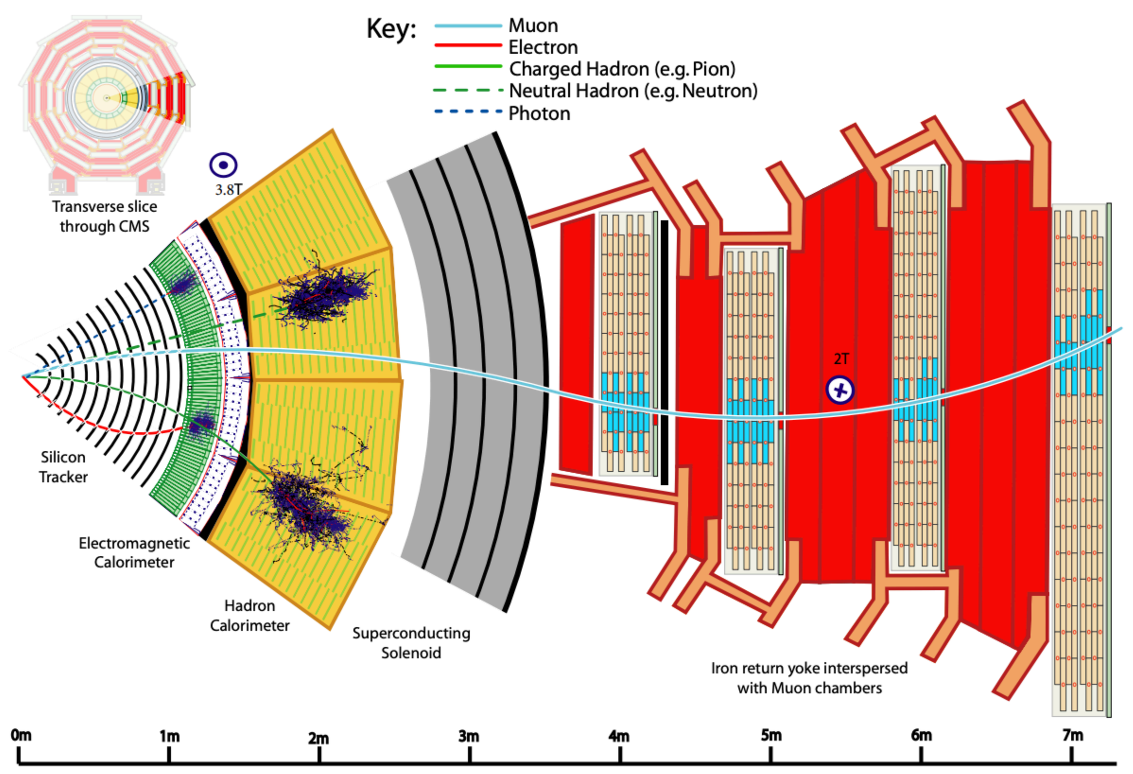
Progress towards an improved particle-flow algorithm at CMS with machine learning

Joosep Pata[1], **Farouk Mokhtar[2]**, Javier Duarte[2], Eric Wulff[3], Maurizio Pierini[3], Jean-Roch Vlimant[4]

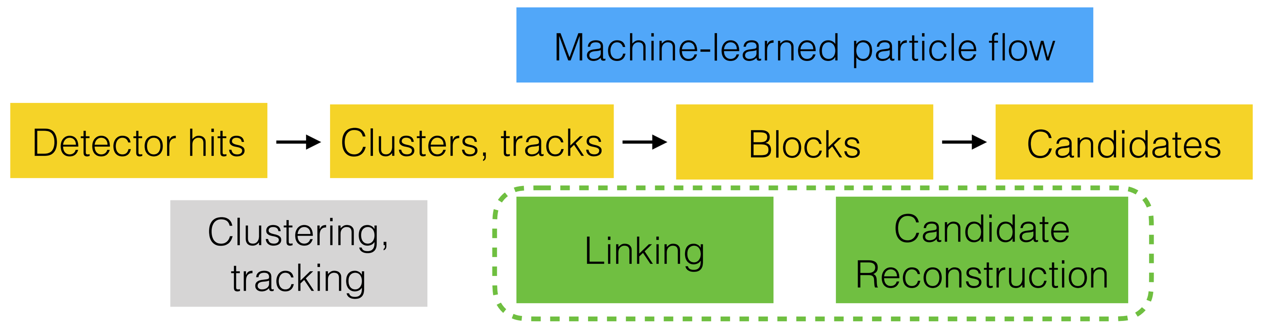
[1]National Institute of Chemical Physics and Biophysics, [2]UC San Diego, [3]CERN, [4]California Institute of Technology

PF and MLPF

- Particle-flow (**PF**) reconstruction is a global event reconstruction that combines information from calorimeter **clusters** and **tracks** to reconstruct stable particles.



- Machine-learned particle-flow (**MLPF**) algorithm is a graph neural network trained to perform particle-flow (PF) reconstruction via supervised learning.
- MLPF does the linking of tracks and clusters, after clustering/tracking has been performed. See figure below.



- Advantages of MLPF include the possibility of deployment on heterogeneous computing accelerators (e.g. GPUs) and reoptimizing the algorithm in light of new experimental conditions.
- **We can now train MLPF in CMS on a gen/sim-level target (i.e. without referencing an existing PF algorithm)** and get results that are largely compatible, and in some cases better, than standard PF.

Datasets

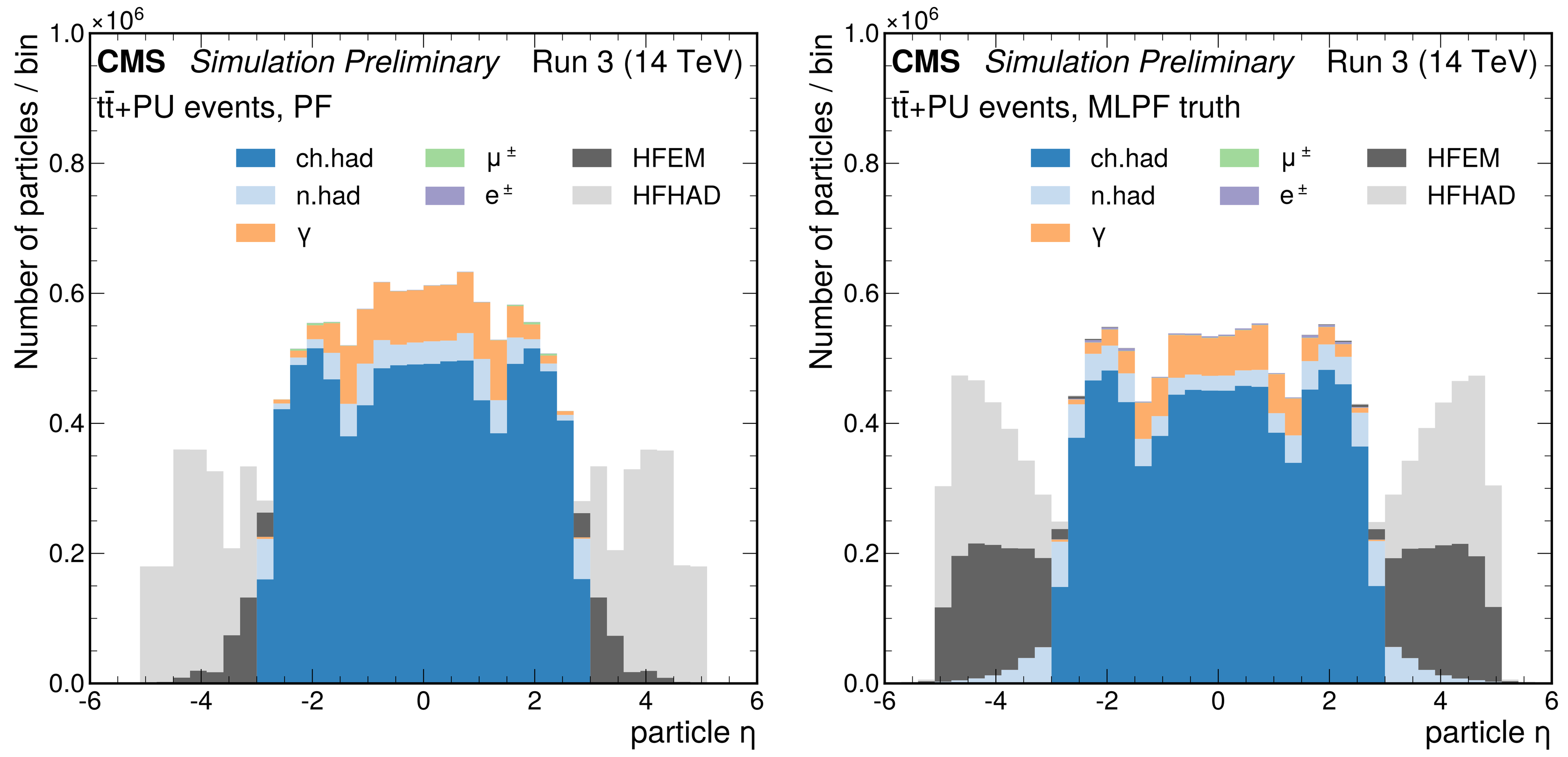
- Datasets are generated using CMSSW under Run 3 conditions.

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.

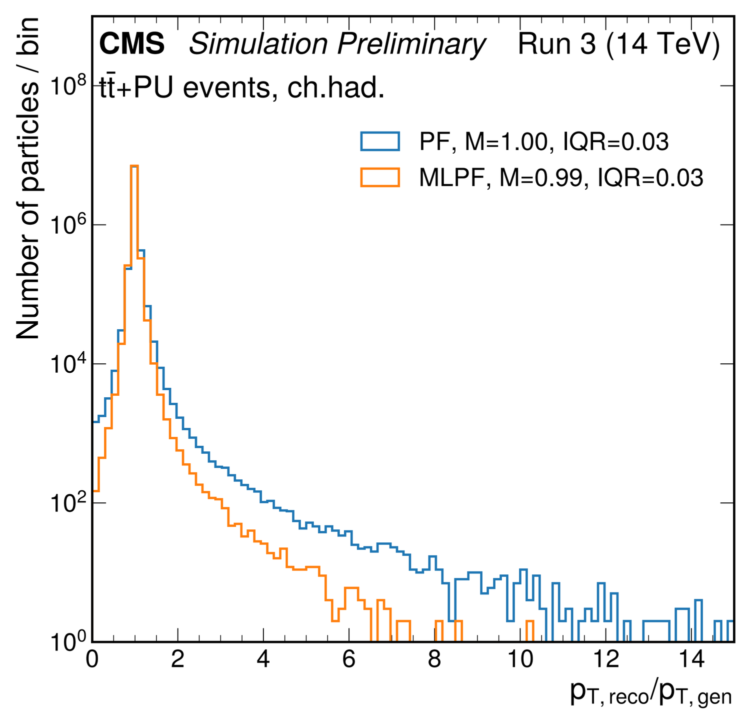
MLPF truth definition

- MLPF training truth defined based on detector simulation information to closely approximate the input the simulation receives from the generator.



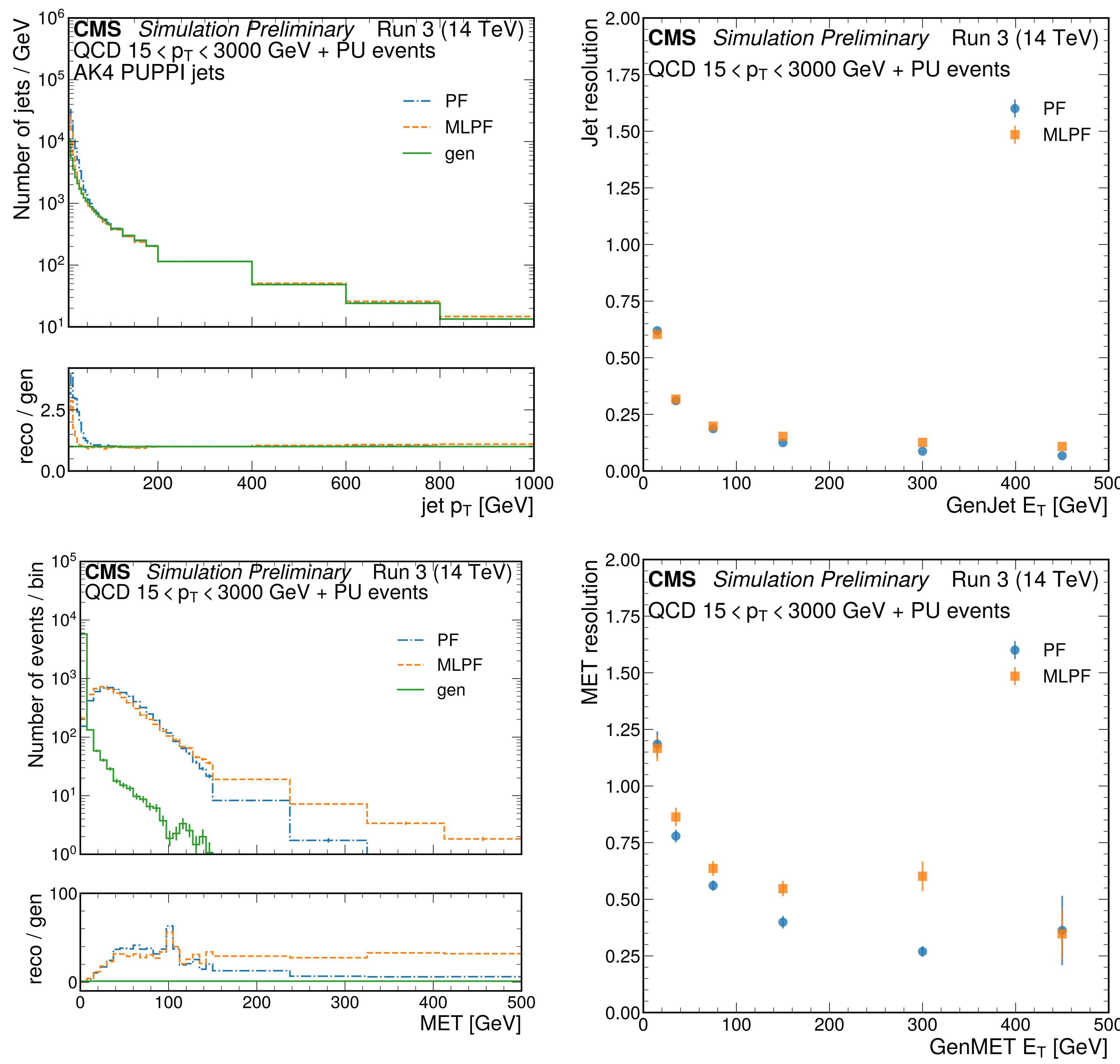
Particle-level validation

- Comparable response for charged hadron reconstruction between MLPF and PF, with slightly better efficiency/fake-rate and p_T resolution for MLPF.



Event-level validation

- Reconstruct PUPPI jets with either PF or MLPF and compare them with gen-level jets: full distributions and response relative to gen-level.
- Comparable performance between PF and MLPF for jet and MET reconstruction (shown for the QCD+PU sample).



Event loss scans

- Compare the usefulness of additional event-level loss terms in improving the reconstruction of jets and MET.
- Baseline approach (no additional event loss term) performs best.