.03605

Network Architecture

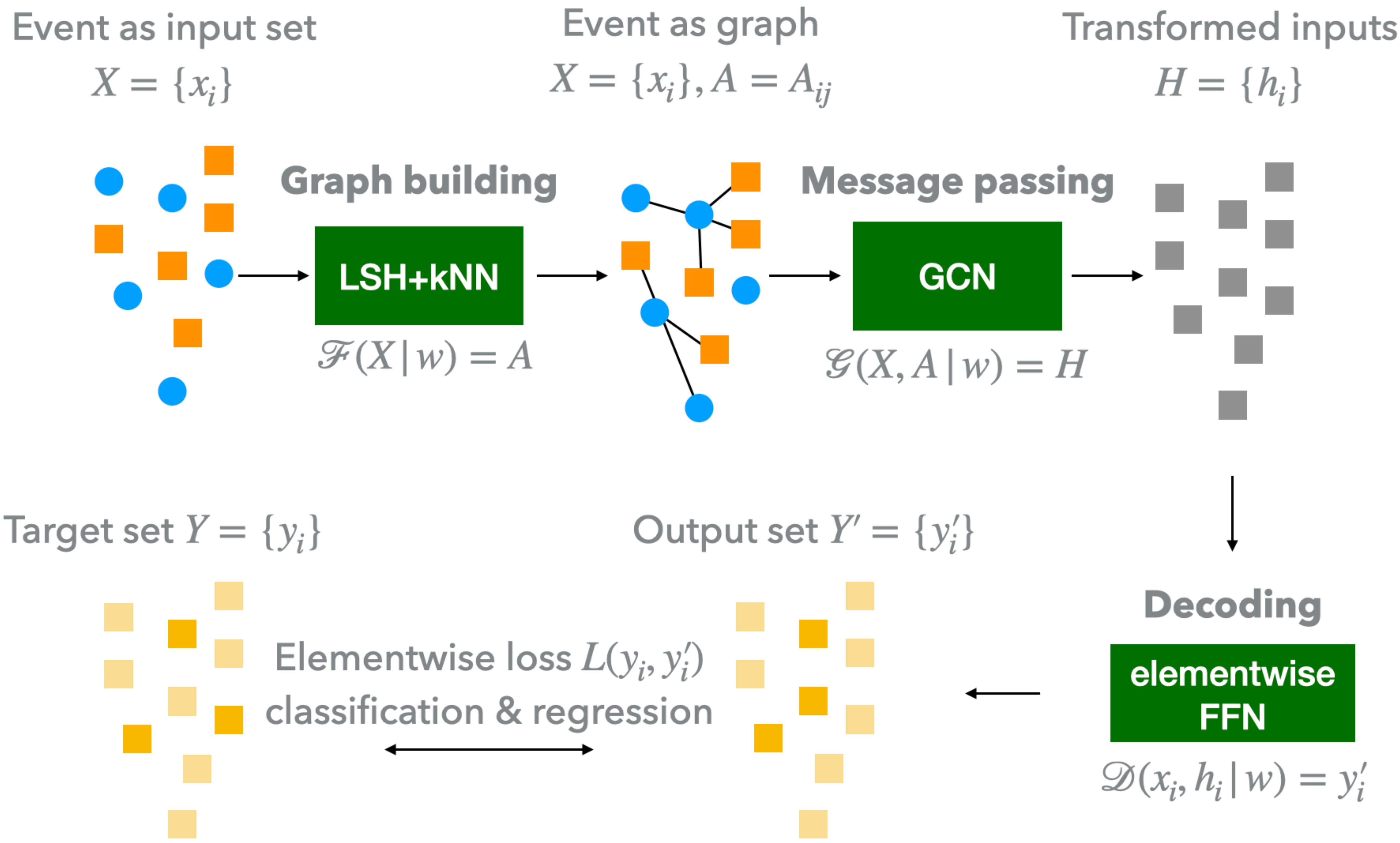
- Graph neural network-based architecture
  - Graph construction performed in local neighborhoods to improve scalability → no N^2 allocation/computations





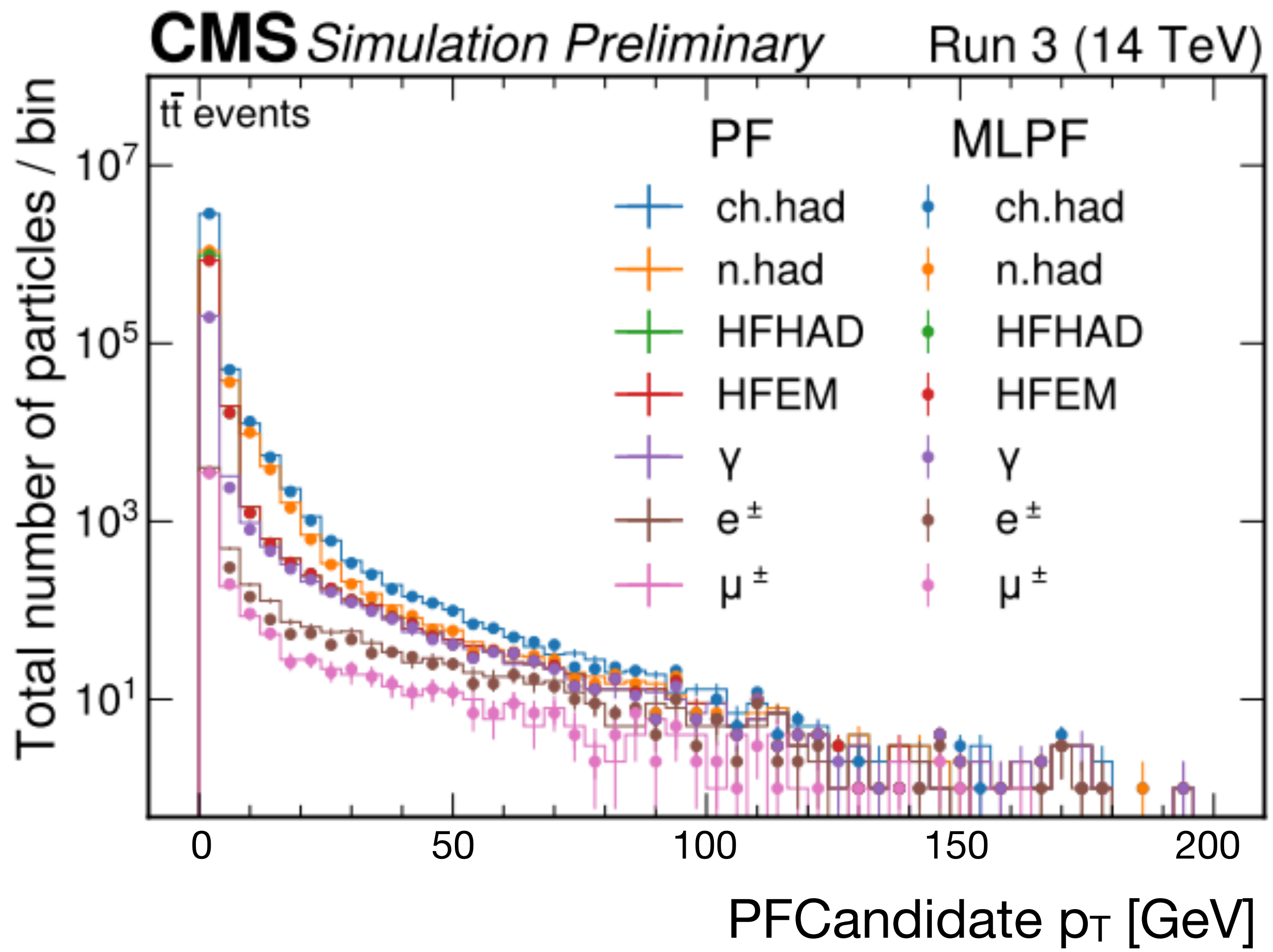
Stacked CombinedGraph (cg) layers, each builds a new graph in a learnable way and propagates information using graph convolutions.

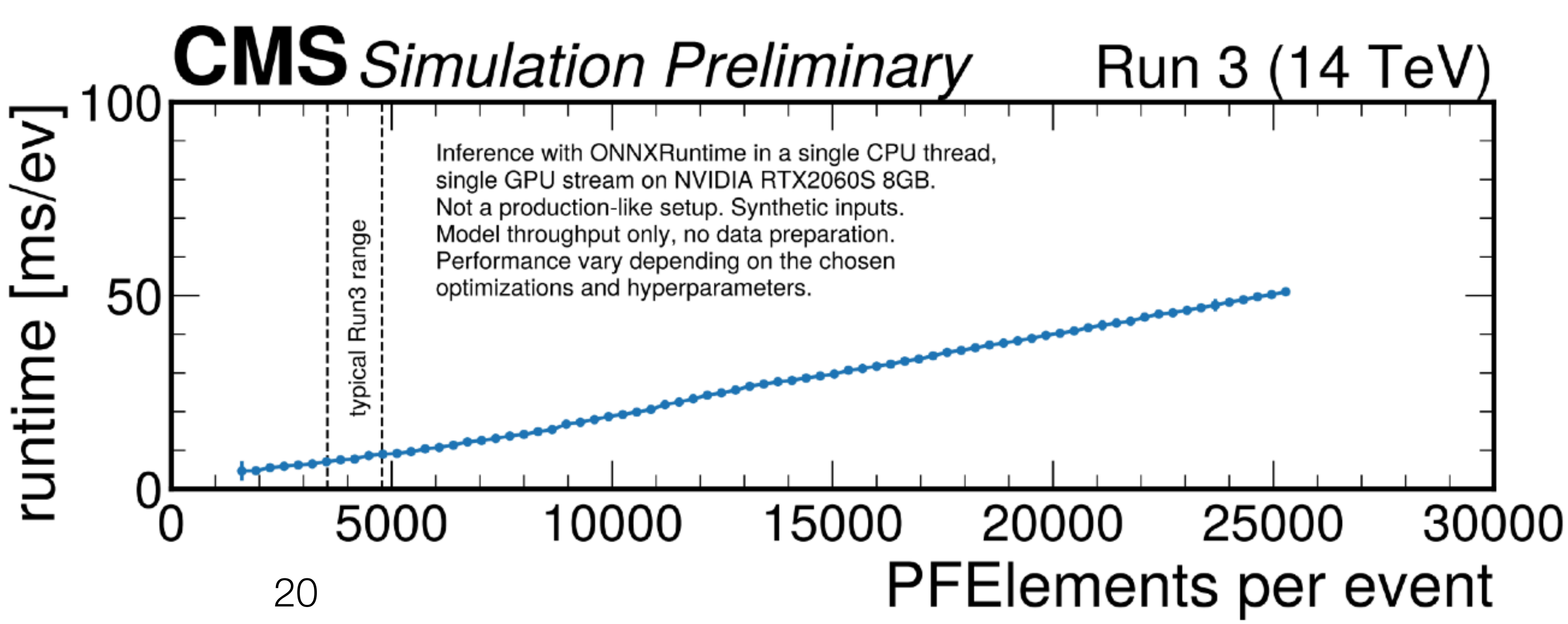
Network Architecture



MLPF v1

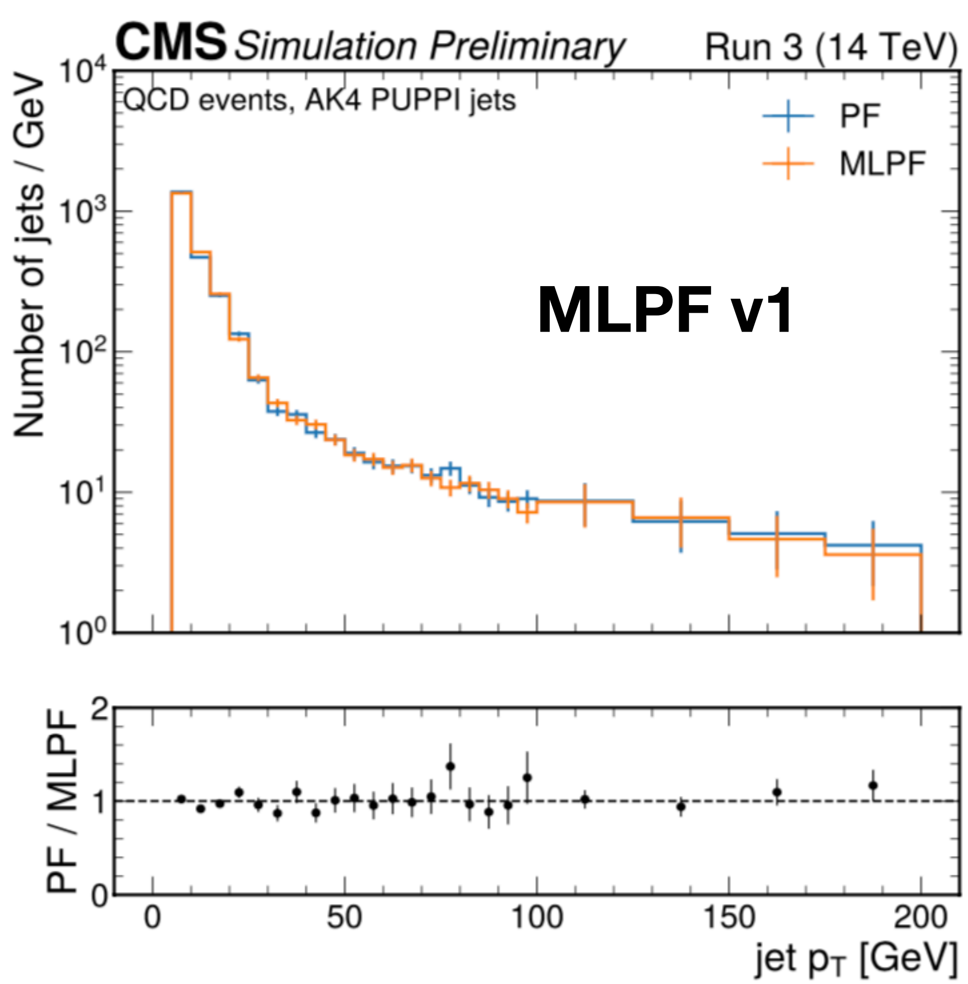
- First version trained using PF as target
  - Can't exceed PF performance, but useful proof-of-concept
- Very promising results (both for physics performance and computational scaling)

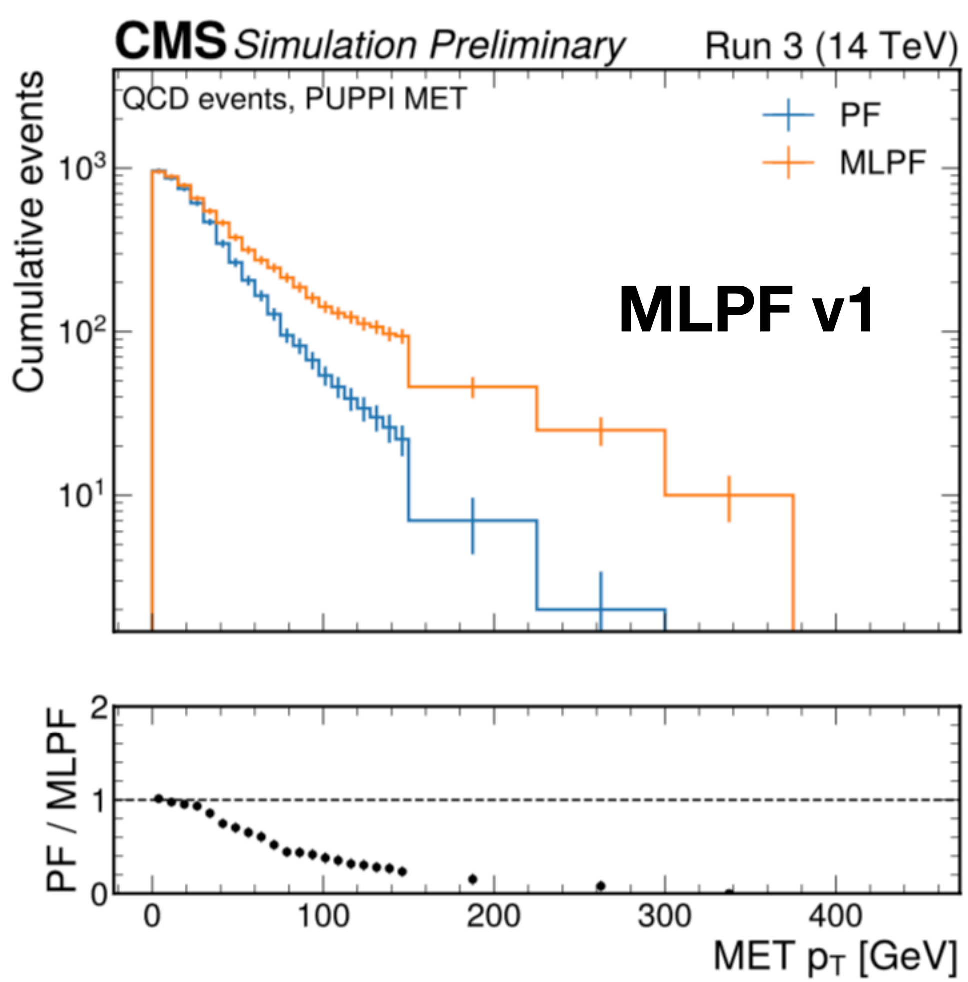




MLPF v2

- How could MLPF improve on standard PF algorithm?
  - Train with truth particles as target
  - Additional terms in loss for physics quantities (eg. jet/MET response)

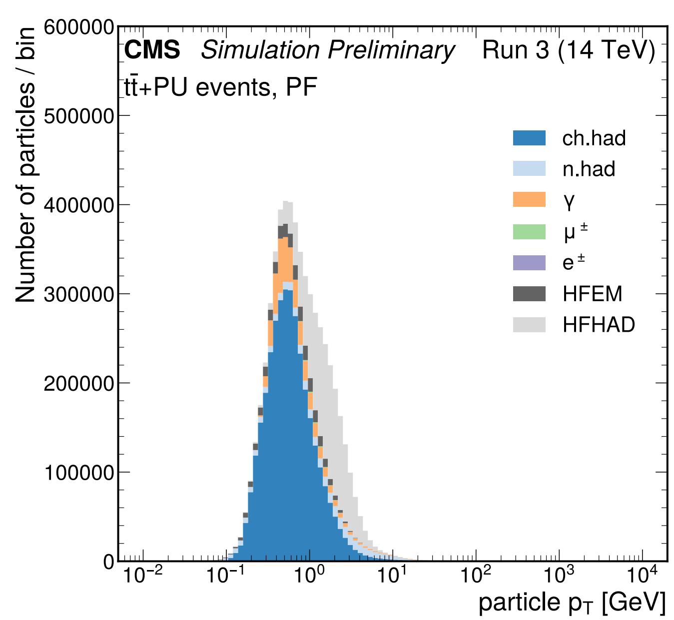


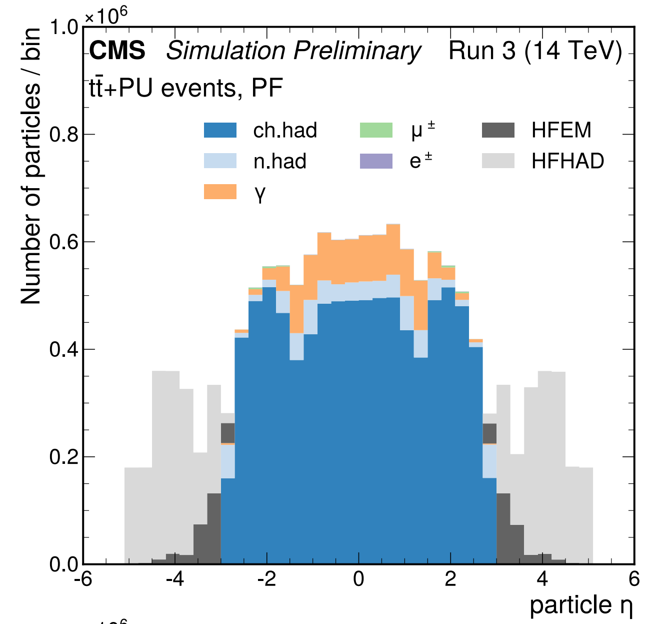


Samples

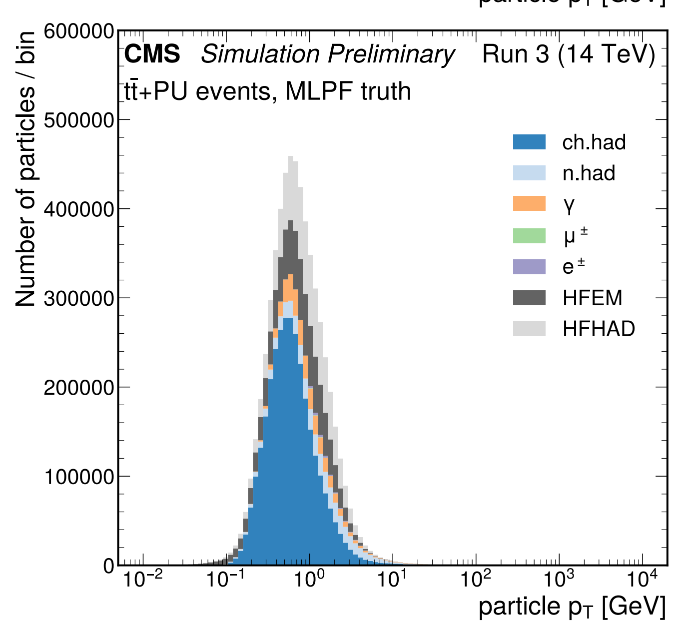
- Mix of physics samples and particle gun, range of PU configurations
  - Run 3 conditions, ~500k events in total

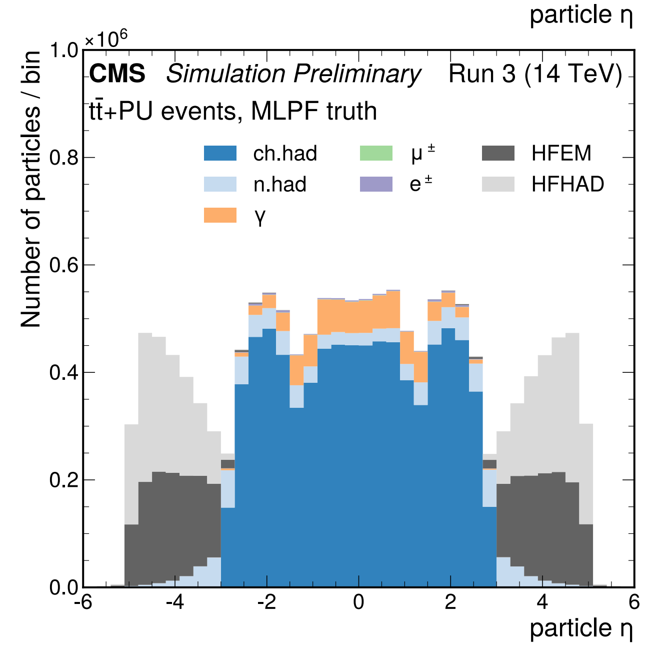
PF





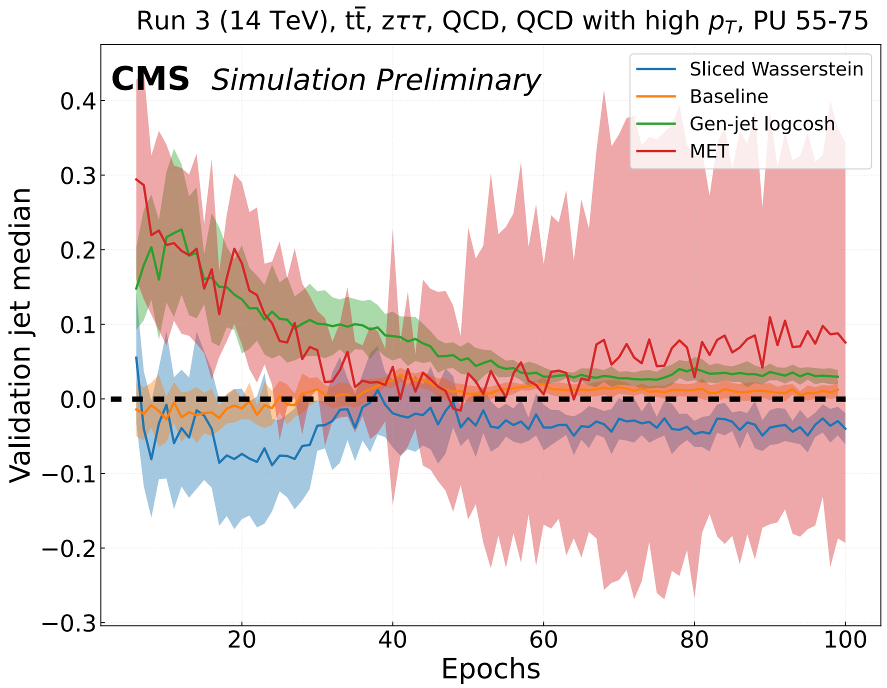
truth

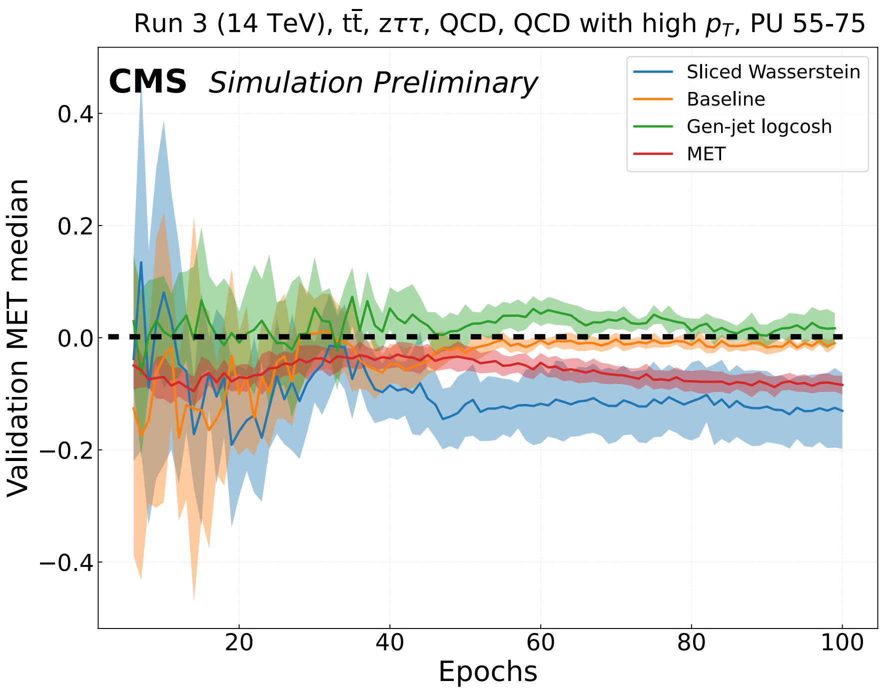


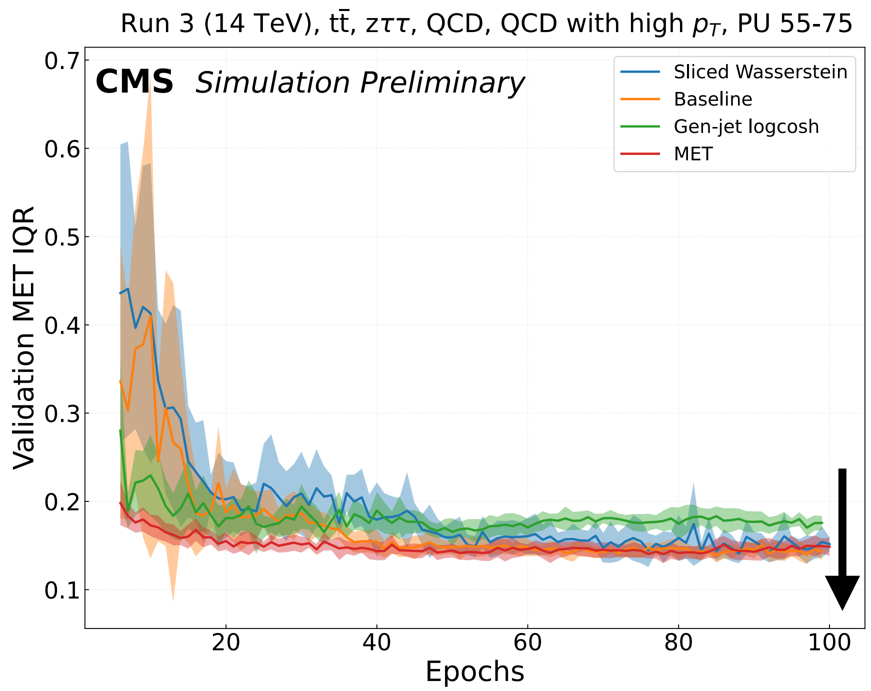


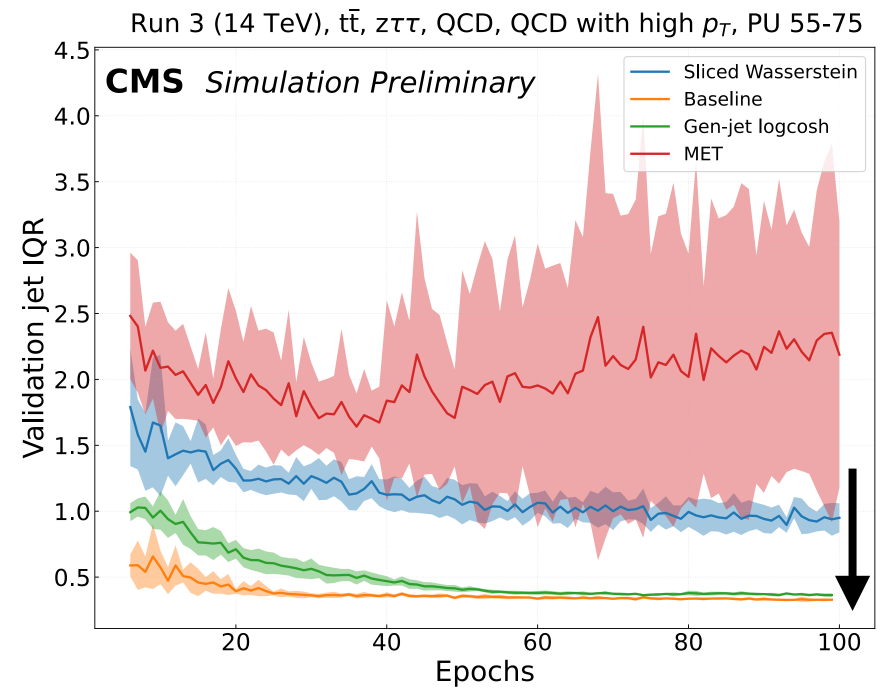
Optimization

- Multiple variations on standard loss studied
  - Attempt to target **jets**, **MET**, **local particle densities**
- **Baseline loss** appears to still perform best overall

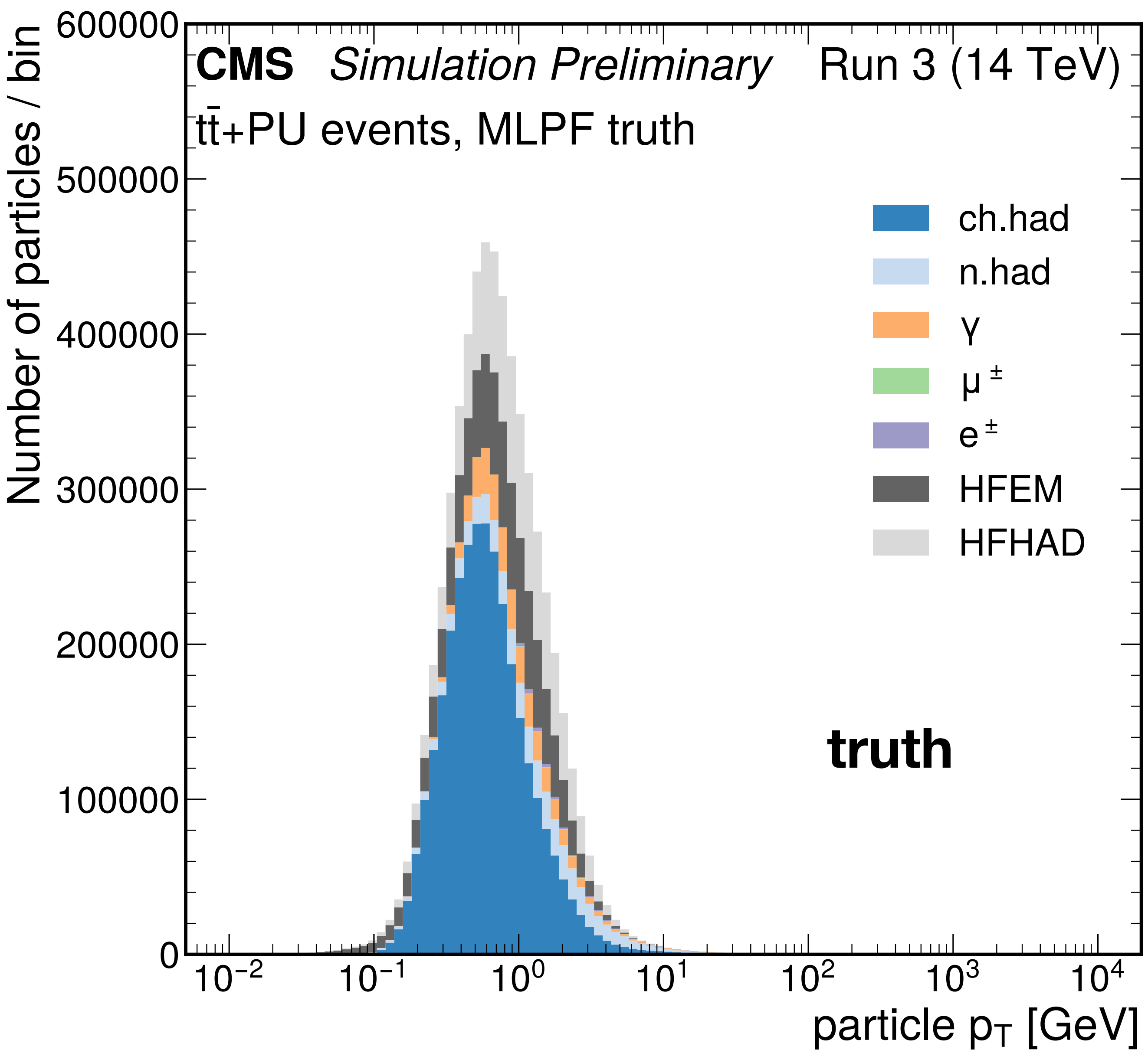


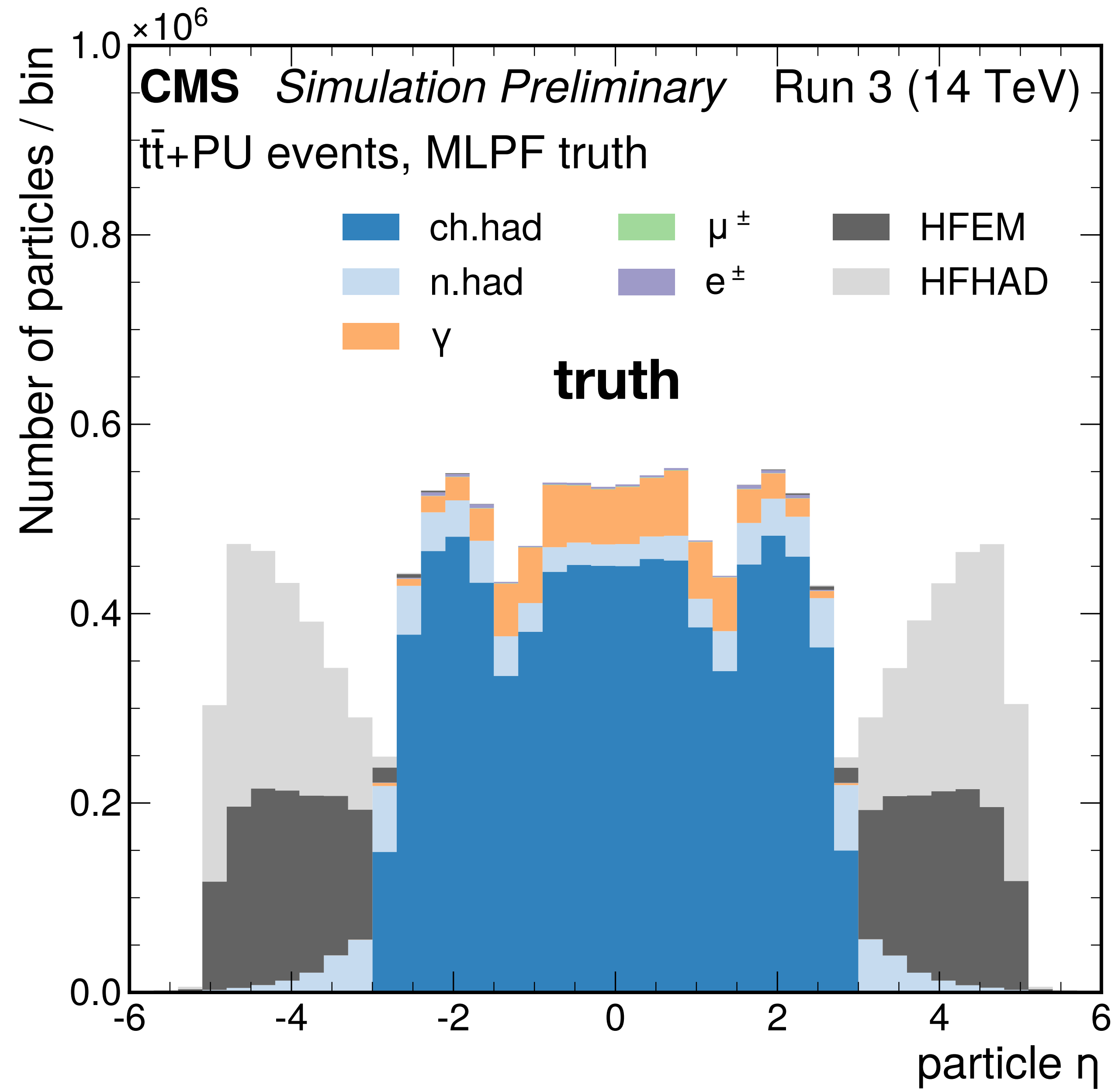






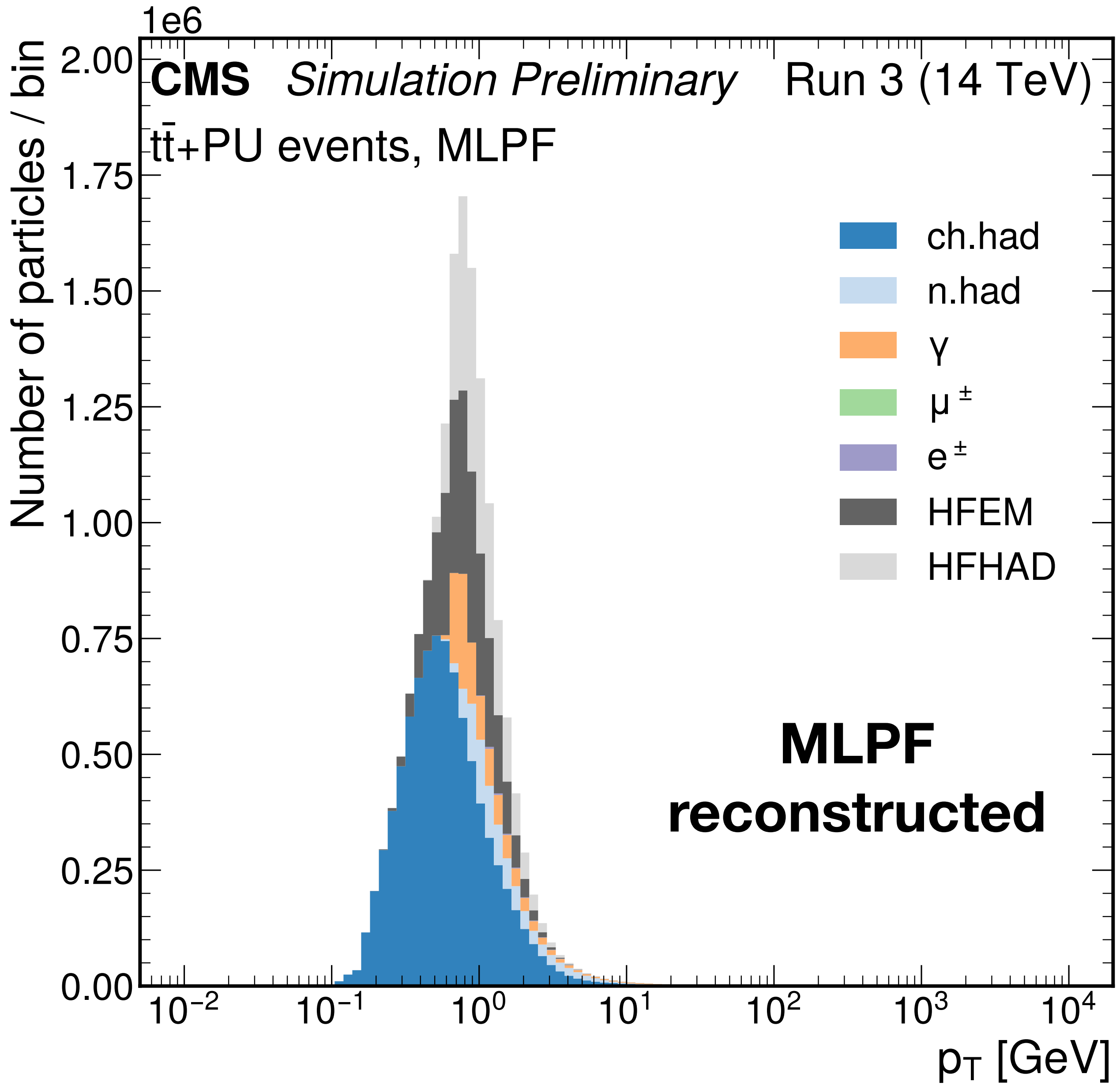
Performance

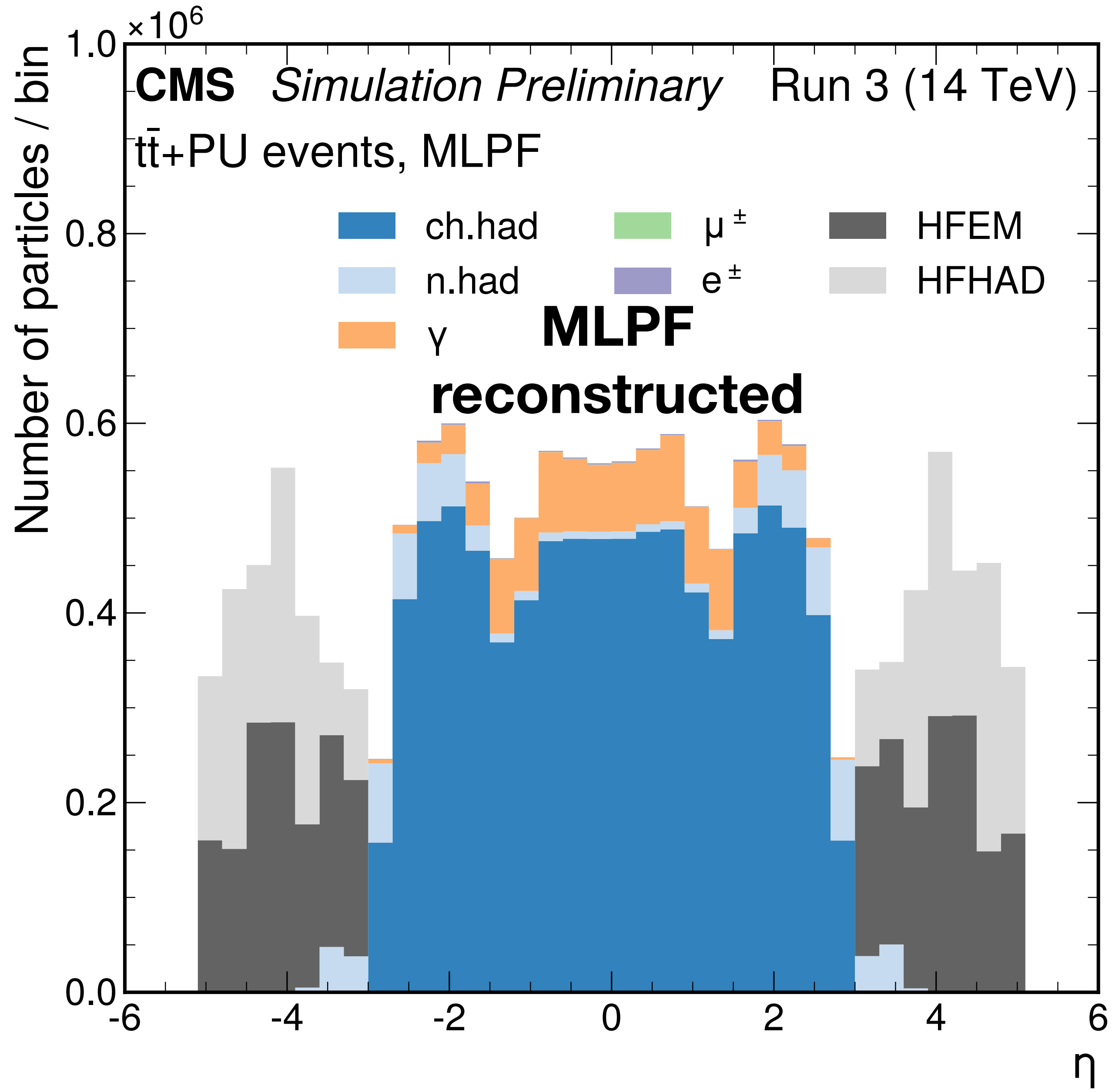




- MLPF is able to predict truth p_T and labels well

Performance

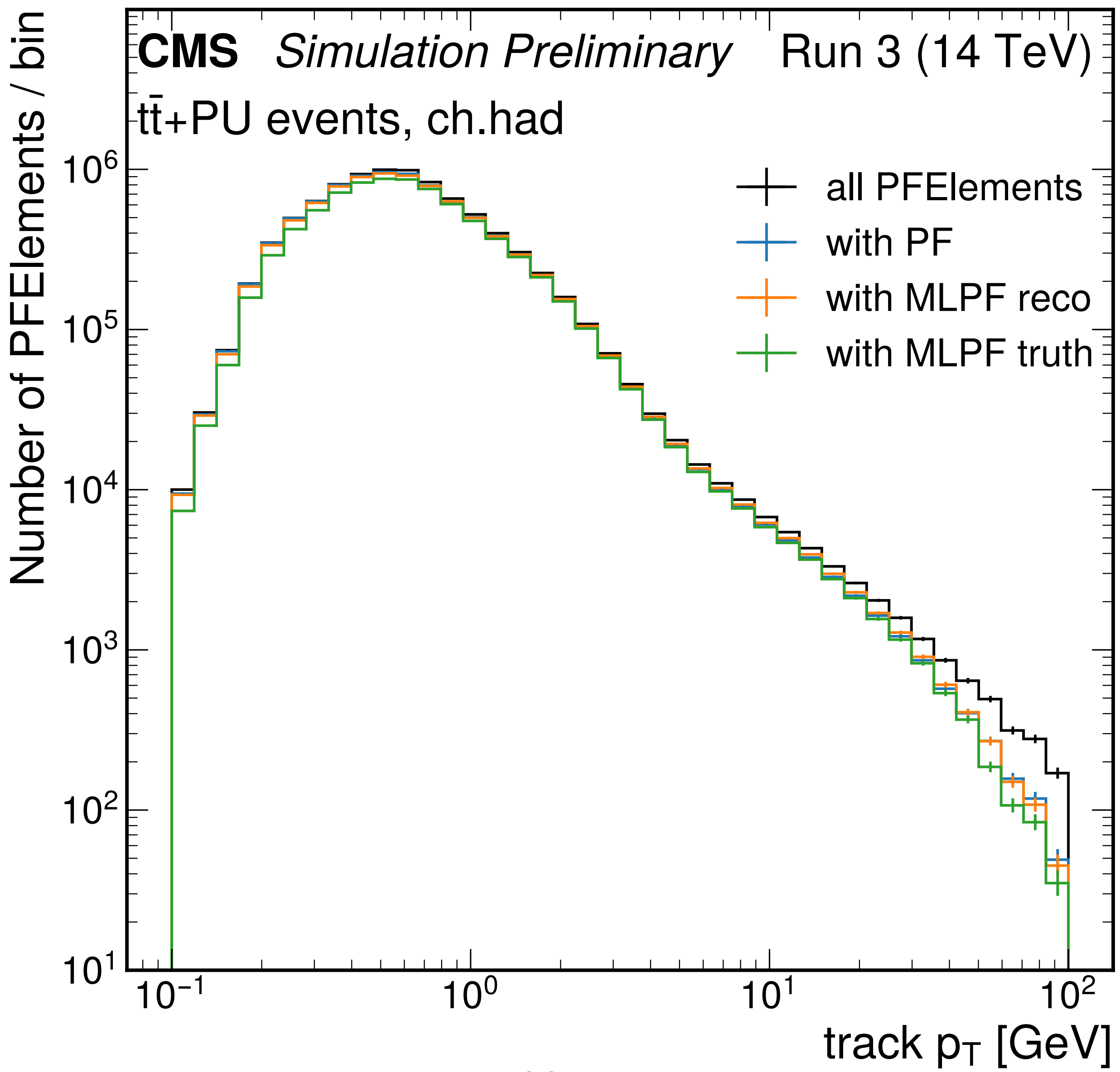




- MLPF is able to predict truth p_T and labels well

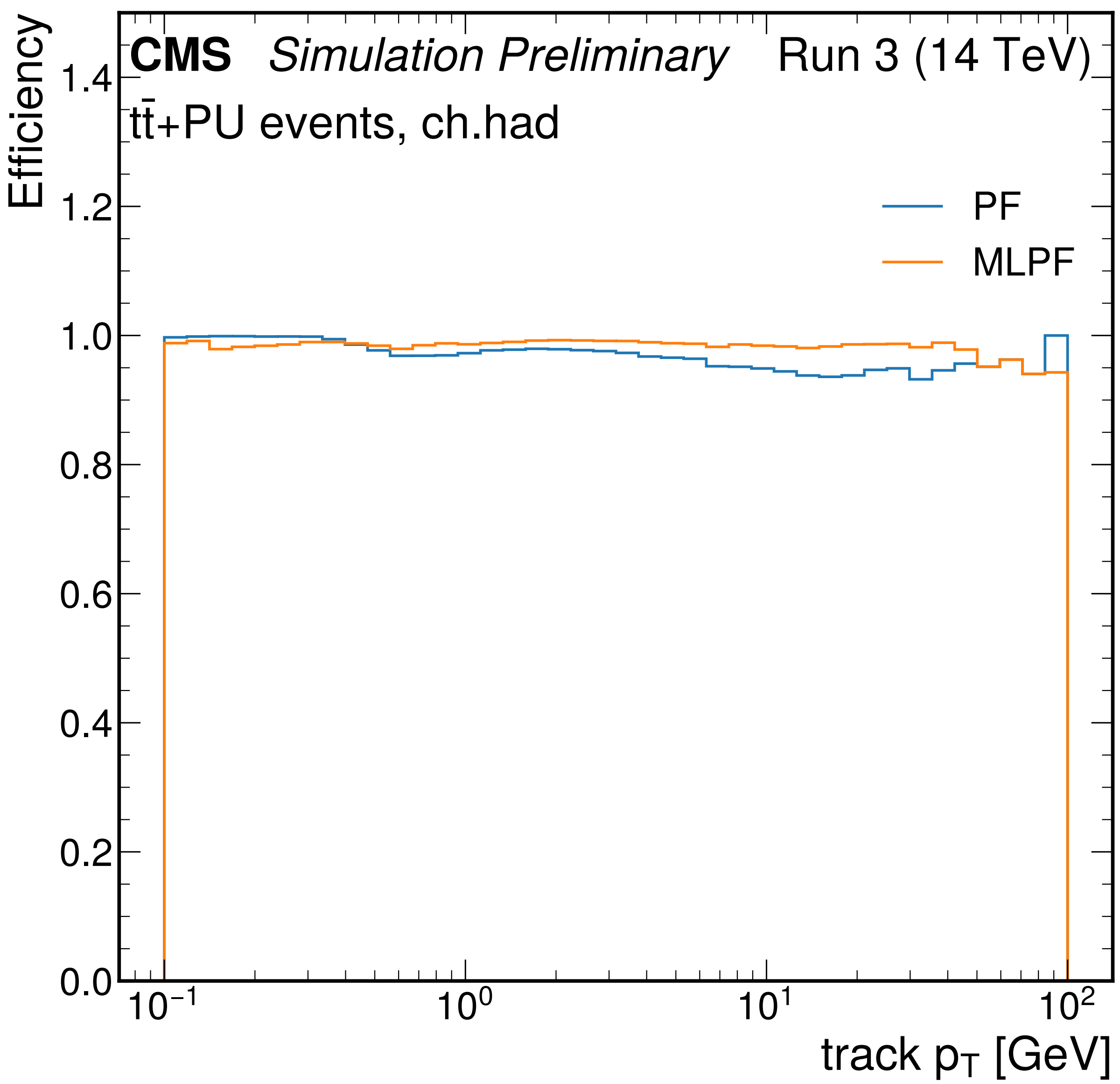
Performance (CH)

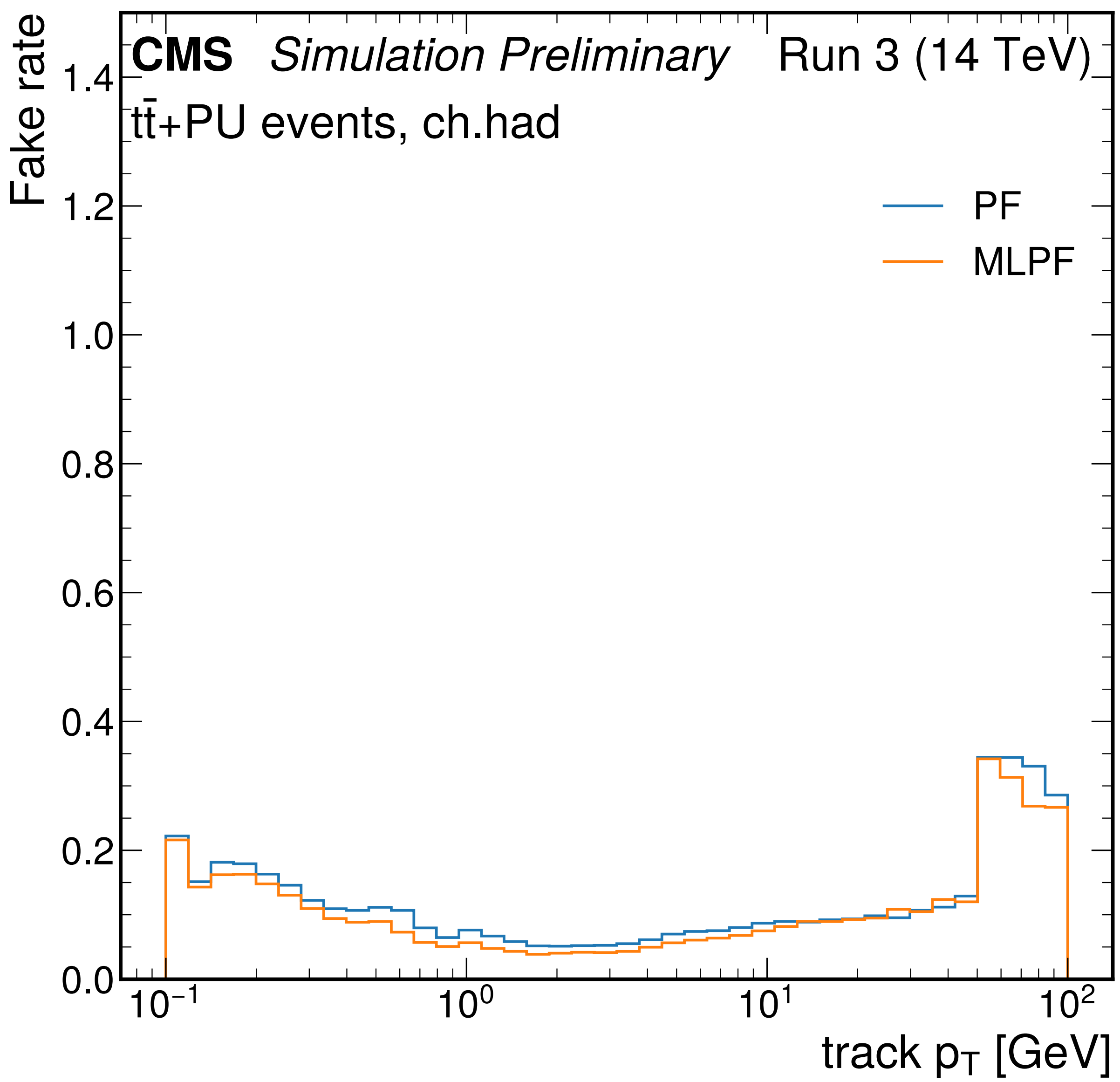
- Similar distributions from **PF** and **MLPF** for charged hadrons
- Similar efficiency & fake rate, small improvements from **MLPF**



Performance (CH)

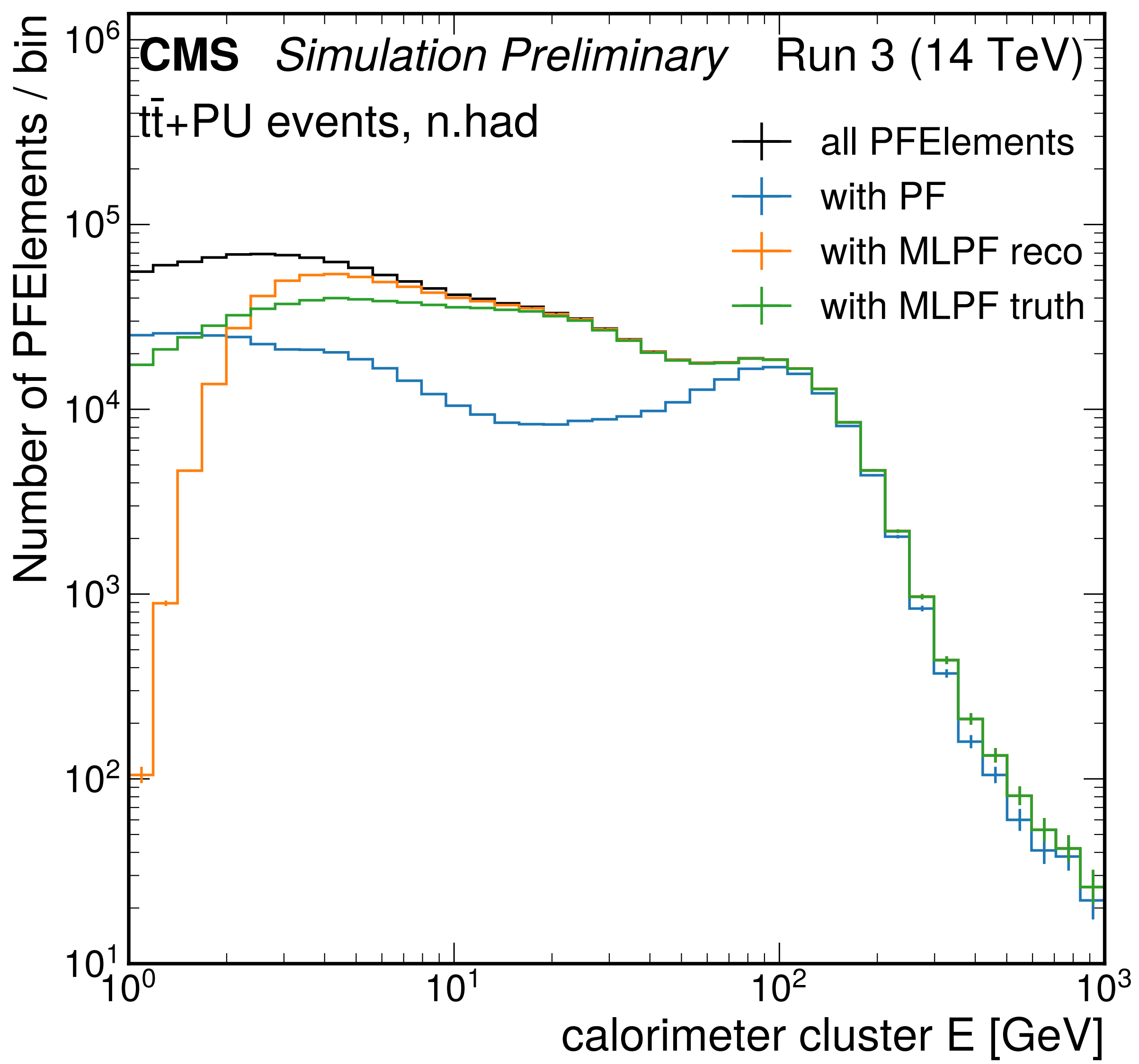
- Similar distributions from **PF** and **MLPF** for charged hadrons
- Similar efficiency & fake rate, small improvements from **MLPF**





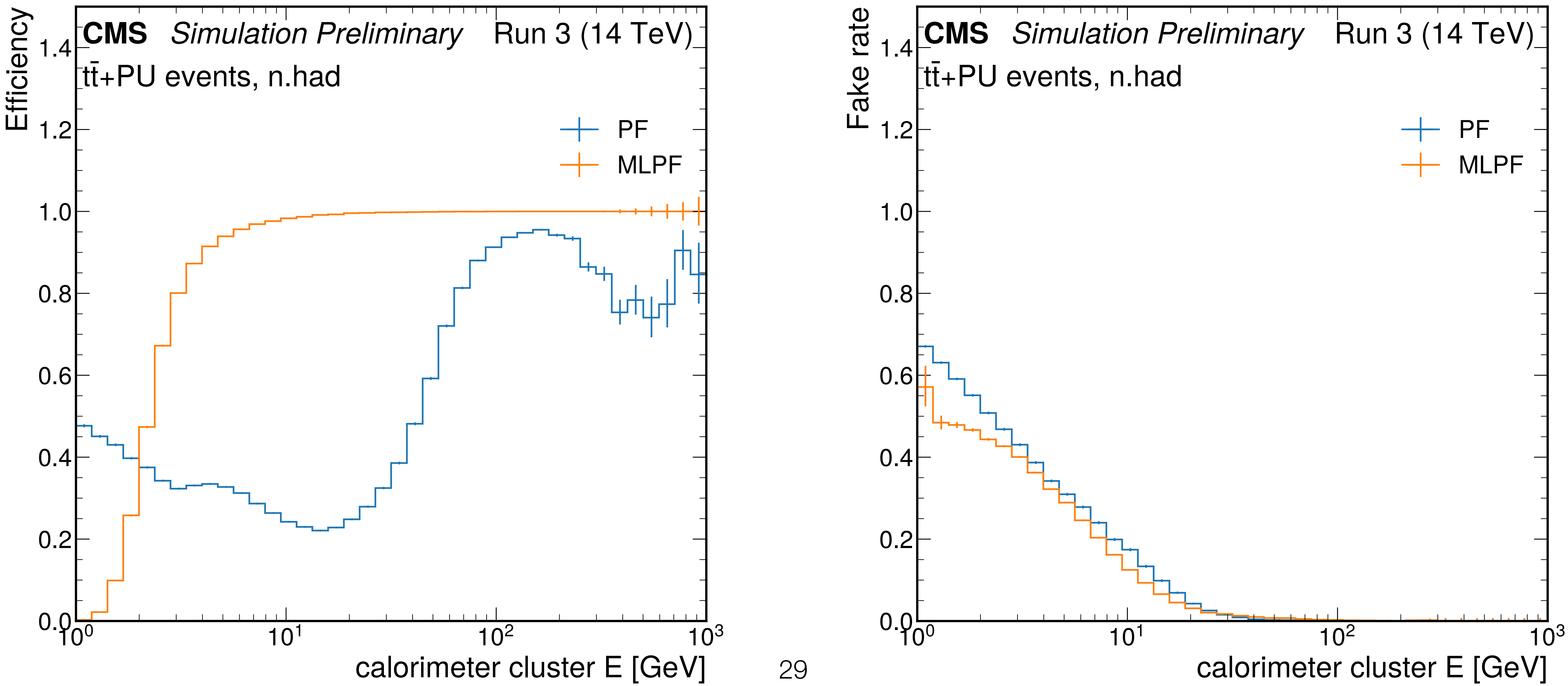
Performance (NH)

- Quite different distributions from **PF** and **MLPF** for neutral hadrons, improved efficiency from **MLPF**
- **PF** operates at high efficiency at the cost of high fake rate for low energy neutrals

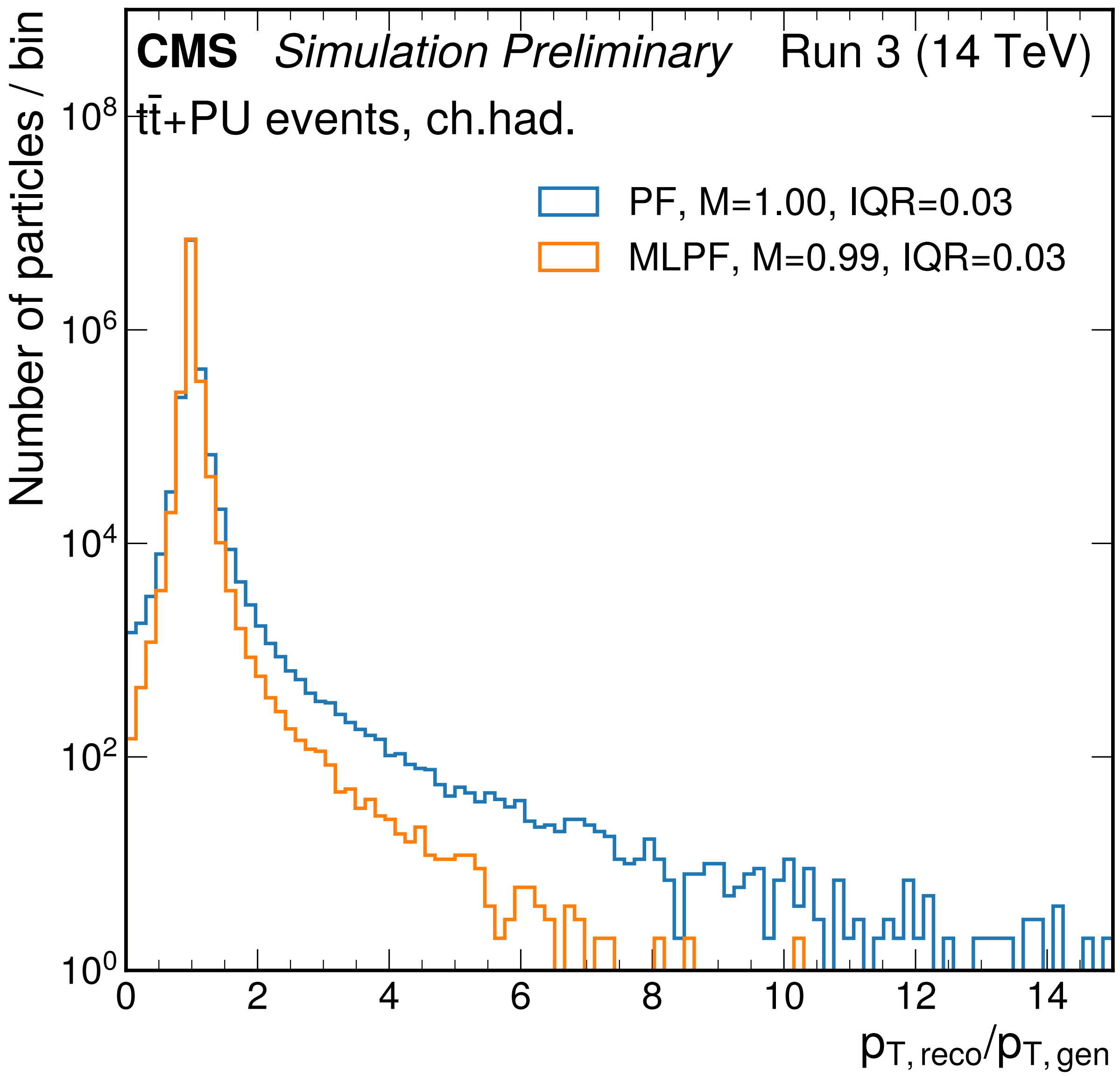


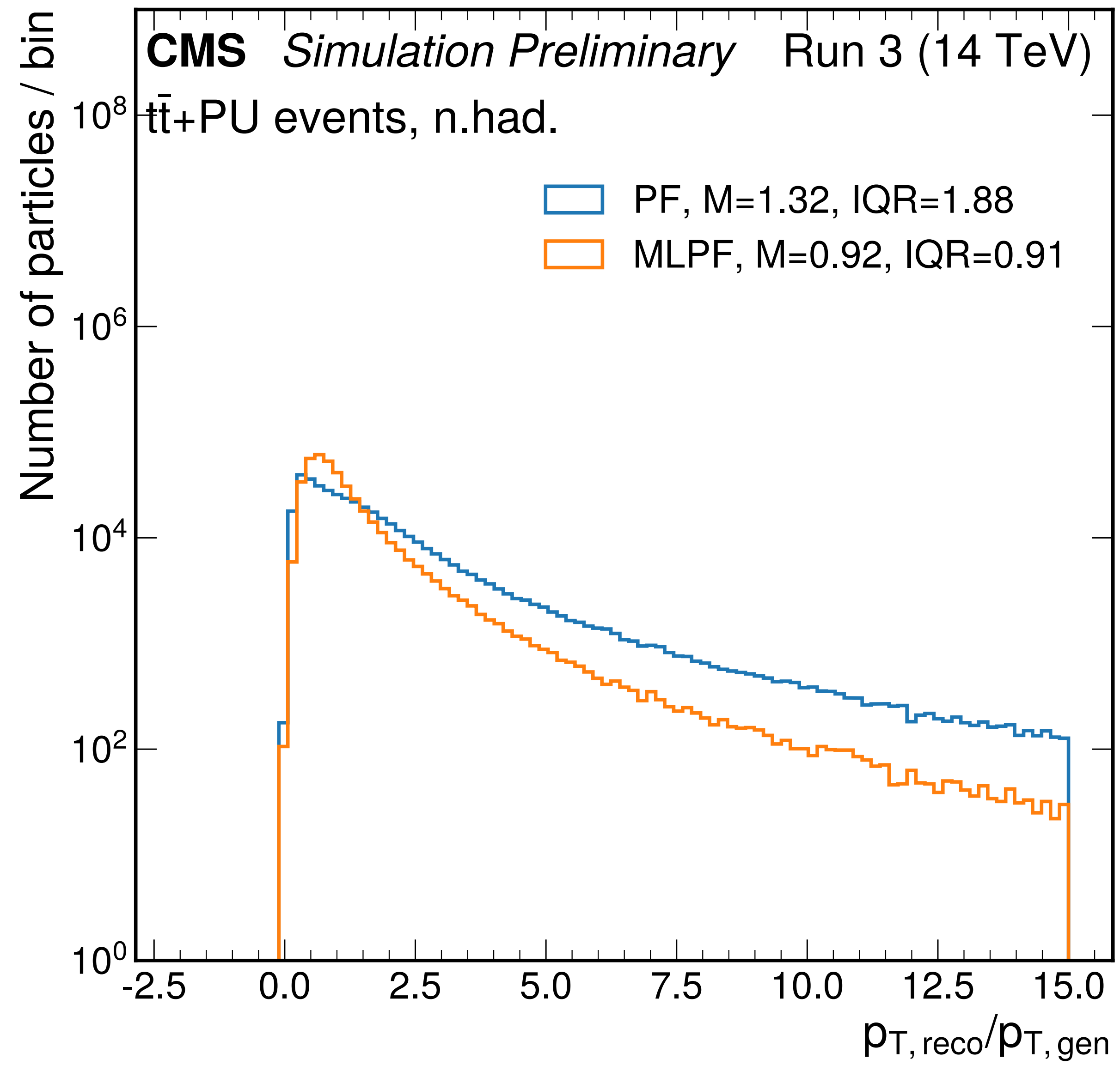
Performance (NH)

- Quite different distributions from **PF** and **MLPF** for neutral hadrons, improved efficiency from **MLPF**
- **PF** operates at high efficiency at the cost of high fake rate for low energy neutrals



Performance (p_T)

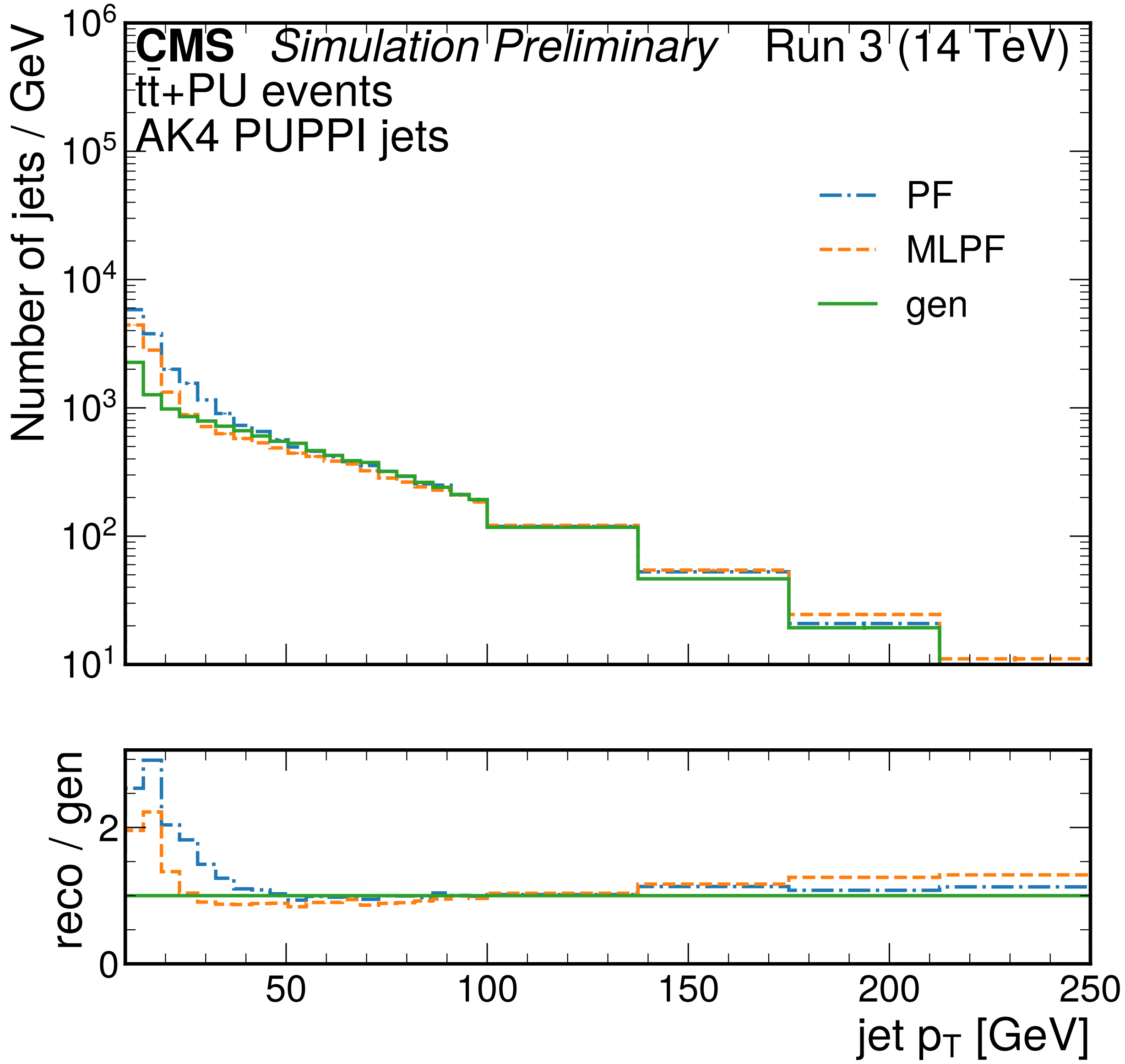


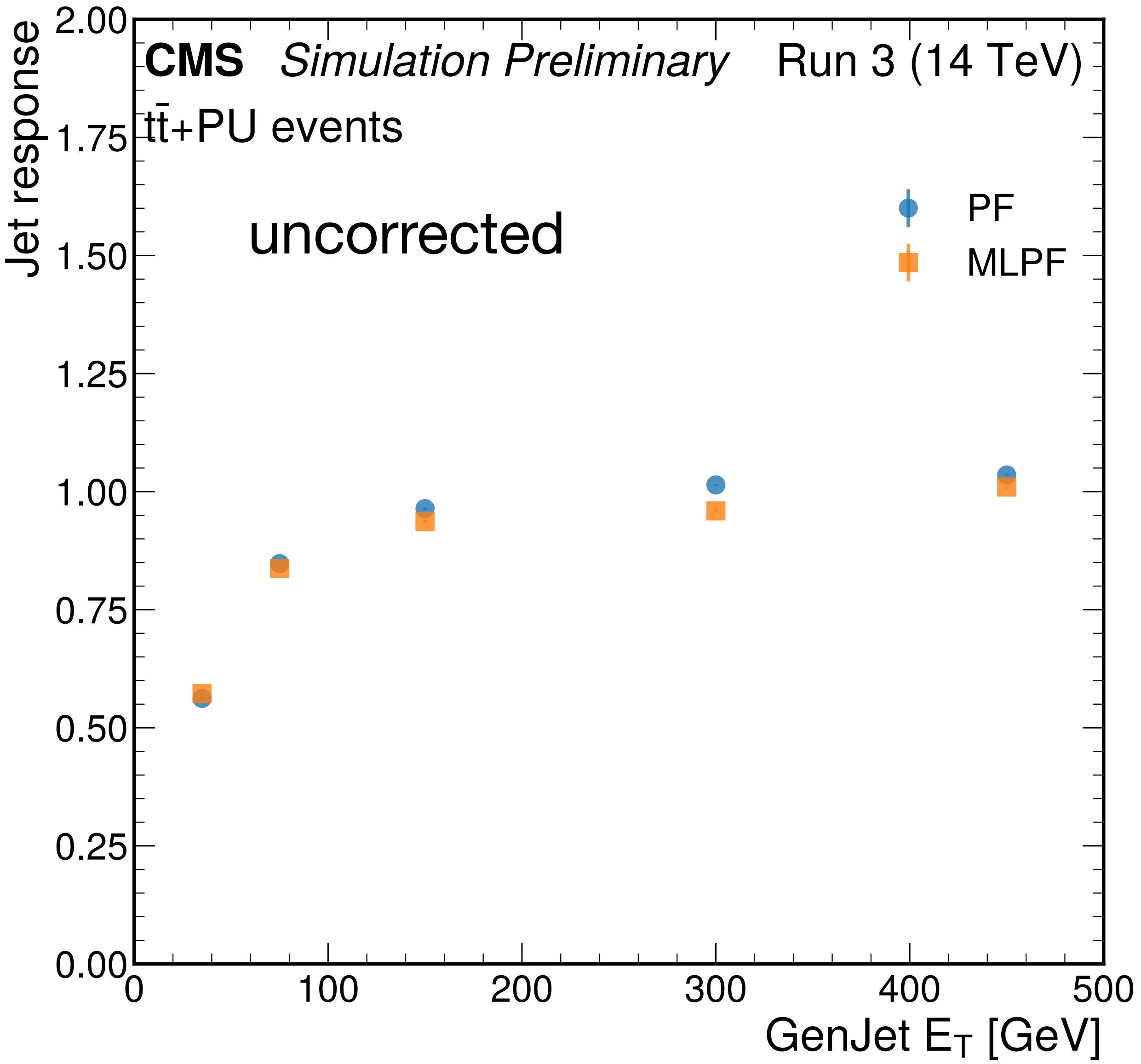


- Slight improvement in charged and neutral particle p_T resolution

M = median
IQR = interquartile range ($Q_{75\%}$-$Q_{25\%}$)

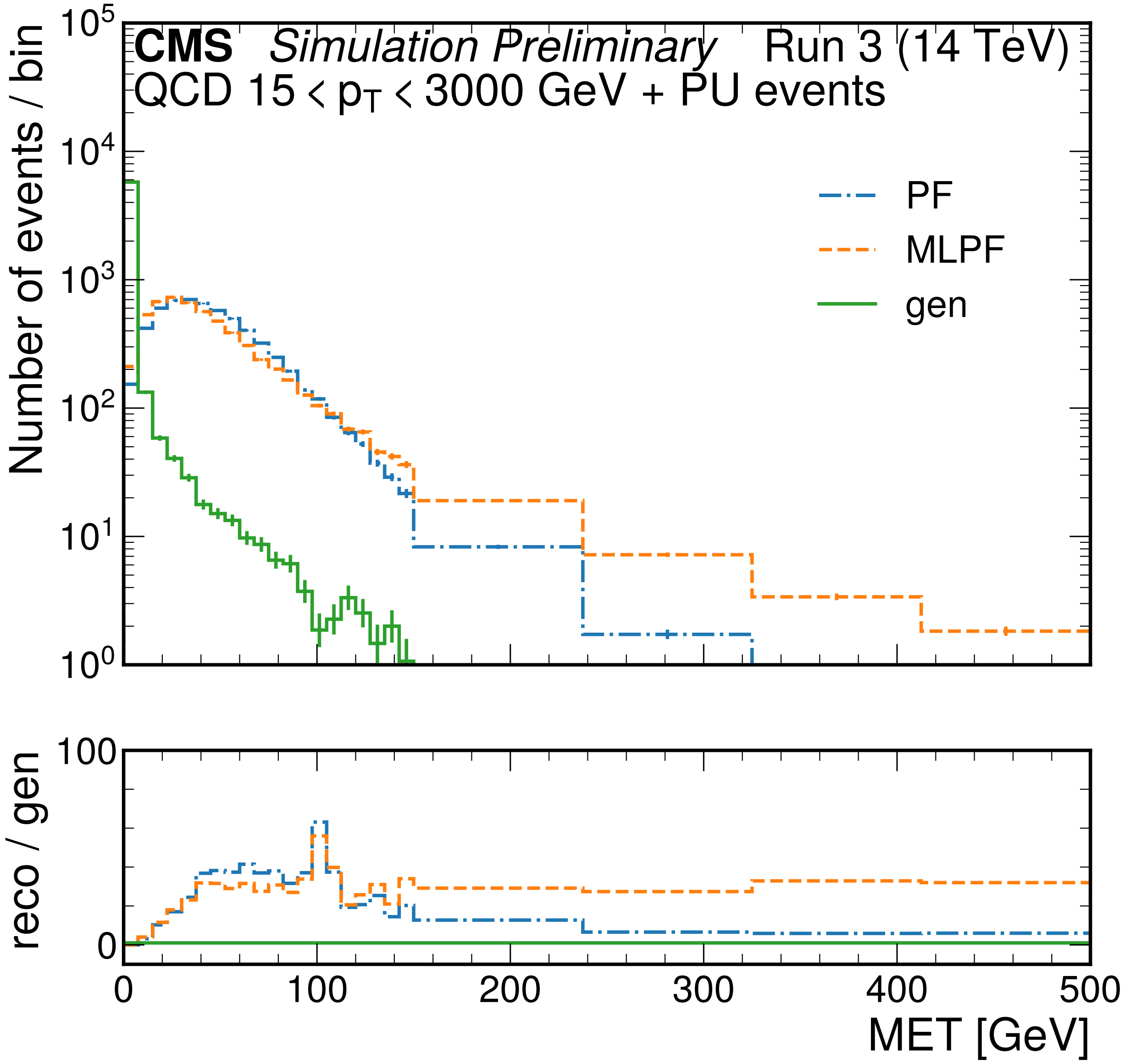
Performance (Jets)

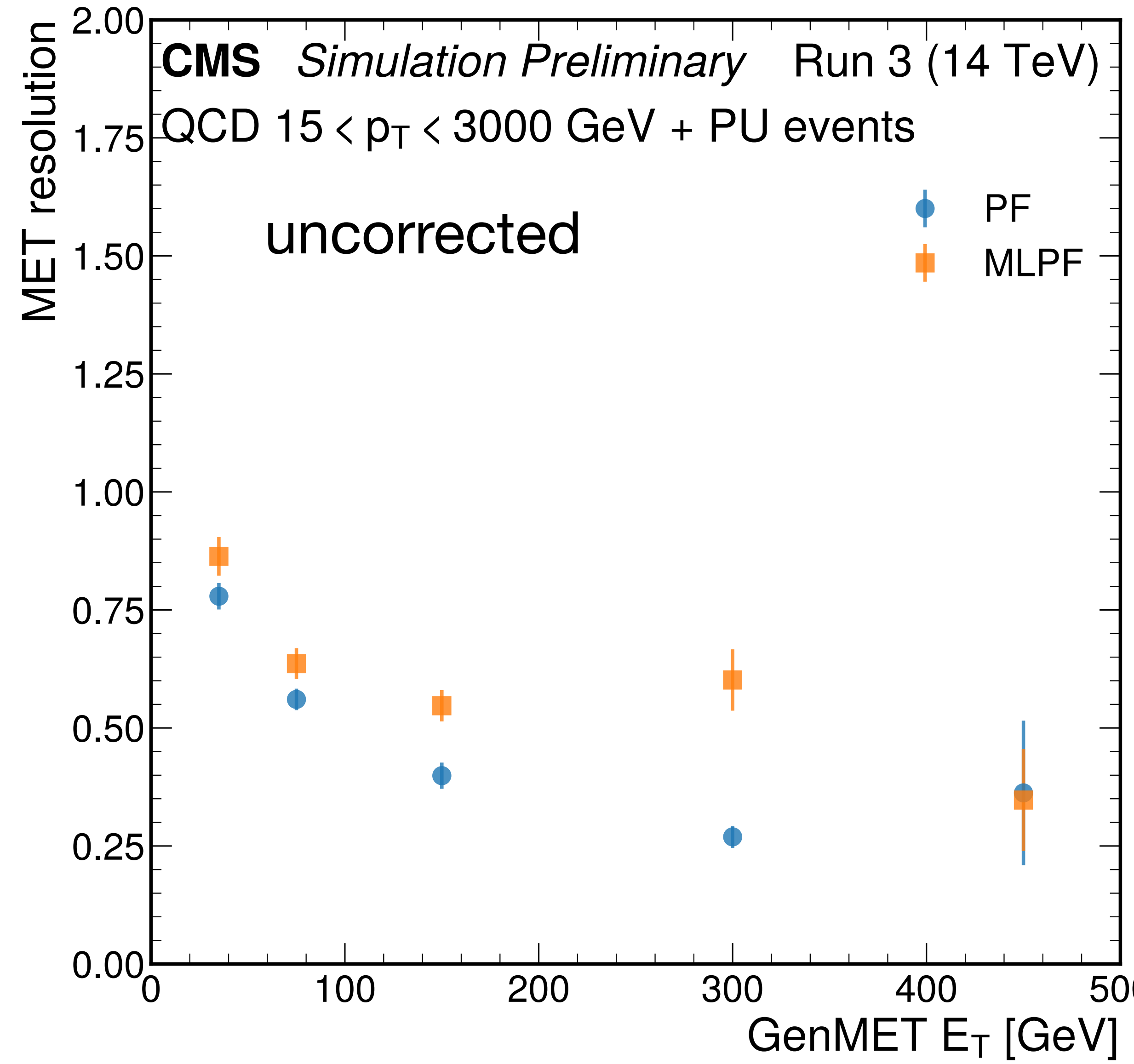




- Similar performance for jets from PF and MLPF

Performance (MET)





- Some large MET tails from MLPF (observed also with MLPF v1)
- Appears to originate from many nearby inputs all from same truth particle

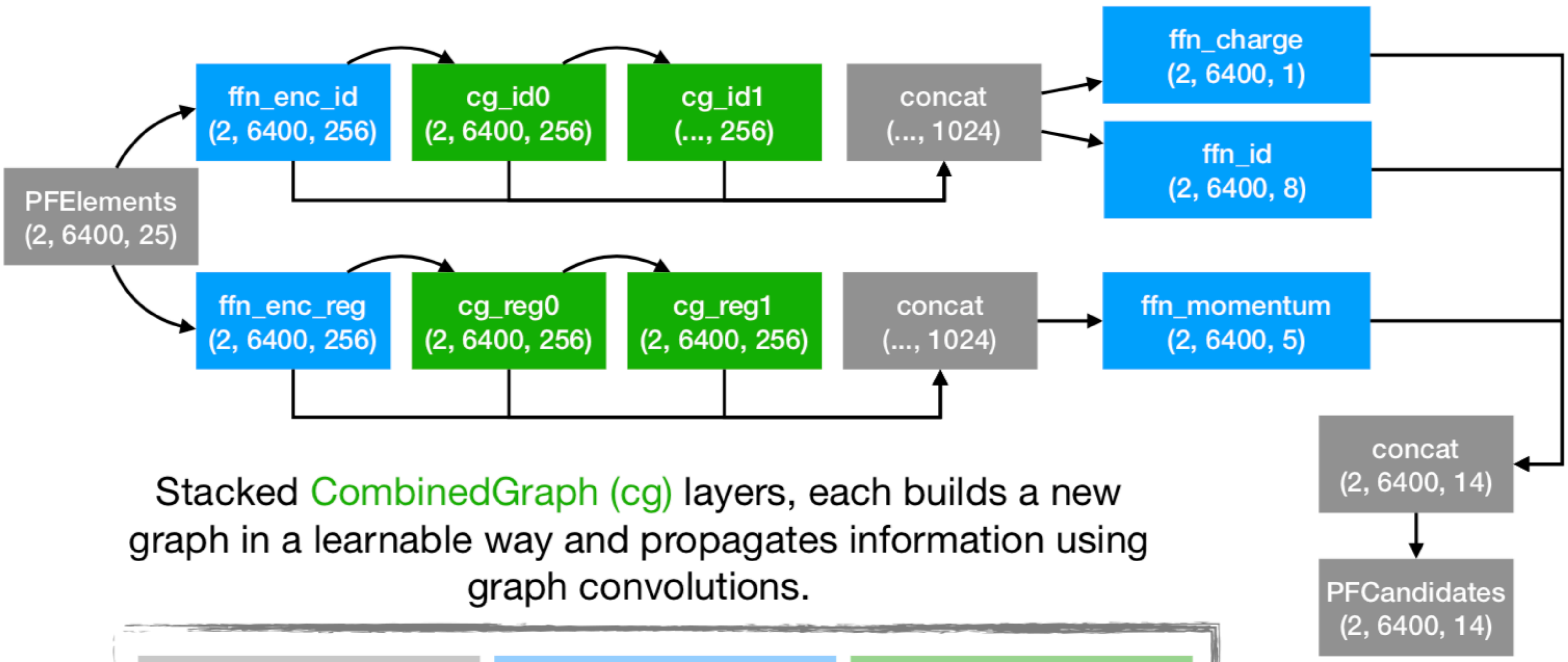
Conclusions

- MLPF algorithm continues to show promise
- Similar or better performance to PF in many regimes
  - Some ongoing investigations (eg. MET tails)
- Computationally stable scaling with number of particles
- Further developments in the pipeline

BACKUP

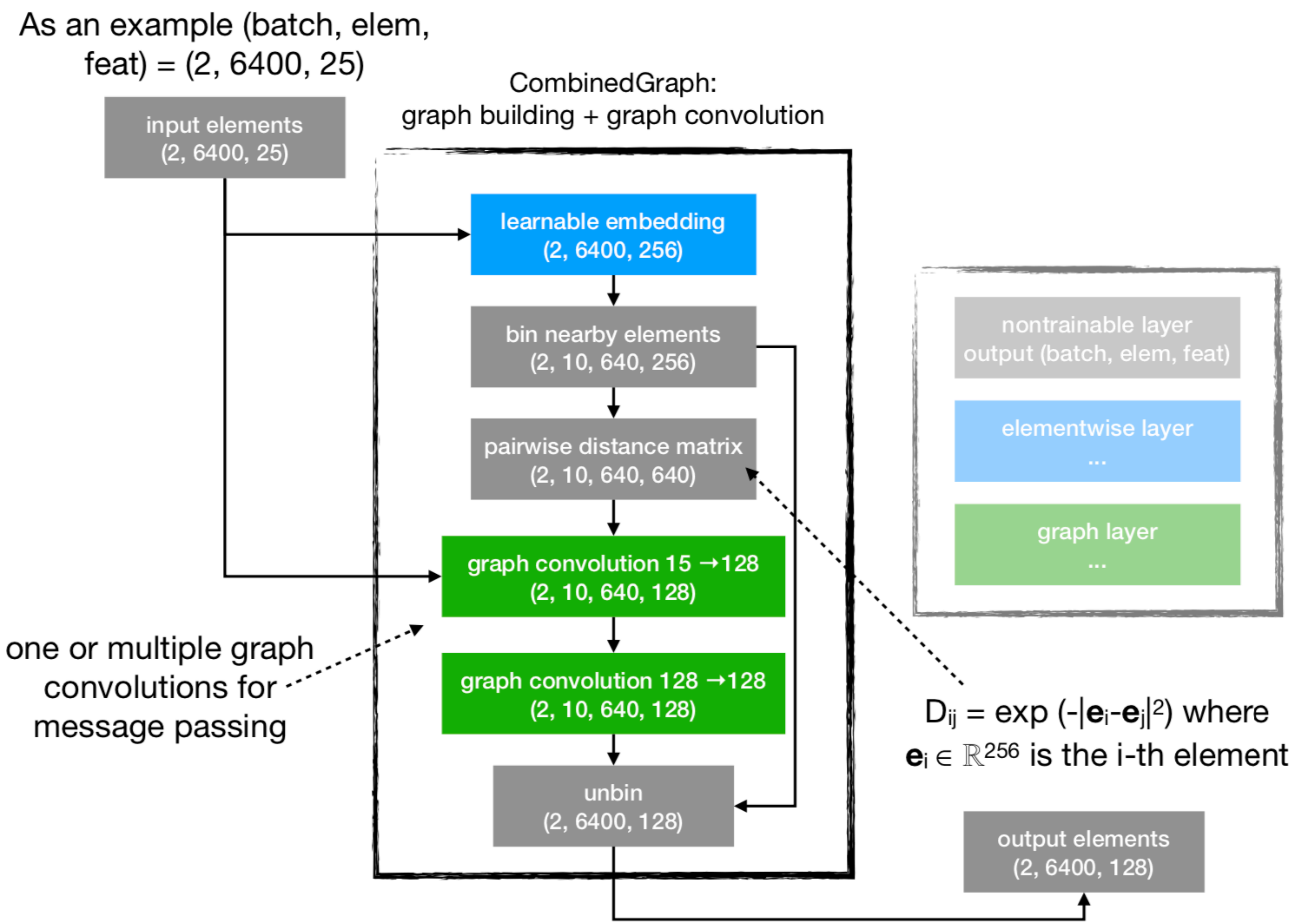
MLPF

As an example (batch, elem, feat) = (2, 6400, 25)



Stacked CombinedGraph (cg) layers, each builds a new graph in a learnable way and propagates information using graph convolutions.

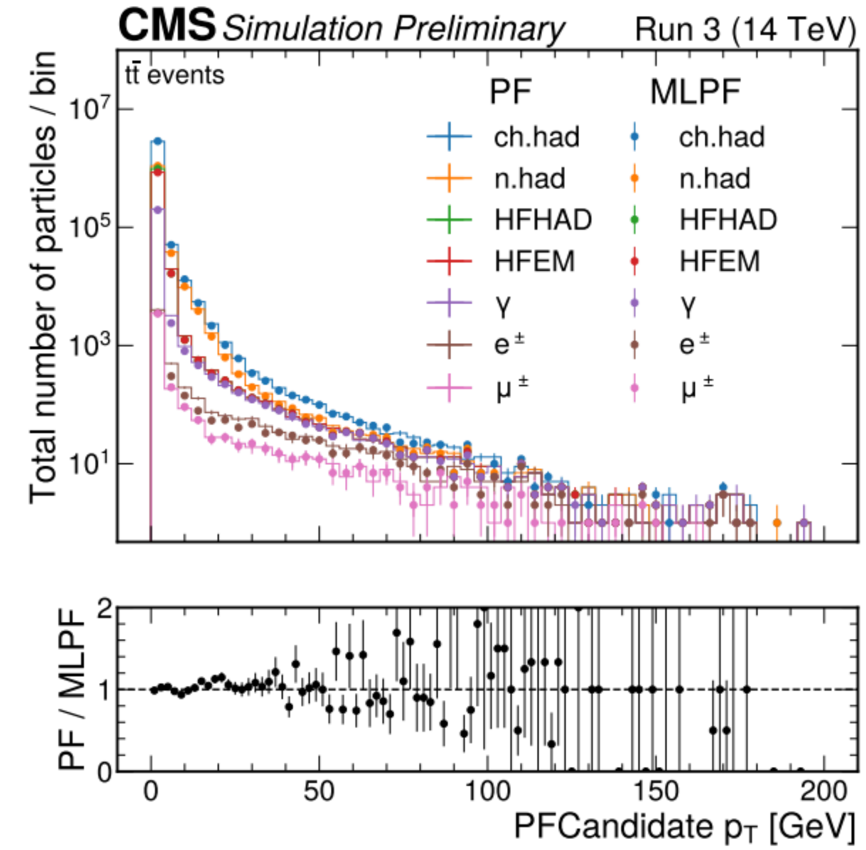
CombinedGraph

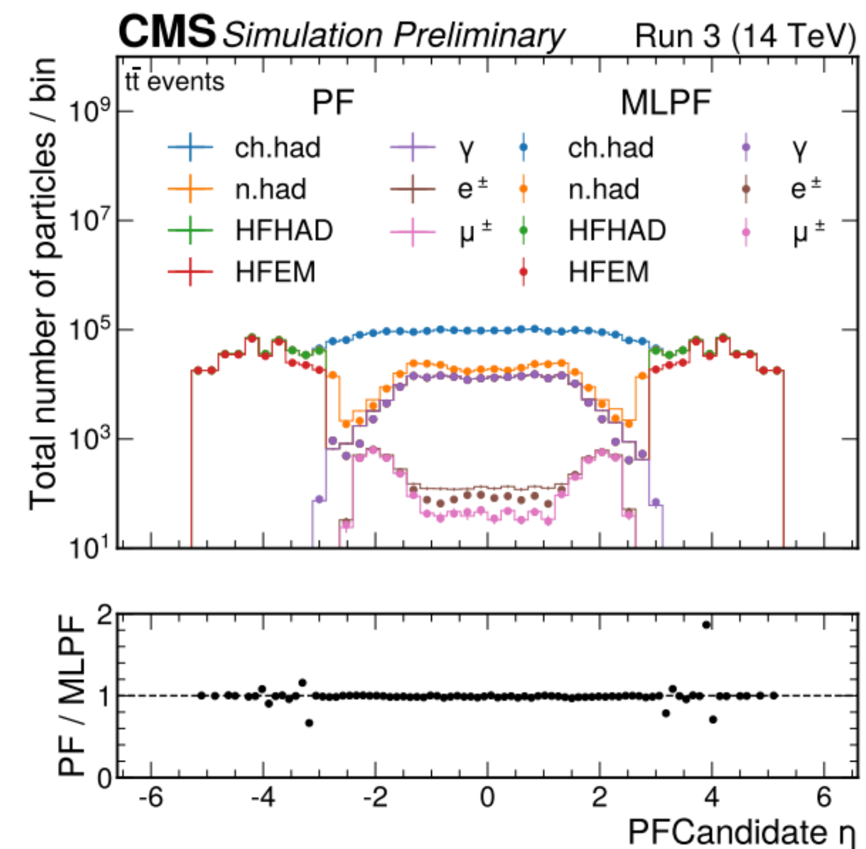


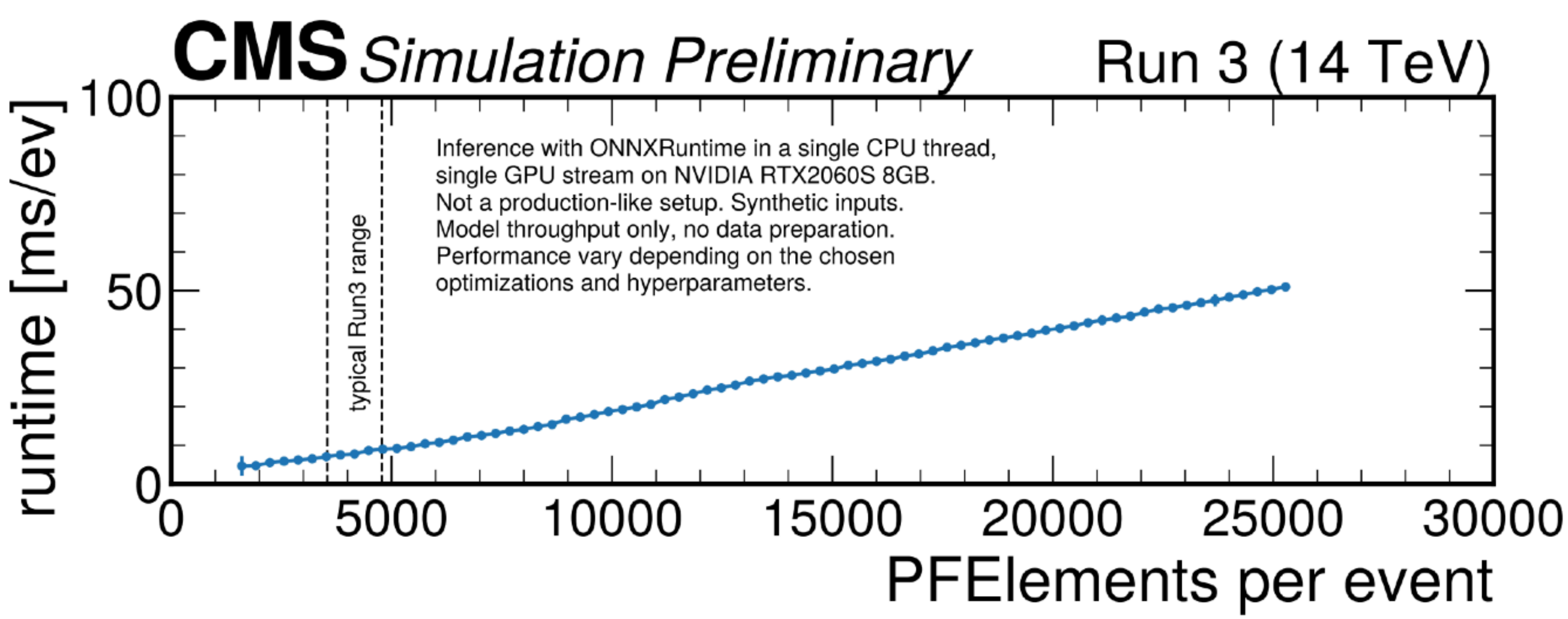
Uses built-in dense matrix, reshape and scatter/gather operations in TF.
Requires batch-mode graphs. No N^2 allocation or computation needed.

MLPF v1

- First version trained using PF as target
    - Can’t exceed PF performance, but useful proof-of-concept
- Very promising results (both for physics performance and computational scaling)





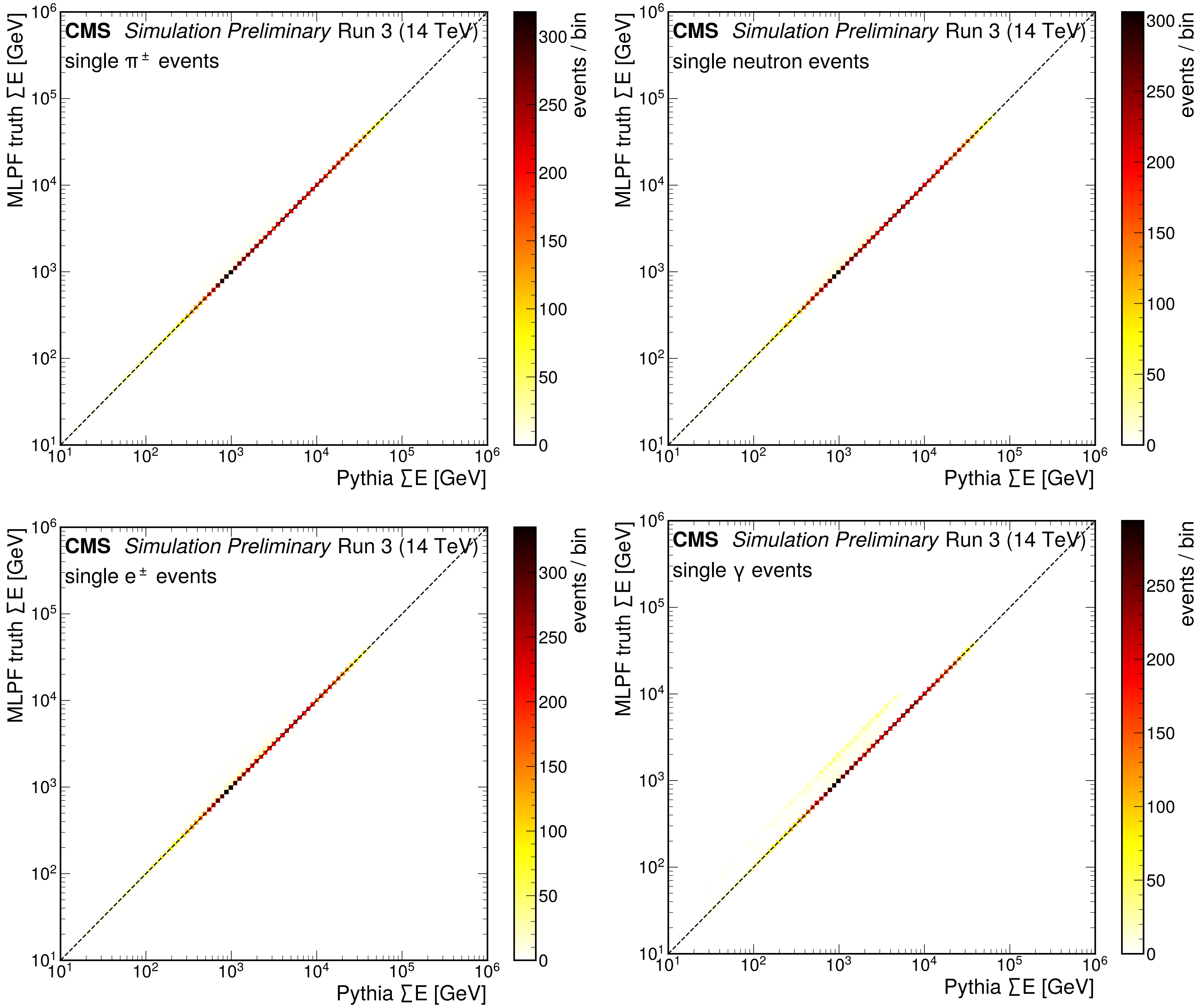


Samples

physics process	PU configuration	MC events
top quark-antiquark pairs	flat 55–75	100 k
QCD $\hat{p}_T \in [15, 3000]$ GeV	flat 55-75	100 k
QCD $\hat{p}_T \in [3000, 7000]$ GeV	flat 55–75	100 k
$Z \to \tau\tau$ all-hadronic	flat 55–75	100 k
single e flat $p_T \in [1, 1000]$ GeV	no PU	10 k
single μ log-flat $p_T \in [0.1, 2000]$ GeV	no PU	10k
single π^0 flat $p_T \in [0, 1000]$ GeV	no PU	10 k
single $\pi^\pm$ flat $p_T \in [0.7, 1000]$ GeV	no PU	10 k
single τ flat $p_T \in [1, 1000]$ GeV	no PU	10 k
single γ flat $p_T \in [1, 1000]$ GeV	no PU	10 k
single p flat $p_T \in [0.7, 1000]$ GeV	no PU	10 k
single n flat $p_T \in [0.7, 1000]$ GeV	no PU	10 k

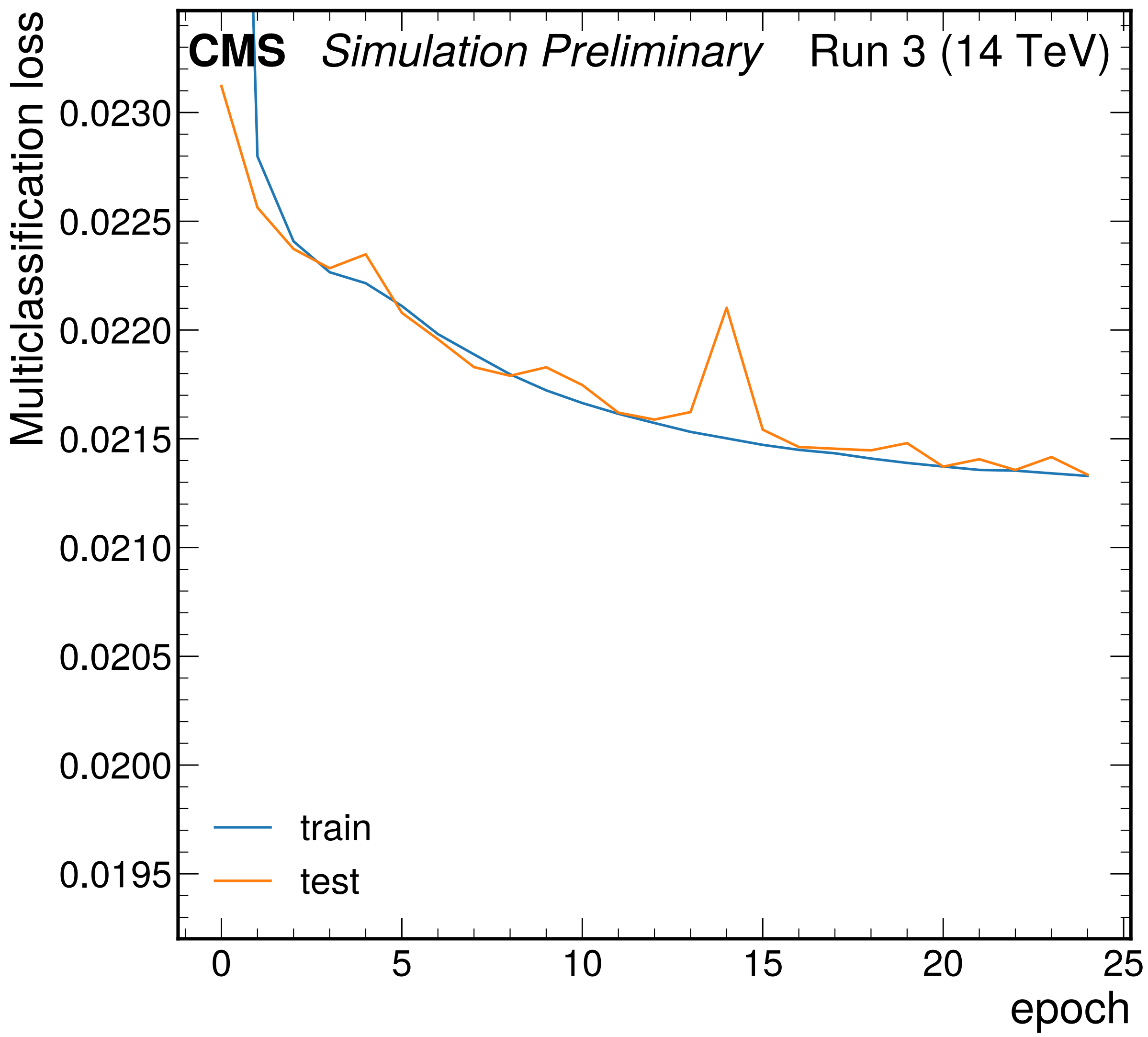
Table 1: MC simulation samples used for optimizing the MLPF model.

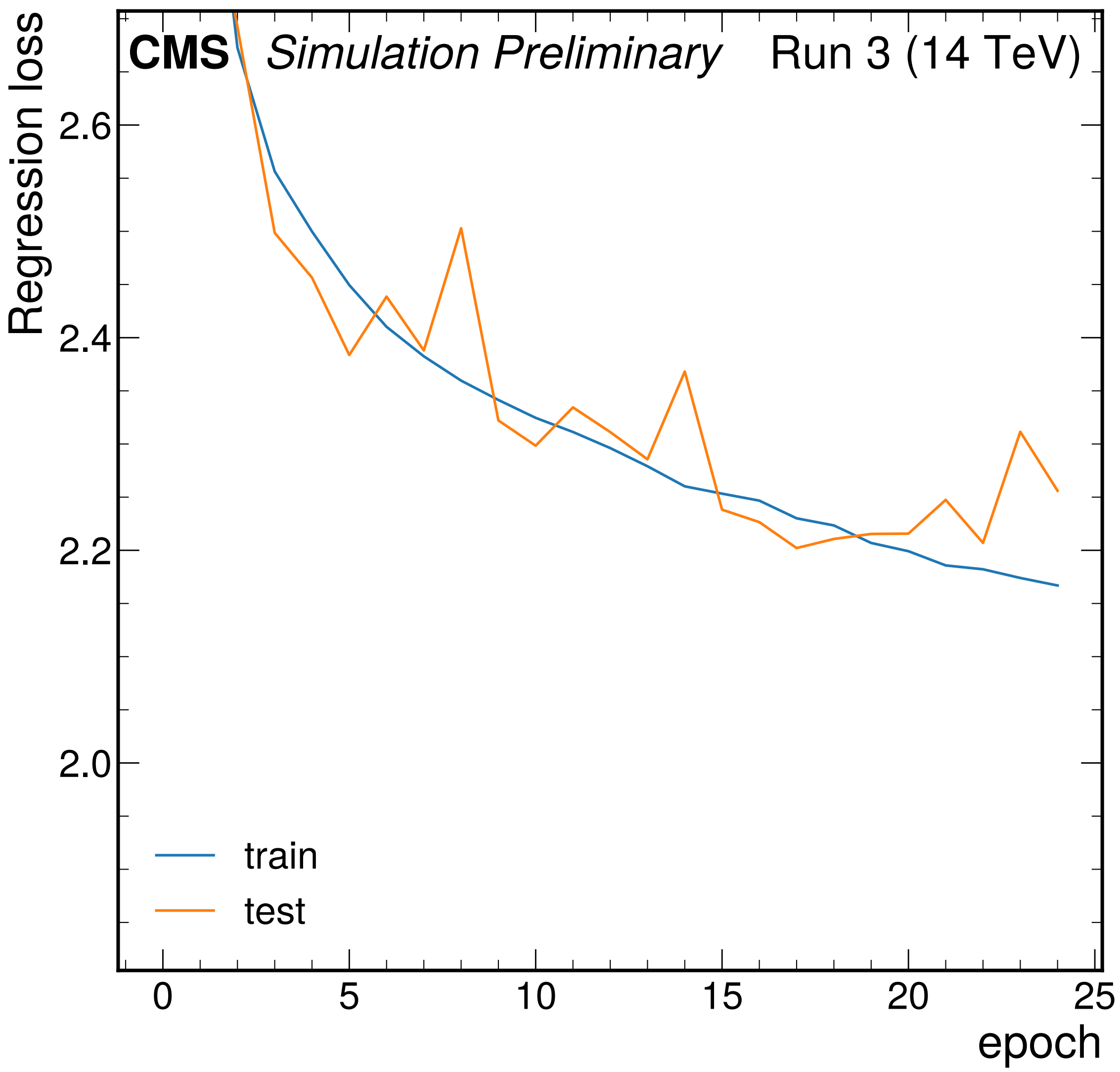
Truth Validation



MLPF truth cross-checked against generator-level info in PU0 particle gun samples

Loss

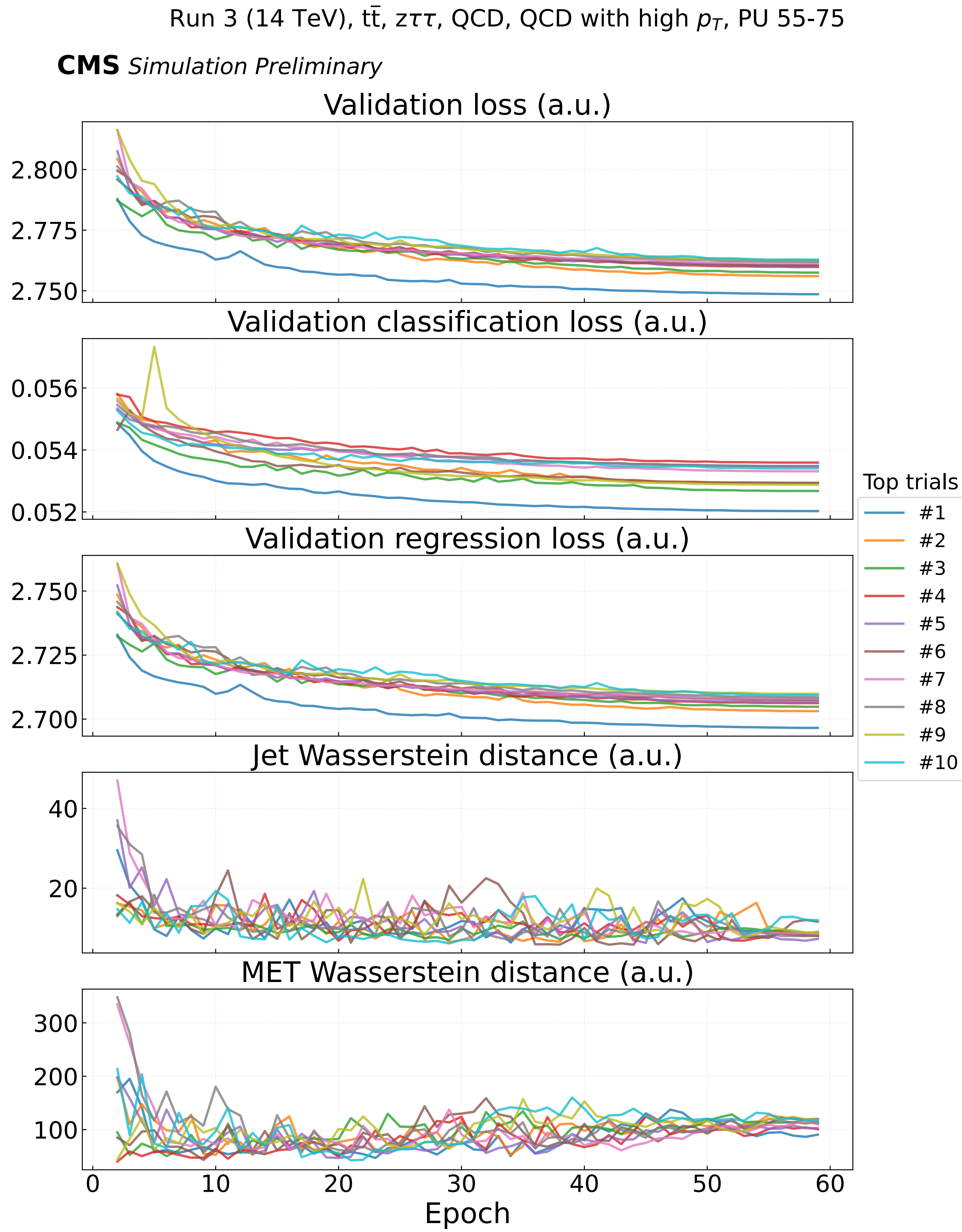

CMS Simulation Preliminary Run 3 (14 TeV)
Multiclassification loss
0.0230
0.0225
0.0220
0.0215
0.0210
0.0205
0.0200
0.0195
train
test
0
5
10
15
20
25
epoch


CMS Simulation Preliminary Run 3 (14 TeV)
Regression loss
2.6
2.4
2.2
2.0
train
test
0
5
10
15
20
25
epoch

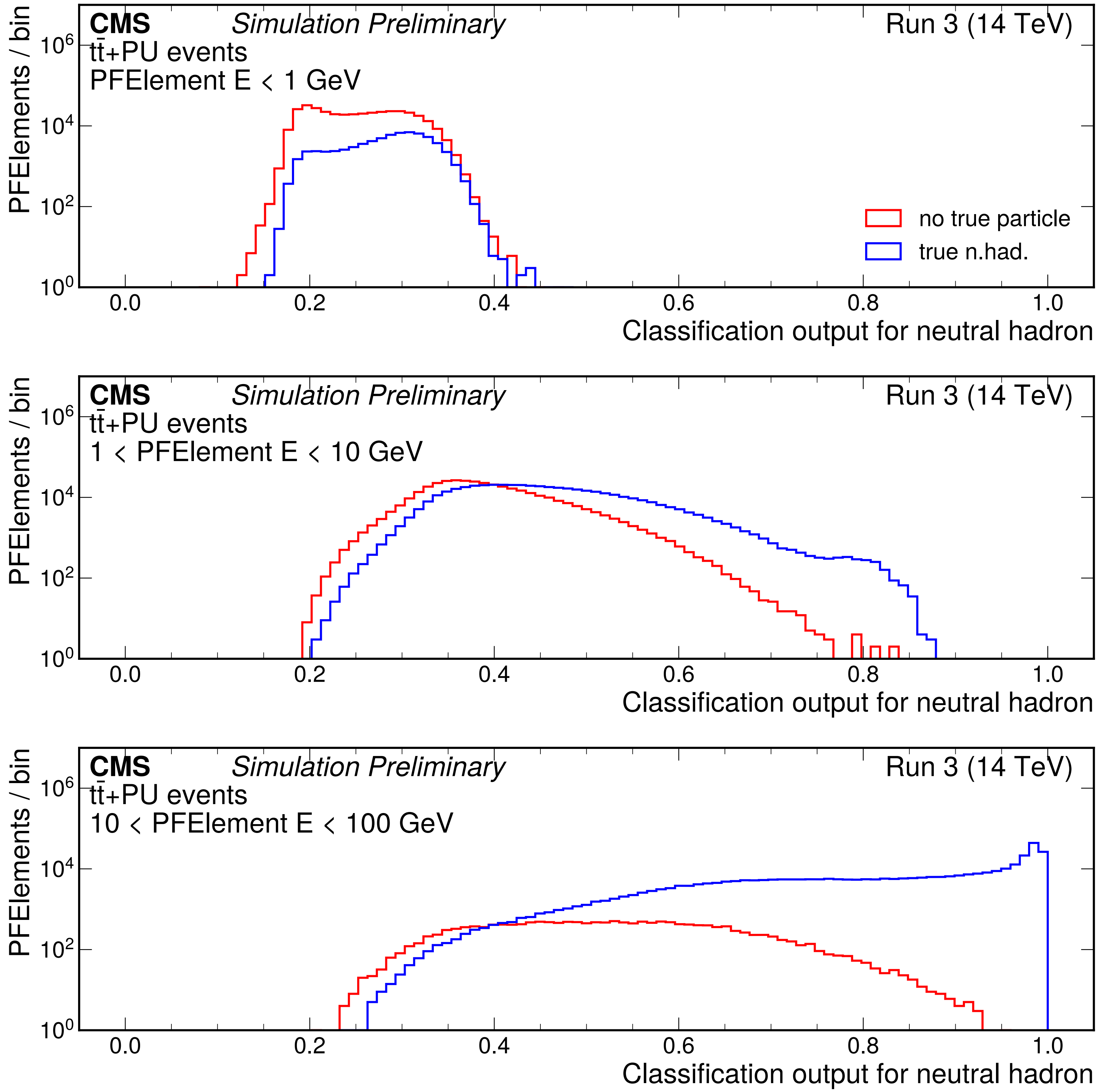
Hypertuning

Hyperparameter	Search space
lr	$\log lr \sim U(10^{-4}, 1 \cdot 10^{-2})$
lr_schedule	{None, cosinedecay}
batch_size	{24, 40}
bin_size	{32, 64, 128, 256}
distance_dim	{32, 64, 128, 256}
ffn_dist_hidden_dim	{32, 64, 128, 256}
ffn_dist_num_layers	{1, 2, 3}
num_graph_layers_id	{0, 1, 2, 3, 4}
num_graph_layers_reg	{0, 1, 2, 3, 4}
num_node_messages	{0, 1, 2, 3}
output_dim	{8, 16, 32, 64, 128, 256}

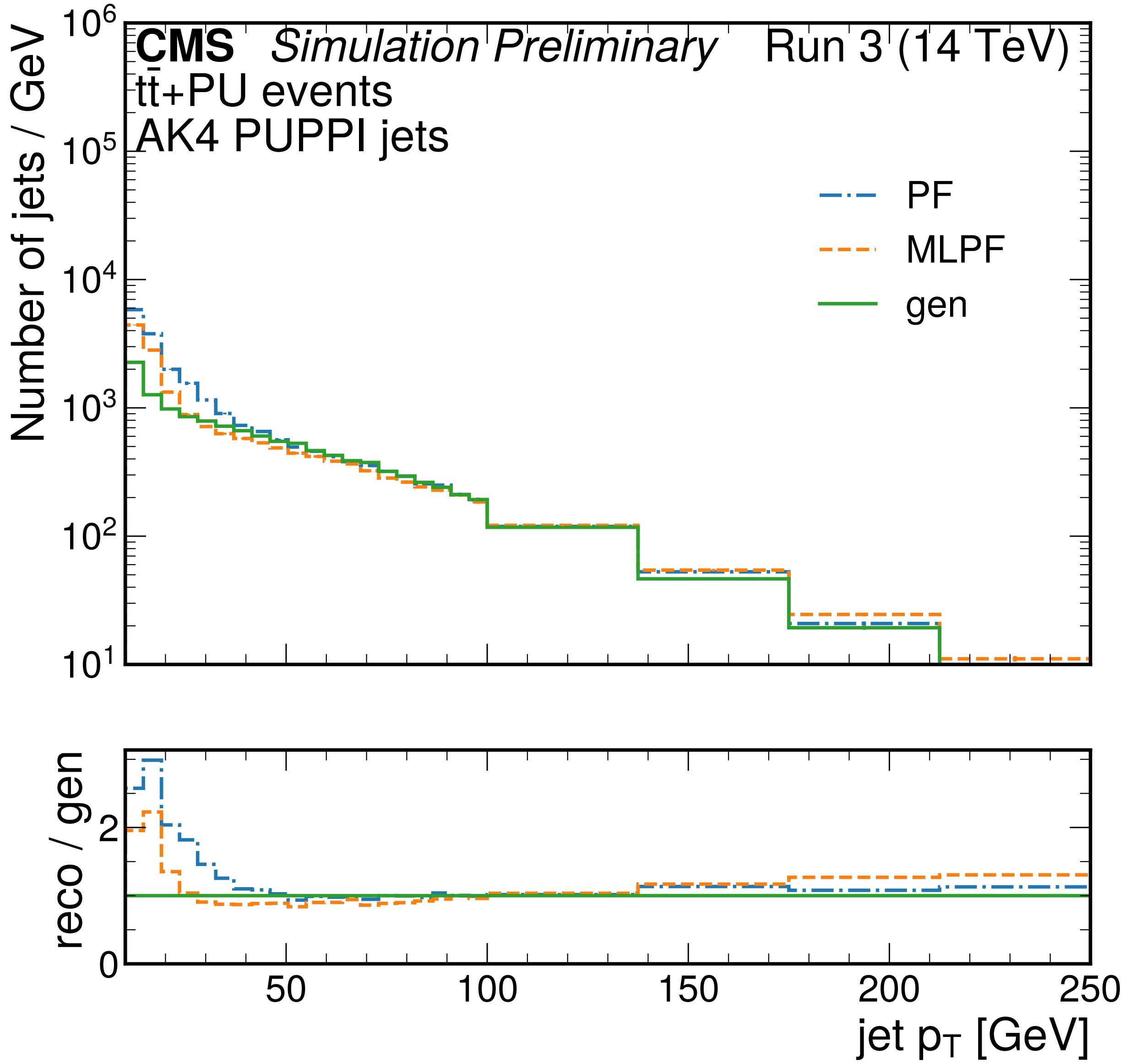
Hyperparameter	Search space
lr	0.001313
lr_schedule	cosinedecay
batch_size	24
bin_size	256
distance_dim	128
ffn_dist_hidden_dim	32
ffn_dist_num_layers	2
num_graph_layers_id	4
num_graph_layers_reg	4
num_node_messages	1
output_dim	256

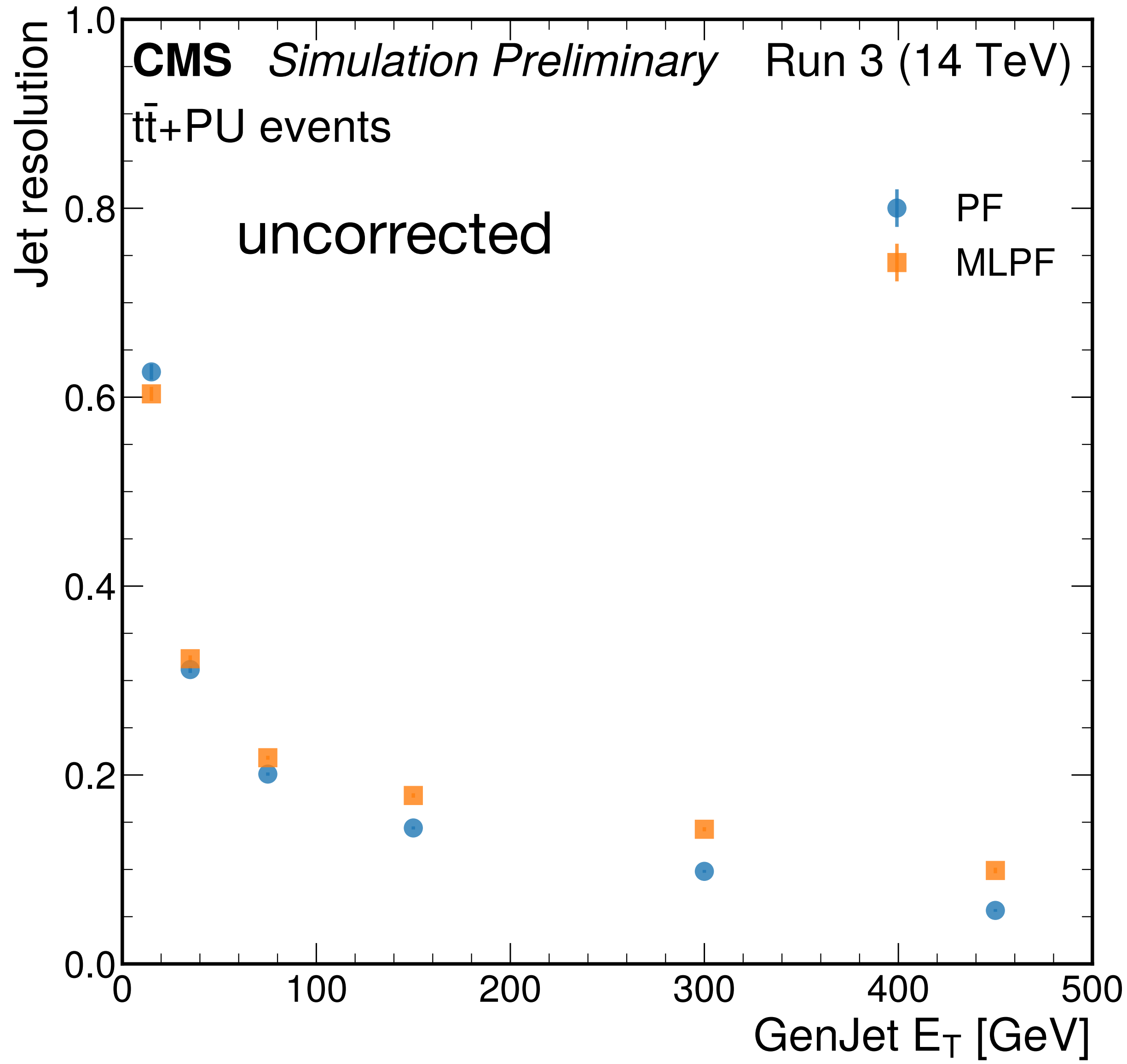


Performance (NH)



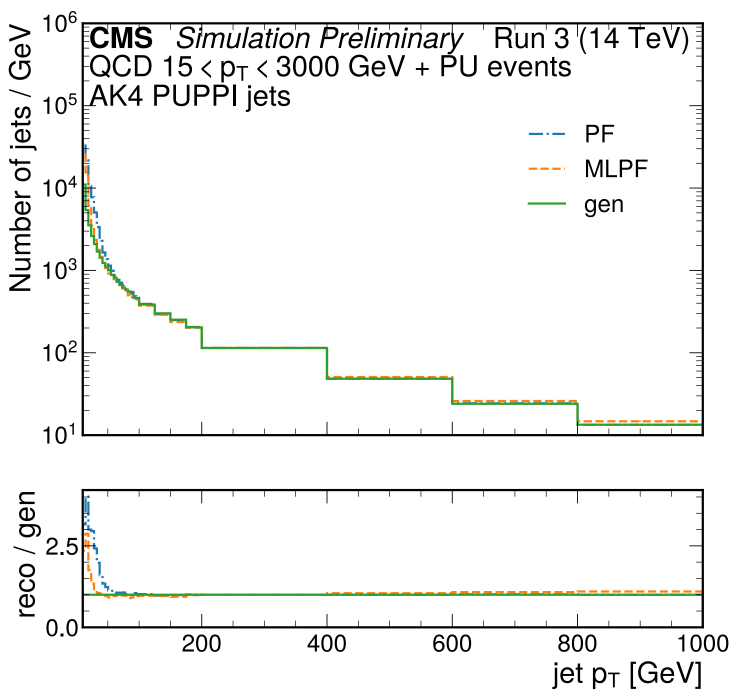
Performance (Jets)

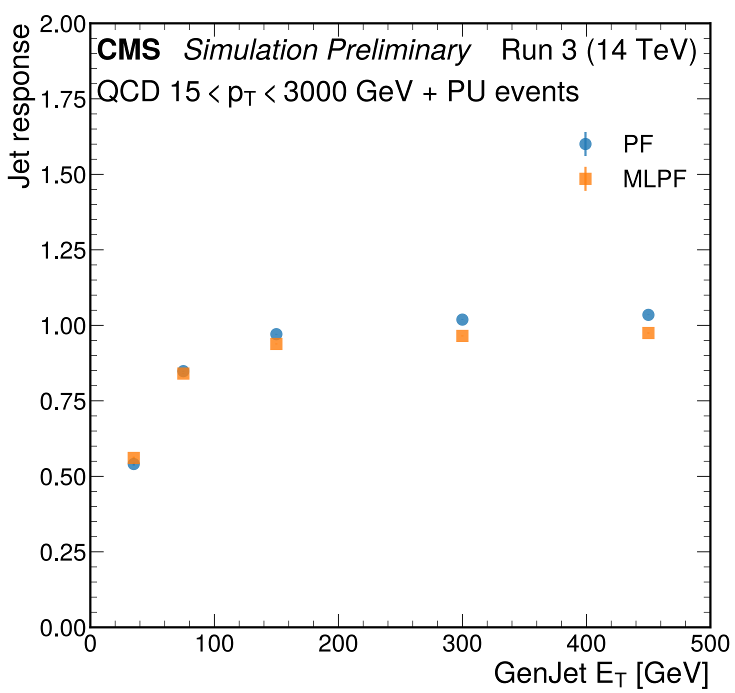


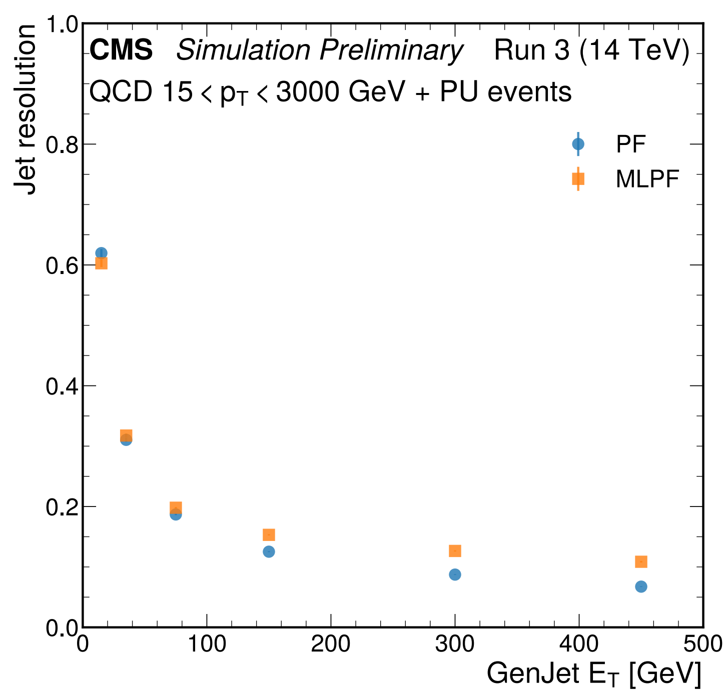


- Similar performance for jets from PF and MLPF

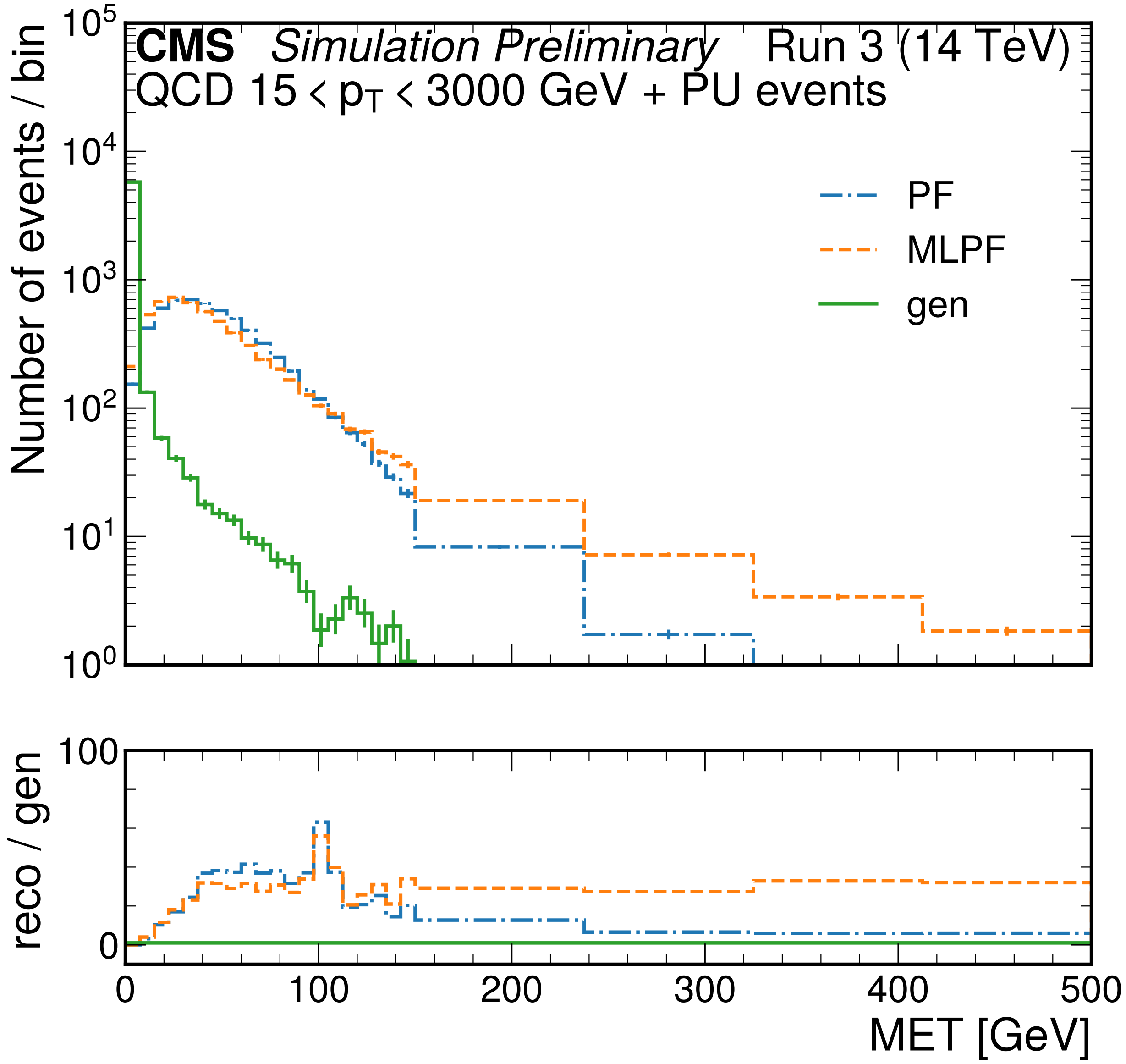
Performance (Jets)

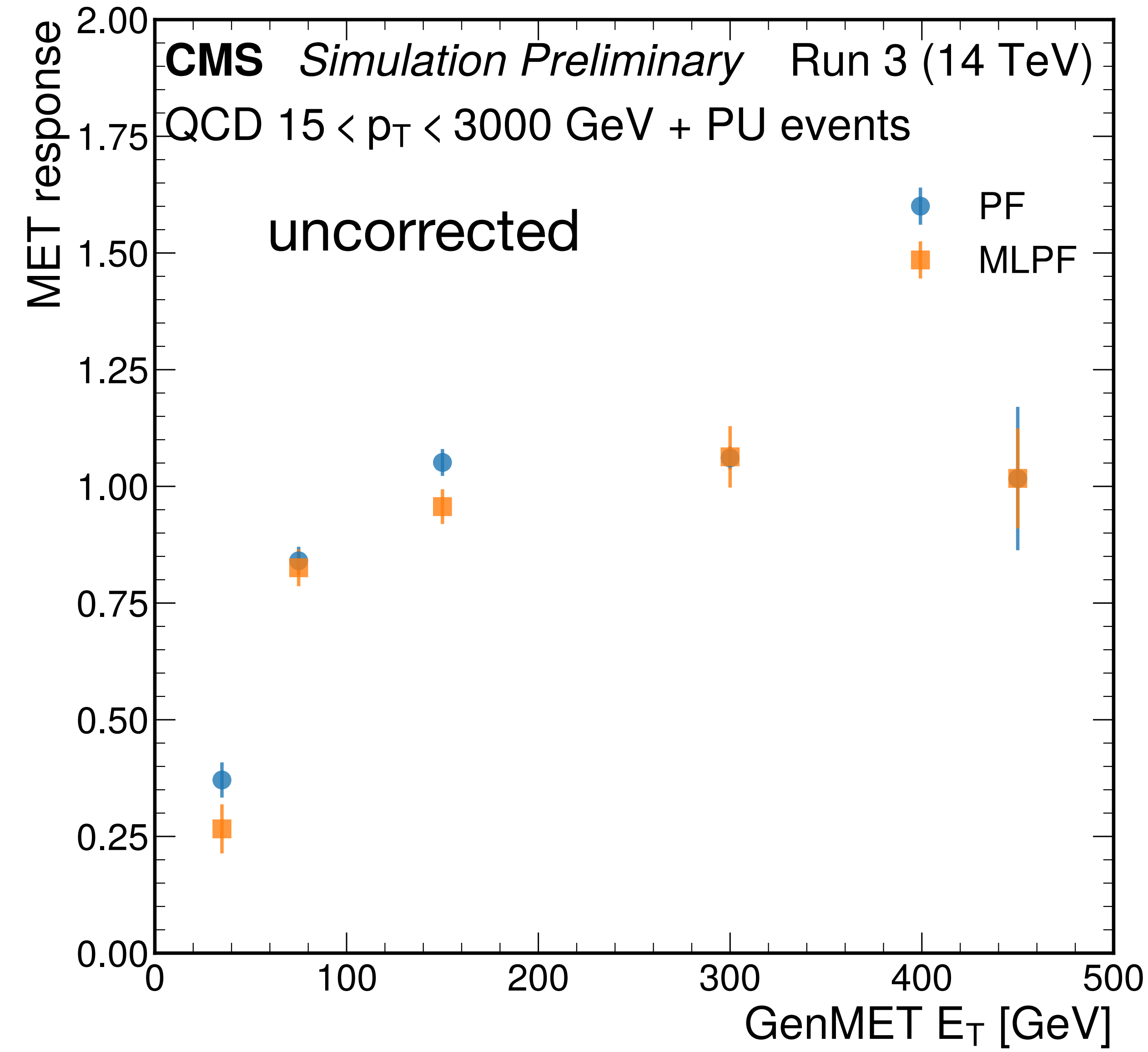






Performance (MET)





- Some large MET tails from MLPF (observed also with MLPF v1)
- Appears to originate from many nearby inputs all from same truth particle

Performance (MET)