Semi-supervised GraphNN for PUPPI

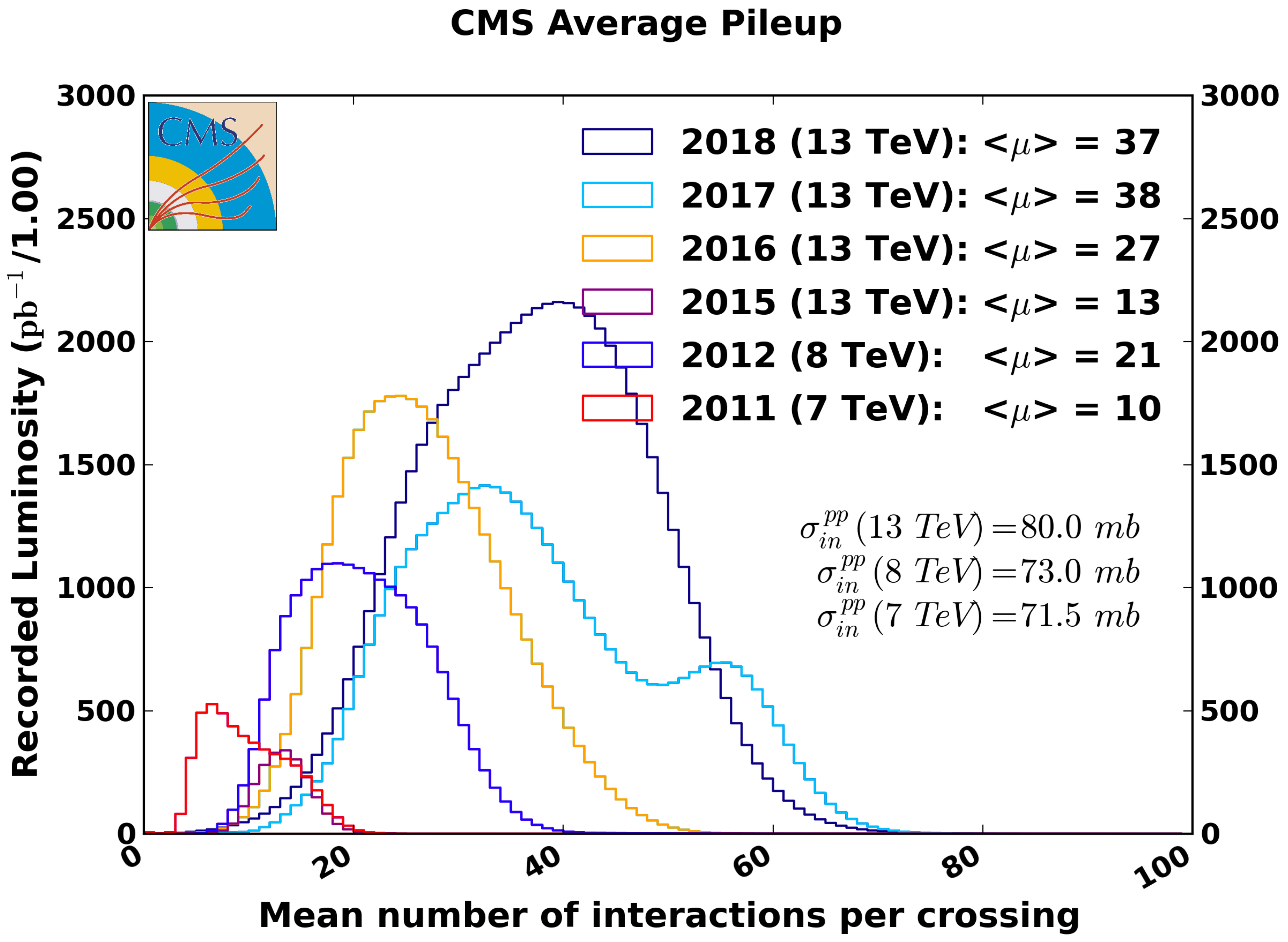
Shikun Liu, Tianchun Li, Pan Li, Miaoyuan Liu (Purdue)

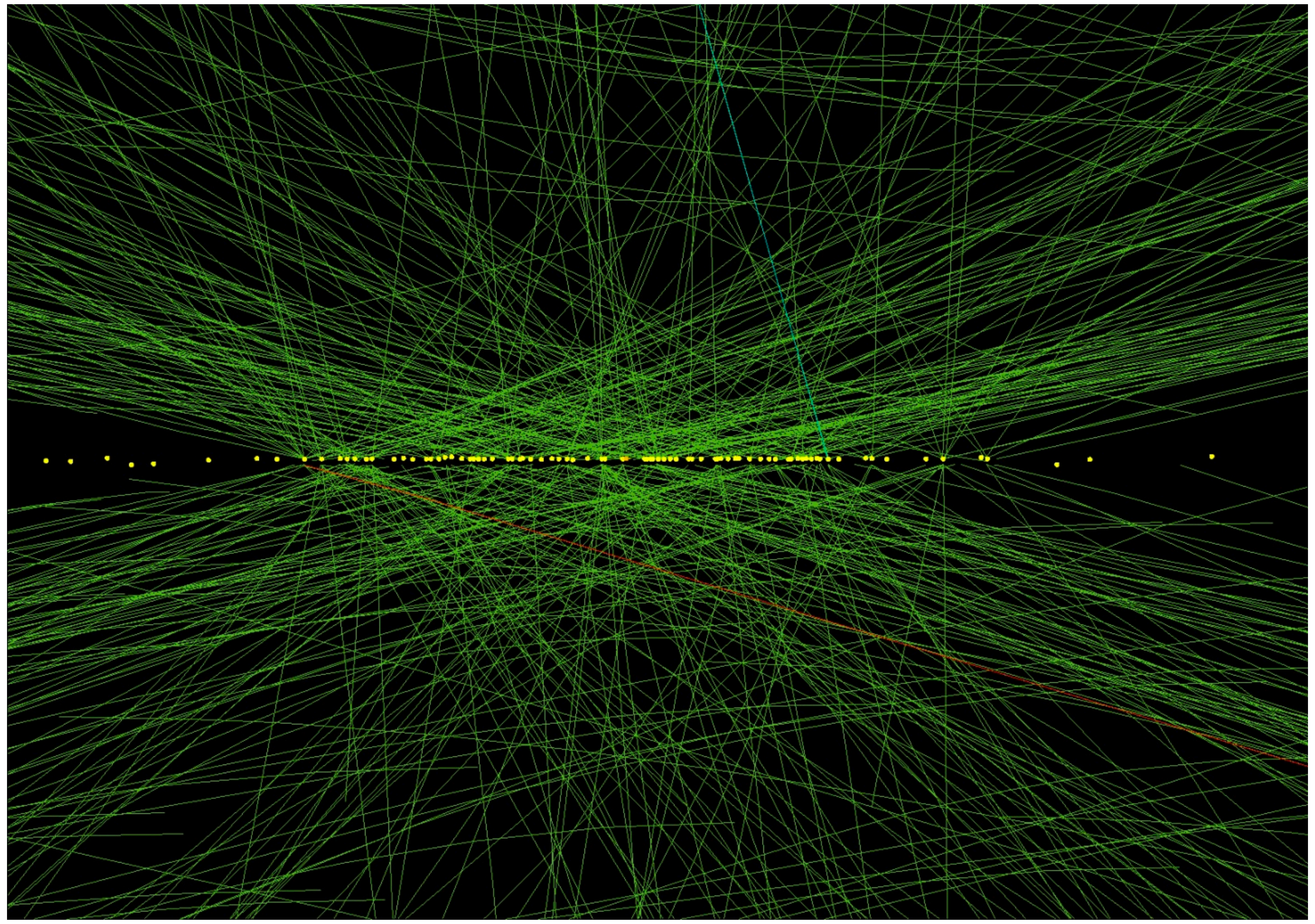
Yongbin Feng, Nhan Tran (Fermilab)

BOOST 2021, Virtual

August 2nd, 2021

Introduction: PileUp Mitigation



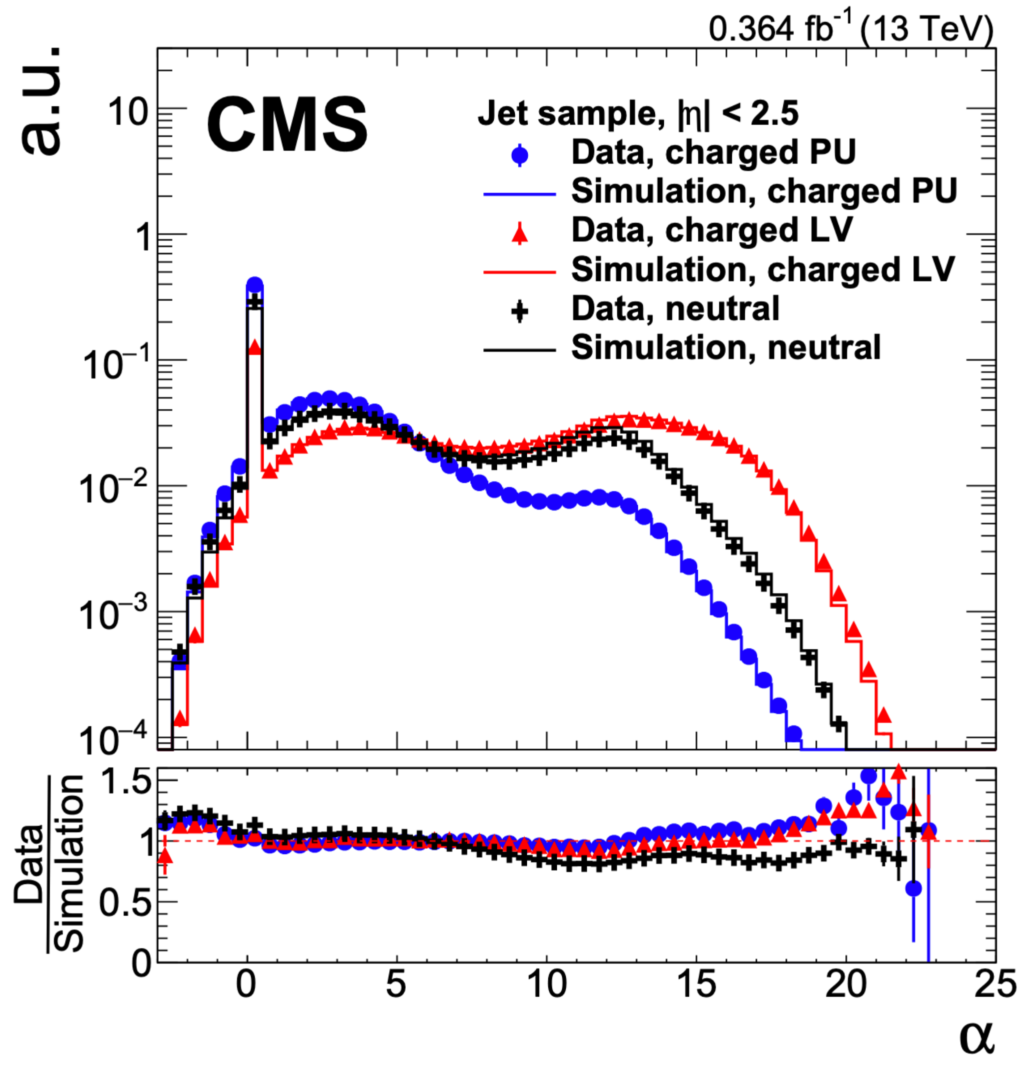


- **Pileup (PU): additional proton-proton interactions in the same or nearby bunch crossings**
  - PU at run2: ~30-40; and is expected to be around 150 at HL-LHC
  - Pileup can significantly affect the reconstruction of many physics variables - jet mass, jet momentum, missing transverse momentum (MET)
  - Pileup mitigation is needed

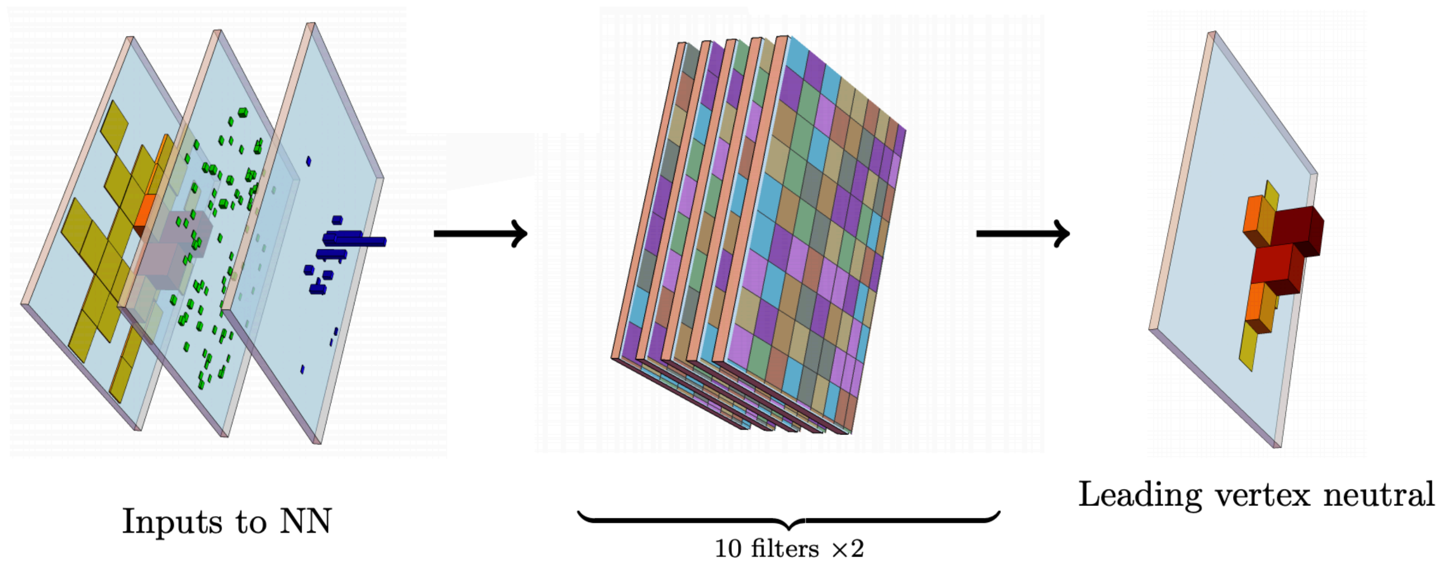
Introduction: Pileup Mitigation Studies

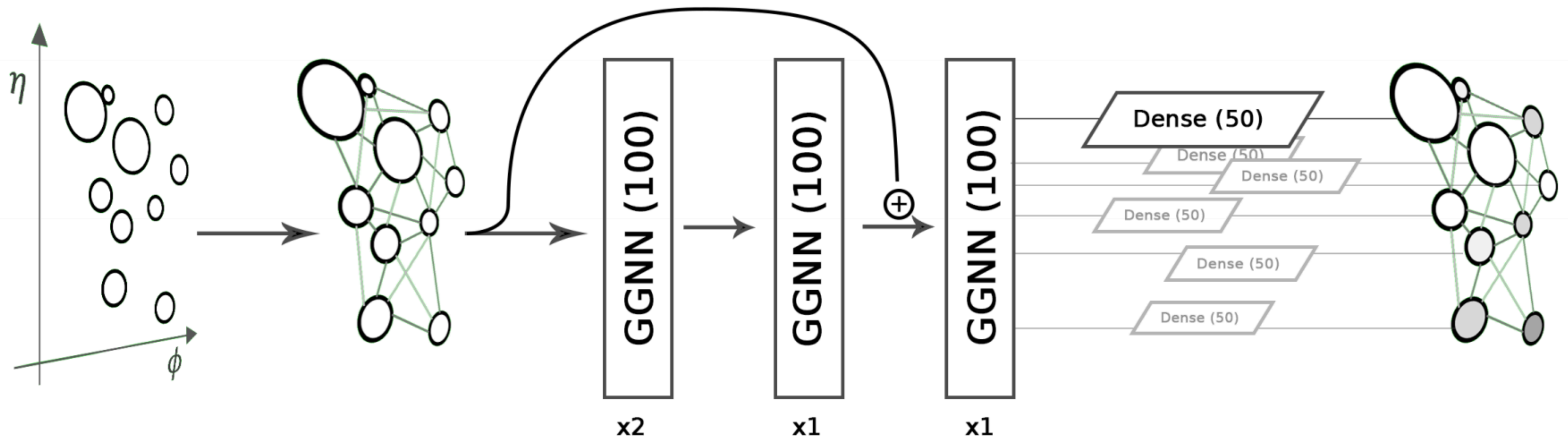
- Most charged leading vertex (LV) and pileup (PU) particles can be identified (**Charged hadron subtraction**, CHS)
- Particle self features: PU particles have lower-pT -> **SoftKiller**
- Particle neighboring features: **PUPPI**. A local shape variable α is defined and PUPPI weights are calculated based on α

$$\alpha_i = \log \sum_{j \in \text{event}} \xi_{ij} \times \Theta(R_{\min} \leq \Delta R_{ij} \leq R_0), \quad \text{where } \xi_{ij} = \frac{p_{Tj}}{\Delta R_{ij}}.$$



- Pileup Mitigation with ML (**PUMML**, arxiv 1707.08600): Convolutional neural network on jet image
- Gated Graph Neural Network for PUPPI (**PUPPIML**, arxiv.1810.07988): GGNN on particle graph
- Other graph/attention models: (**ABCNet**, arxiv.2001.05311), (**PUMA**: with transformer, arxiv.2107.02779)

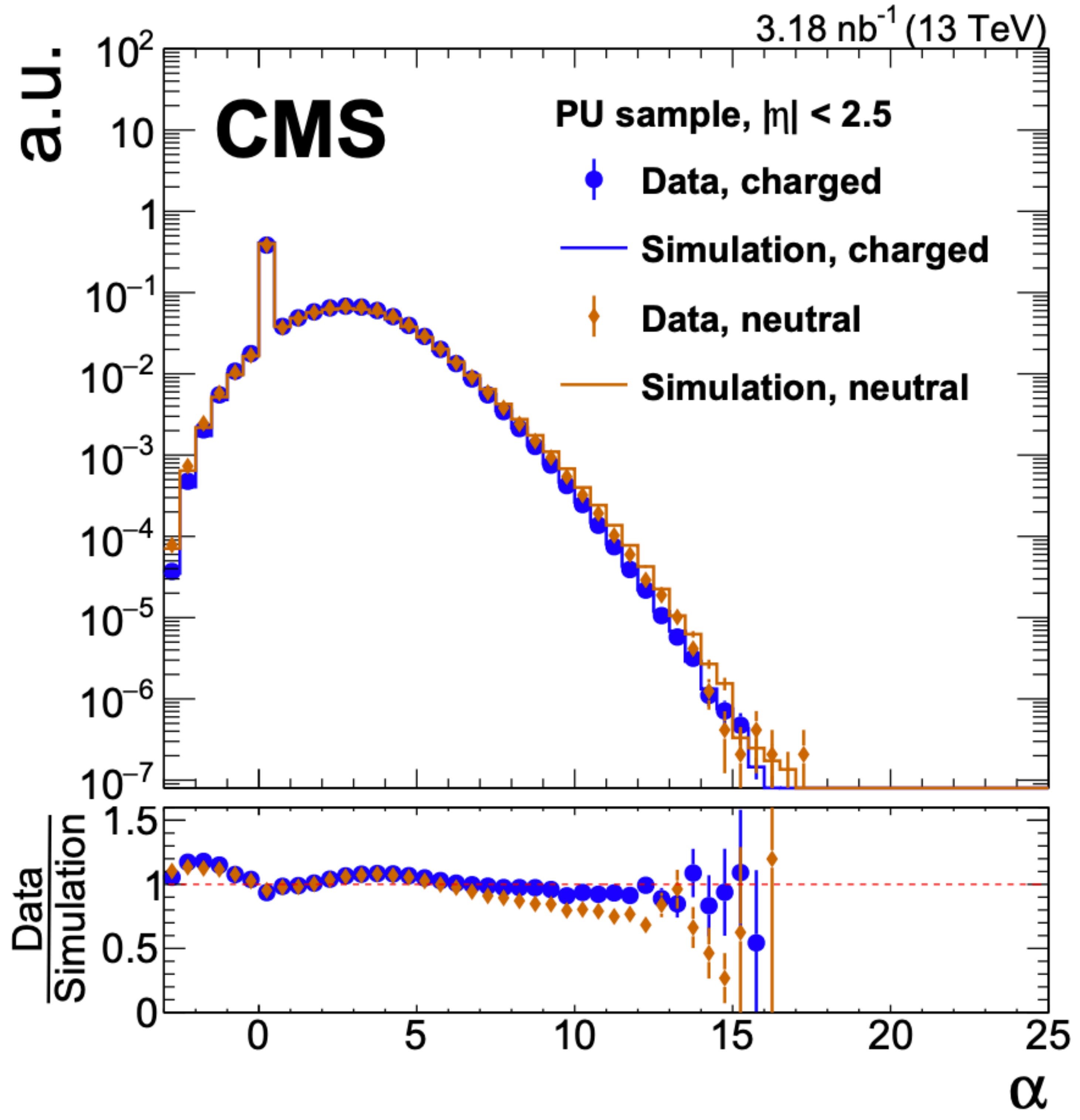




Semi-Supervised Learning

- **To operate at particle level, the current ML models would require the prior knowledge of whether the particle is produced from PU or LV, as the ground truth information**
  - For charged particles, it is easy to retrieve, even in the real data
  - For neutral particles, currently very hard to recover truth information, sometimes mixed LV/PU; no truth information in the real data

- **How about we train the model using the charged particles, and then do inference on neutral particles?**
  - This semi-supervised ML method would allow us to train directly on real data/full simulation, without worrying about the labels for the ground truth information
  - This semi-supervised training strategy would work on different ML models and architectures



Training Datasets

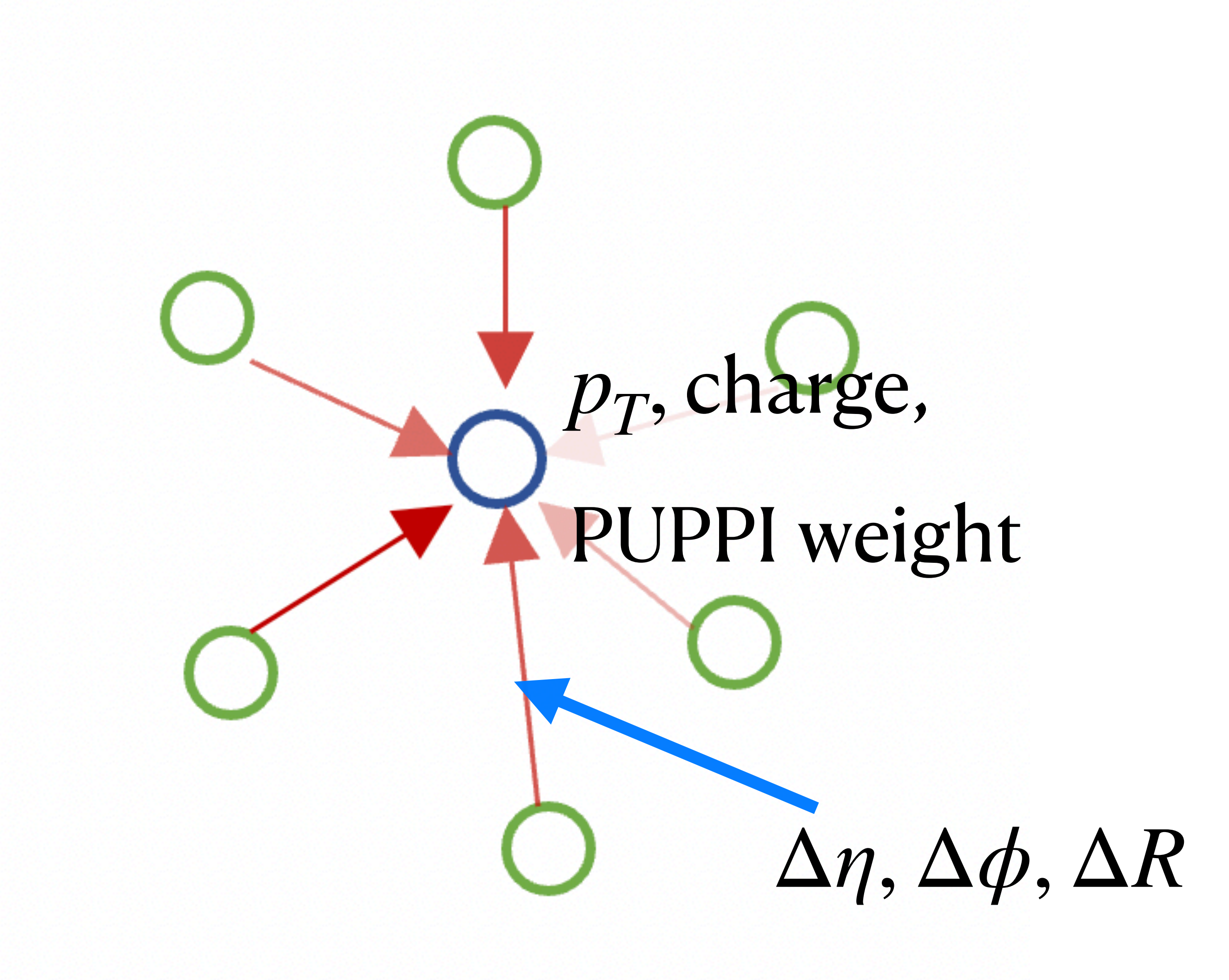
- **Using the PUPPIML datasets as our training datasets (Many thanks!)**
  - Pythia 8.223 + Delphes 3.3.2 for simulation
  - $Z(\nu\nu)$+jets signal processes
  - Pythia-generated QCD events as pileup; Poisson distribution sampled with the average pileup of 80 and 140
  - Charged particle flag for the LV and PU is set to be perfect
- **Number of different particles per event at PU=80 (with pT>0.5GeV cut)**

# Particles	Charged	Neutral
LV	85	50
PileUp	1600	800

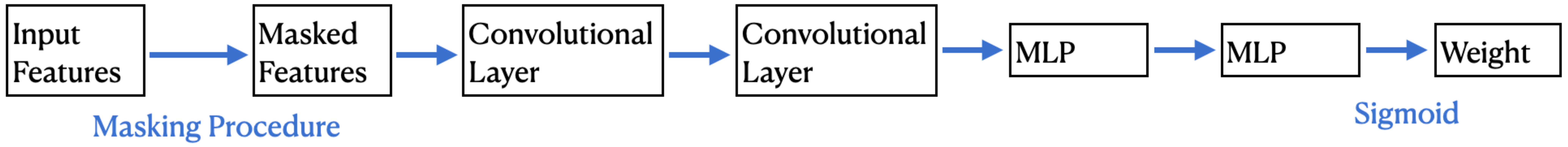
Model Architecture

- Build graph in $\eta - \phi$ space. Connect the particles in the $\Delta R = 0.8$ cone.
- Input features are the p_T , charge, and PUPPI weights for the nodes, and the $\Delta\eta$, $\Delta\phi$, and ΔR between particles for the edges
- Outputs are a weight between 0 and 1, representing the probability that the particle is produced from the LV. Also study the **'Hybrid' algorithm**:

$$\text{Final Score} = \beta \cdot \text{GNN Output} + (1-\beta) \cdot \text{PUPPI Weight, where } 0 < \beta < 1$$



- Model architecture:



✤ Convolutional layer is the convolution on graphs. Here we use the **gated model**:

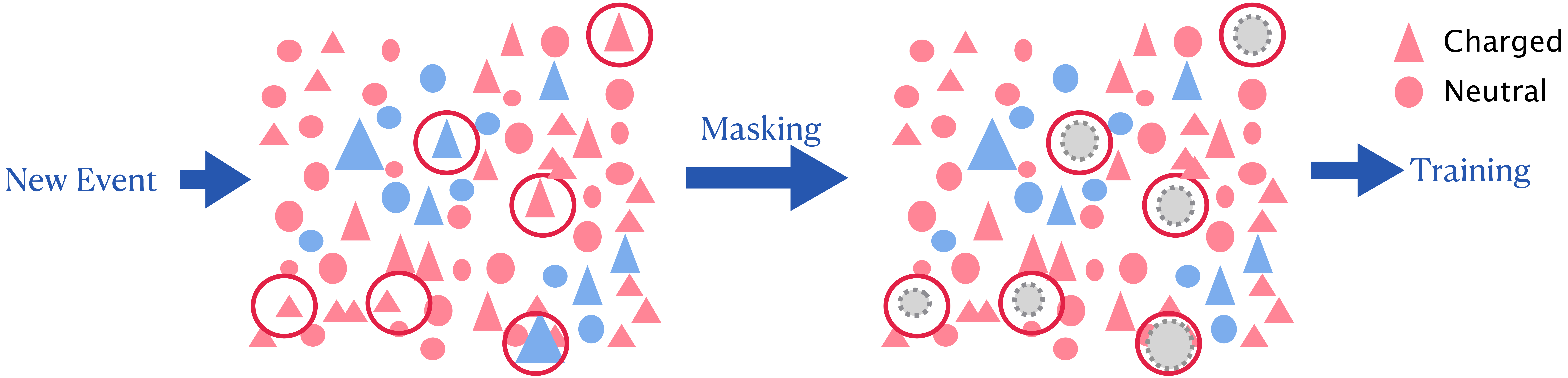
$$h_u^{k+1} = G_u^k \cdot h_u^k + (1 - G_u^k) \cdot M_u^k \quad \text{where } G_u^k = \text{Sigmoid}(h^k \oplus M_u^k) \text{ is the gate controlling the node feature updates}$$

$$\text{and } M_u^k = \frac{1}{N} \sum_v (G_{u,v}'^k \cdot M_{u,v}^k),\ G_{u,v}'^k = \text{Sigmoid}(M_{u,v}^k) \text{ is the gate controlling feature passing from neighbor to the node}$$

✤ Also tested GraphSage (ArXiv:1706.02216) for the convolution, and compare the performances

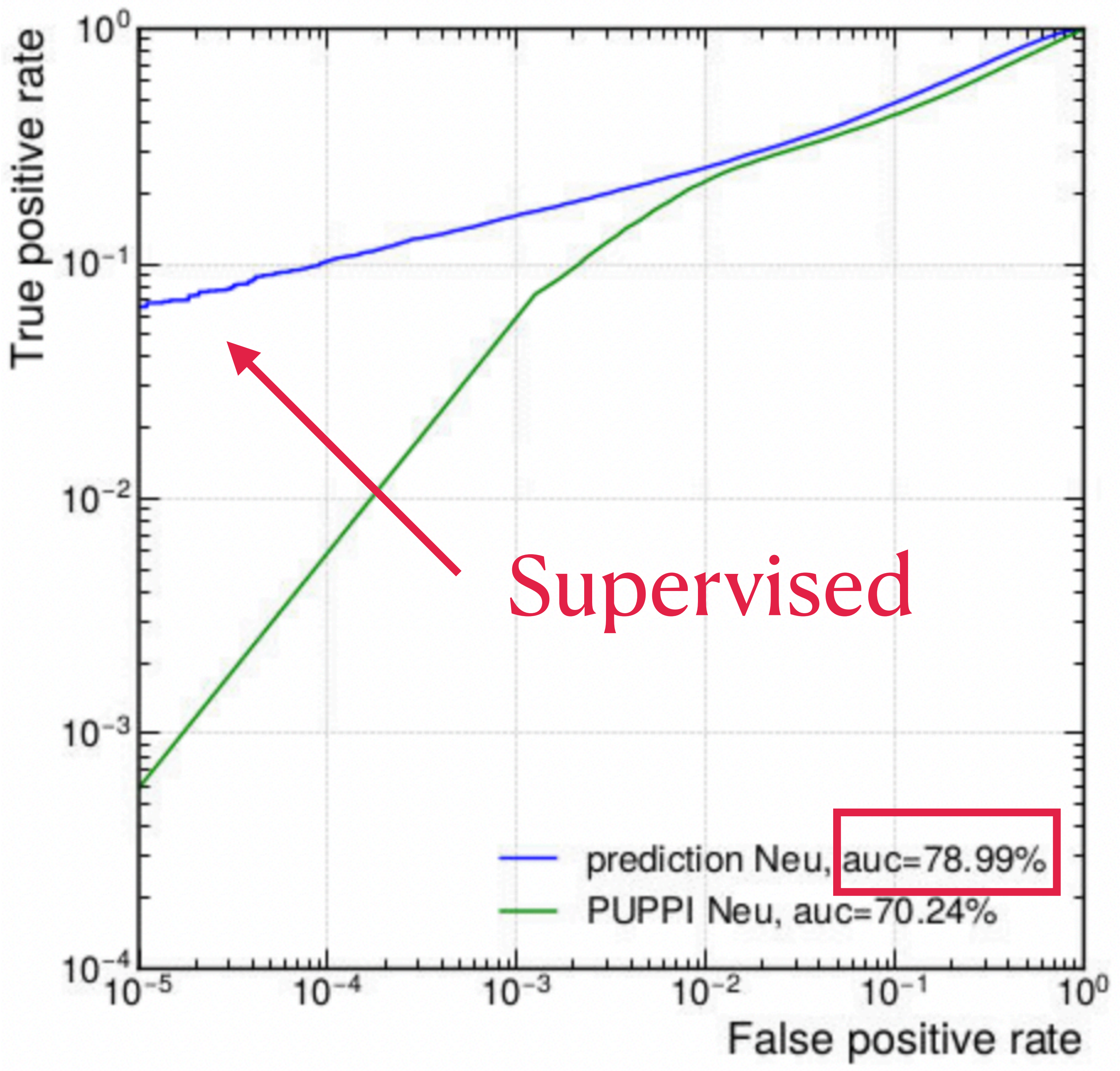
Masking & Training Details

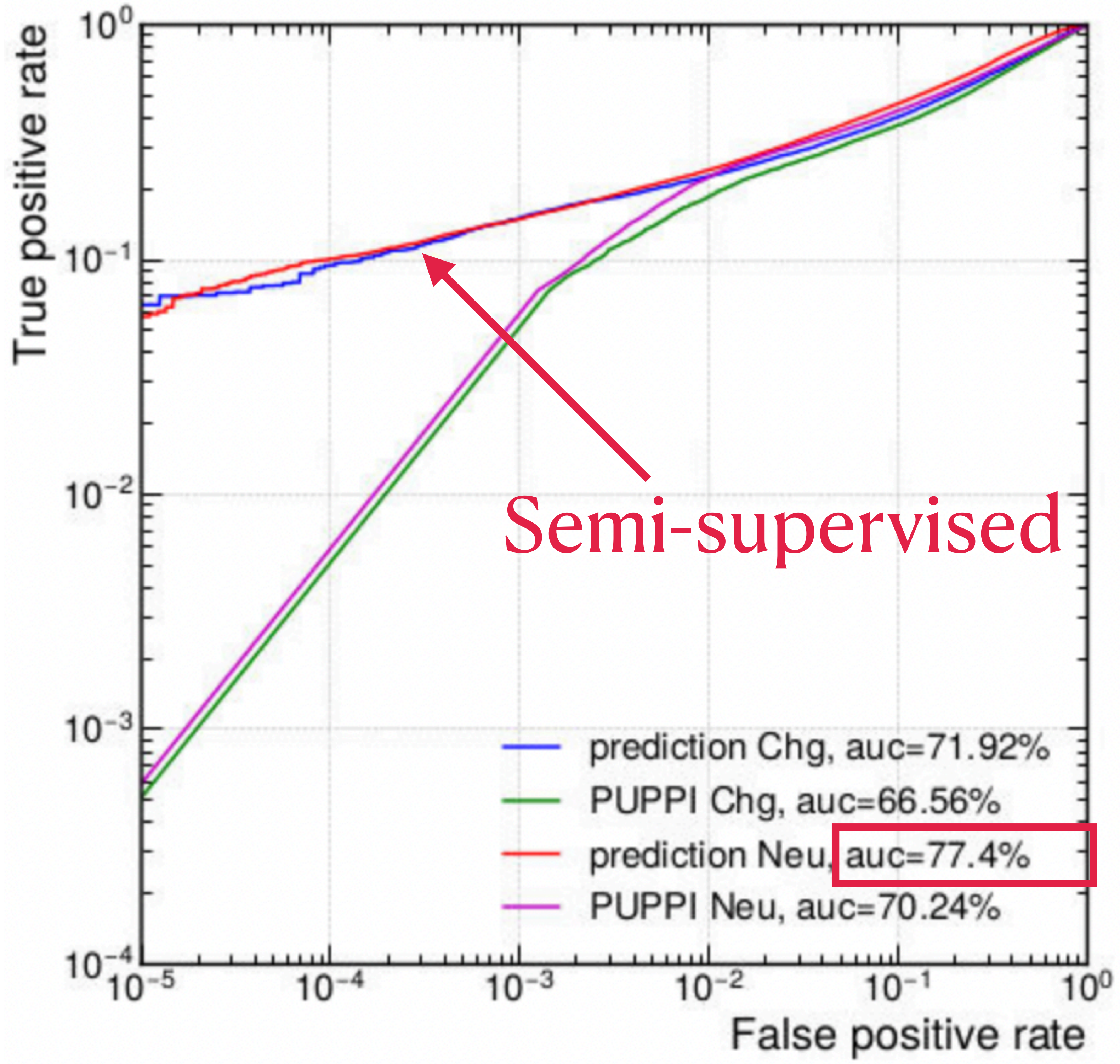
- Total number of trainable parameters is O(1K).
- At PU=80, uses 3K events for training.
- For each event, randomly select 10 LV charged particles and 160 PU charged particles for training. (About the same fraction as the original dataset.) Mask their charged related features so that GraphNN treats them as 'neutral'. Train on these particles.
- Binary cross entropy loss function estimated on these masked particles.



- Repeat the same previous two steps on the next event. Loop over these 3K events until it converges. Usually it takes about ~5 loops in total to converge (Takes about a few hours)

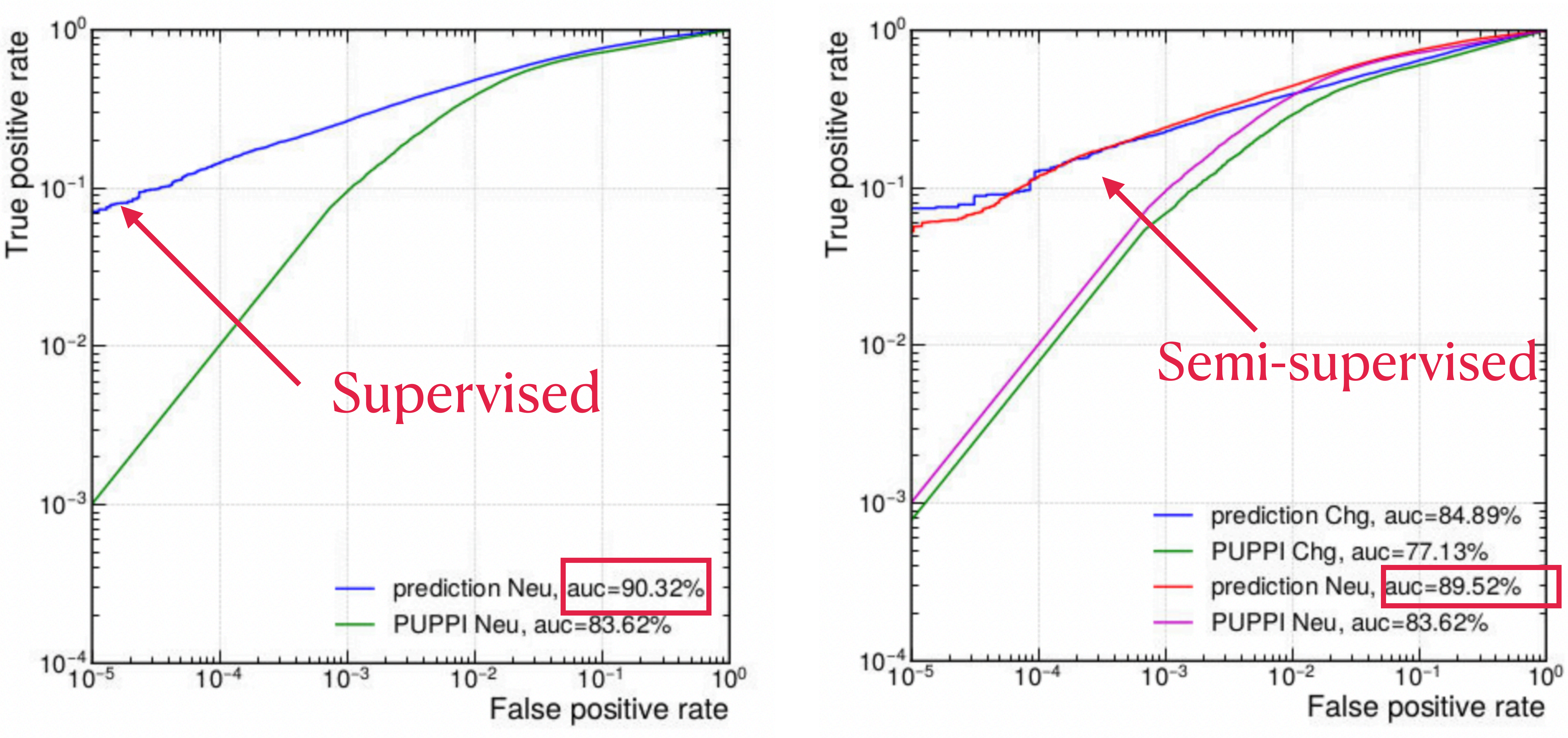
Performance at PU=80





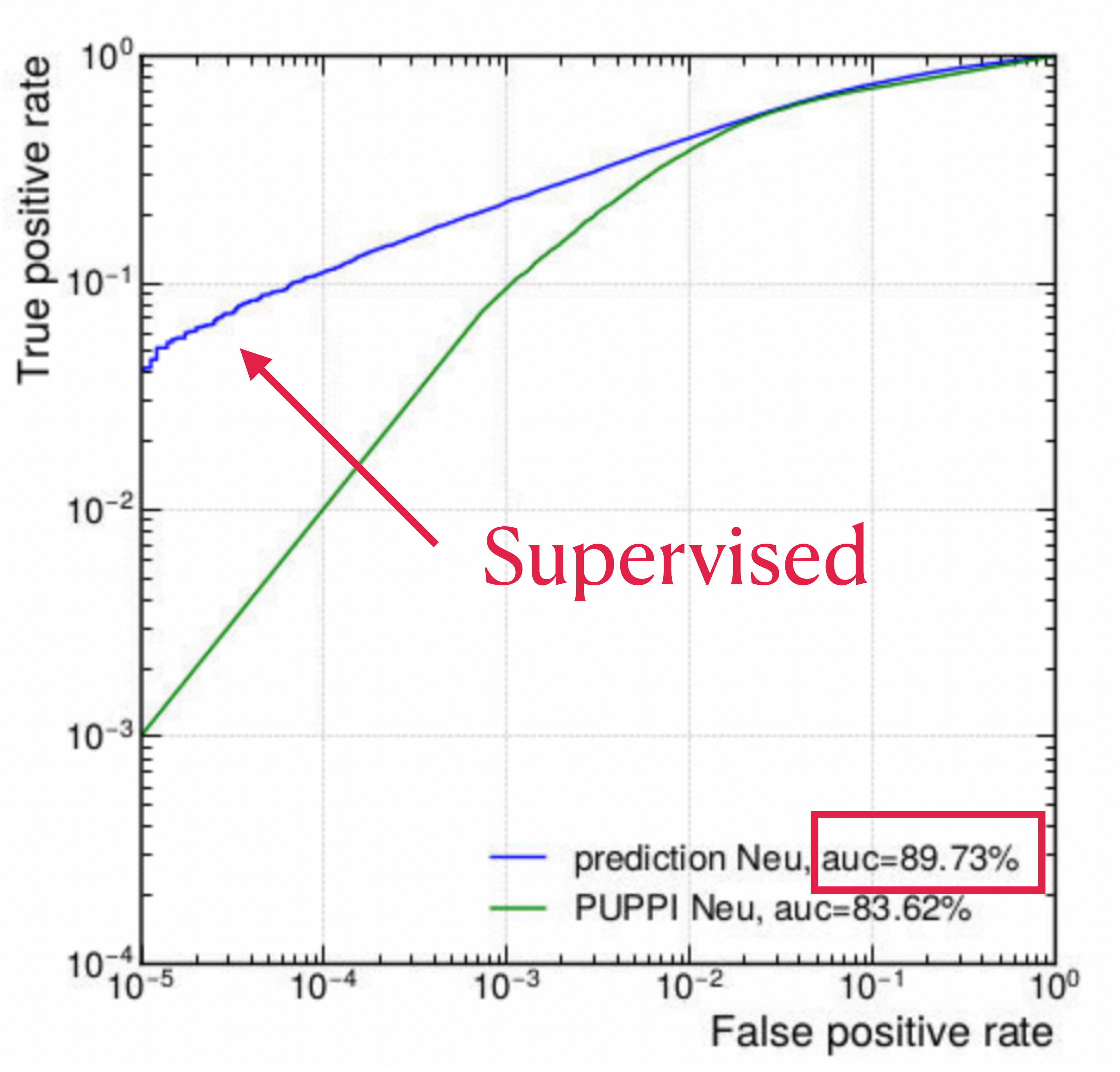
- Evaluate the performance of boosted model with hybrid on the PU=80 samples. (training and testing both at PU80)
- Left plot is the results of supervised training. Right plot is the results of semi-supervised training
- Both supervised and semi-supervised have better performance than PUPPI
- Performance maintains comparing supervised with semi-supervised

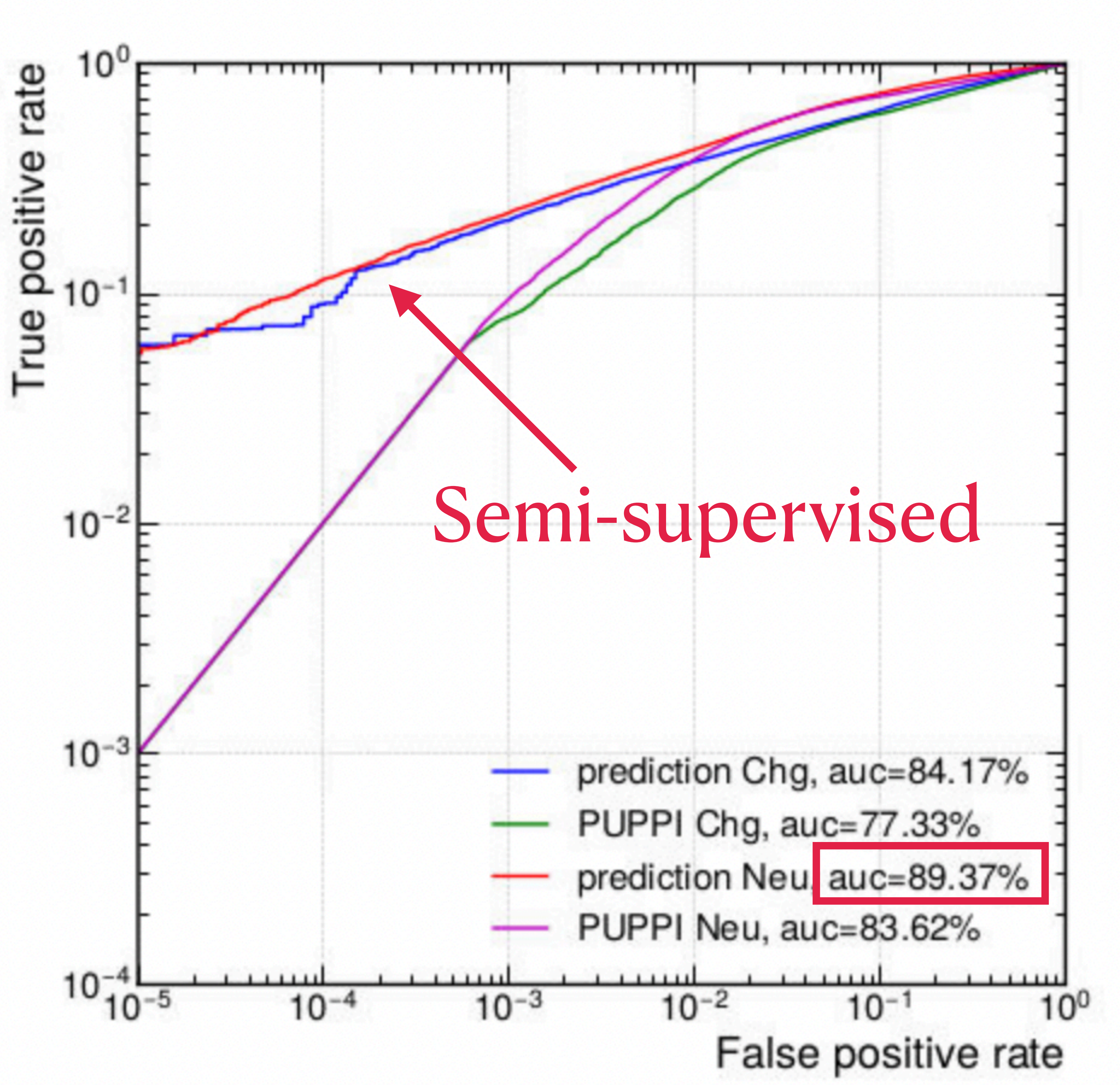
Performance at PU=140



- Evaluate the performance of boosted model with hybrid on the PU=140 samples. (training and testing both at PU140)
- Both supervised and semi-supervised have better performance than PUPPI
- Performance maintains comparing supervised with semi-supervised

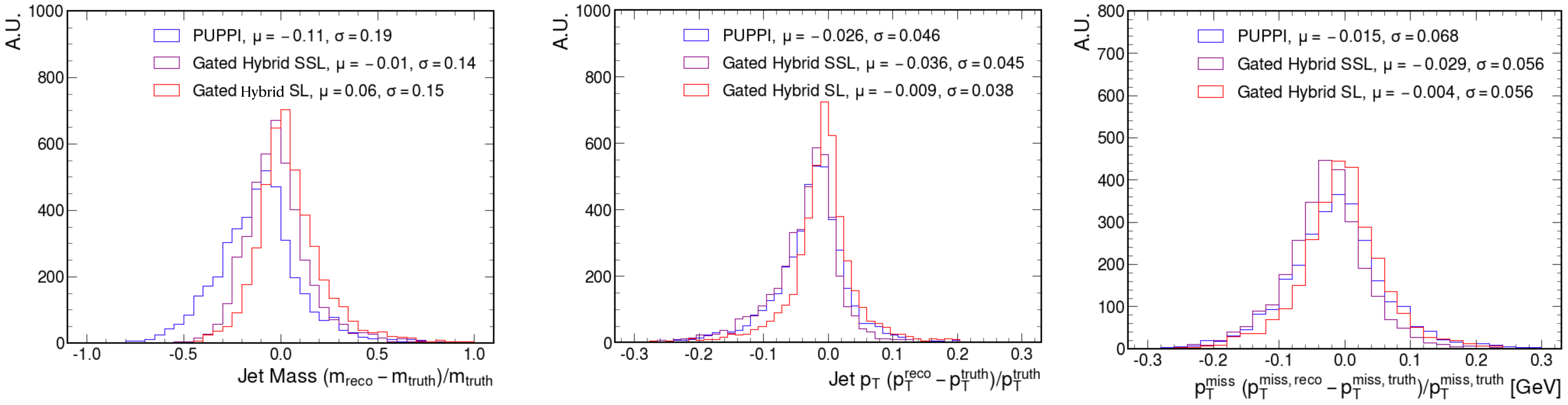
Model Generalization from PU=80 to PU=140





- Train on PU=80, evaluate the performance at PU=140
- Both supervised and semi-supervised have better performance than PUPPI
- Comparing with the previous results in the previous slide, the performances maintains when apply the PU80 training on PU140

Performance on Jet Mass, pT, and MET (PU80)



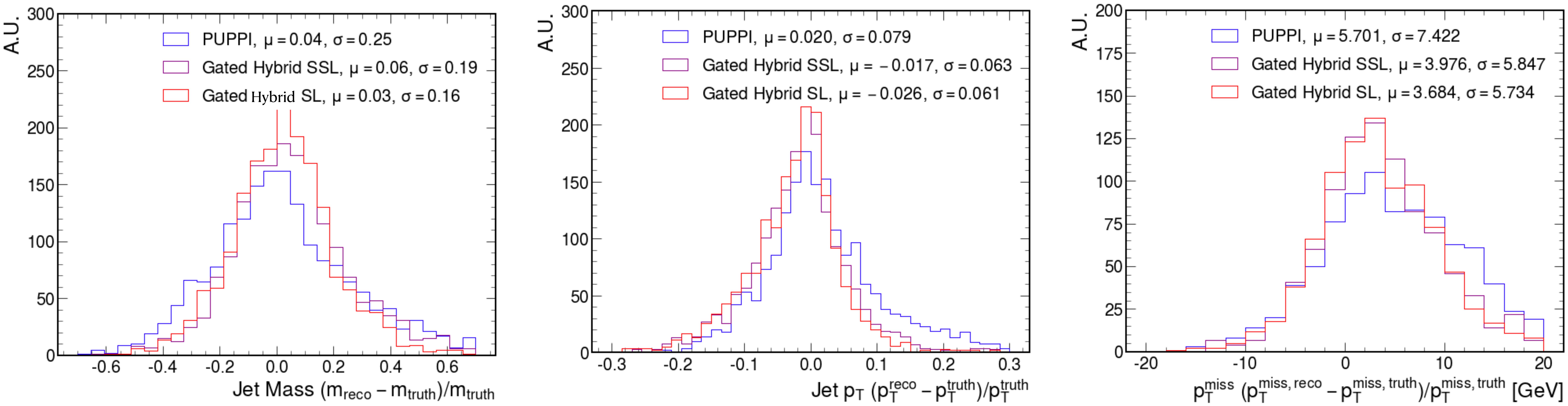
- **Comparison of the jet mass, jet pt, and MET resolutions, at PU=80:**
  - Jets are reconstructed using anti-kT algorithm with deltaR=0.7; truth jets are defined as jets reconstructed using particles from the LV at the truth level; PUPPI jets and GNN predicted jets are jets reconstructed using rescaled particle momentum with PUPPI weights/GNN outputs
- **The resolutions look to be ~20% better comparing supervised GNN with PUPPI**
- **Going from Supervised learning to Semi-supervised learning, the performance drops slightly- still better than PUPPI**

Performance on Jet Mass, pT, and MET

	PUPPI	Supervised Gated Hybrid	Gated Hybrid	Gated NoHybrid	Supervised GraphSage	Graphsage Hybrid	Graphsage NoHybrid
AUC (%)	71.8	77.3	78.6	78.5	77.3	77.8	77.3
Jet mass Bias	-0.002	0.058	-0.011	0.037	0.050	0.063	-0.028
Jet mass Resolution	0.195	0.148	0.144	0.158	0.152	0.164	0.180
Jet pt Bias	0.008	-0.009	-0.037	-0.033	-0.026	-0.015	-0.165
Jet pt Resolution	0.044	0.038	0.045	0.046	0.042	0.043	0.121
MET bias	0.010	-0.004	-0.029	-0.025	-0.010	0.004	-0.159
MET resolution	0.069	0.057	0.055	0.056	0.056	0.060	0.134

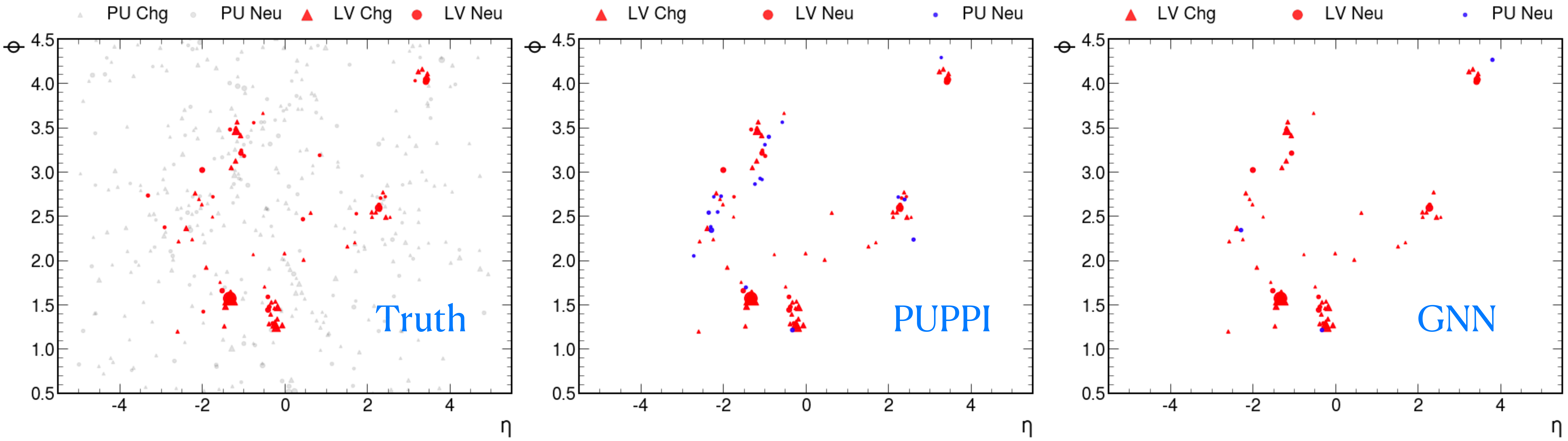
- Comparing the bias and resolution of different models, in general it looks the gated model with hybrid (both supervised and semi-supervised) has the best performance
- No significant performance drop going from supervised to semi-supervised for gated models hybrid with PUPPI

Performance on Jet Mass, pT, and MET (PU140)



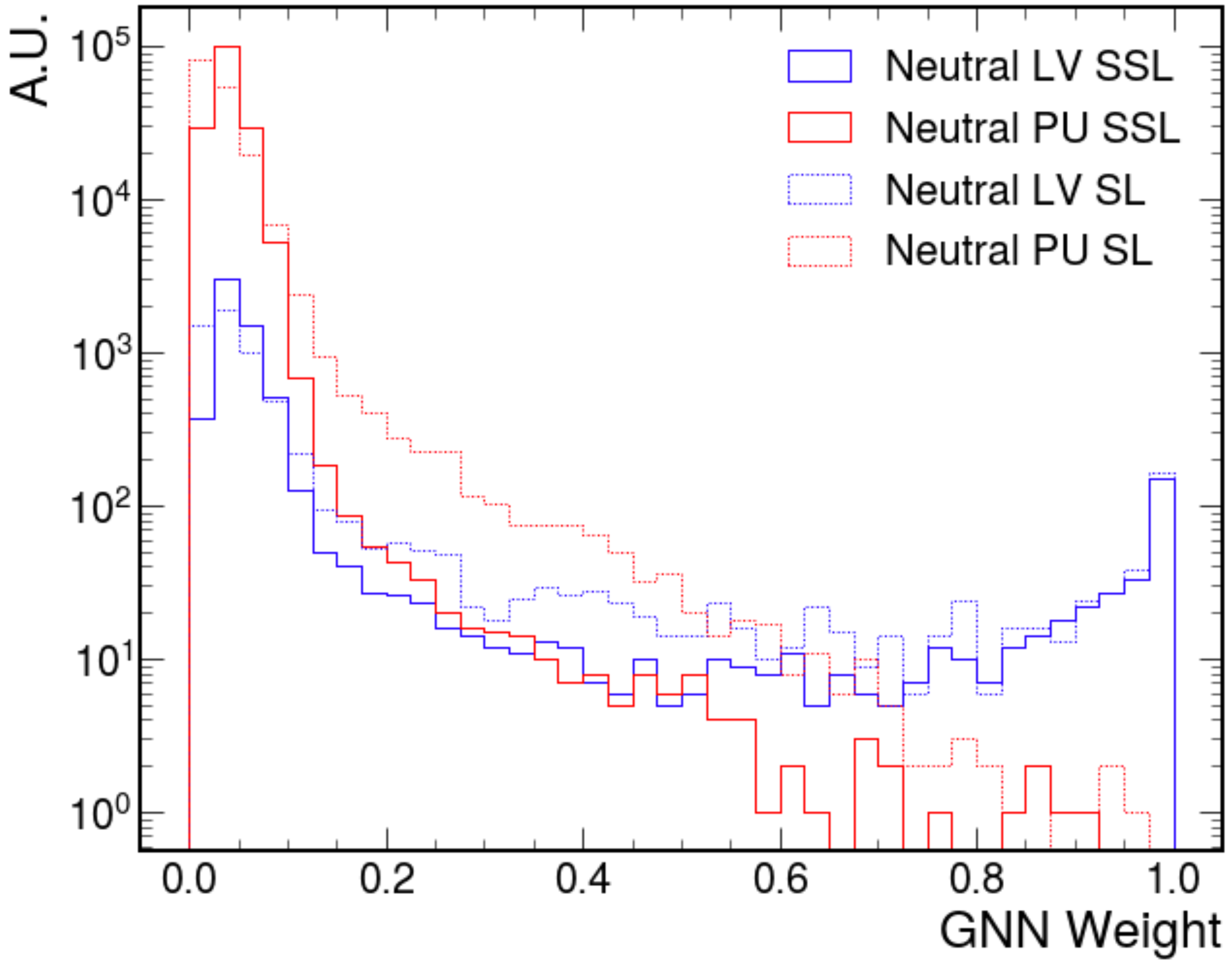
- Comparison of the jet mass, jet pt, and MET resolutions, at PU=140:
- Little performance drops going from supervised training to semi-supervised training.
- Large improvements compared with PUPPI

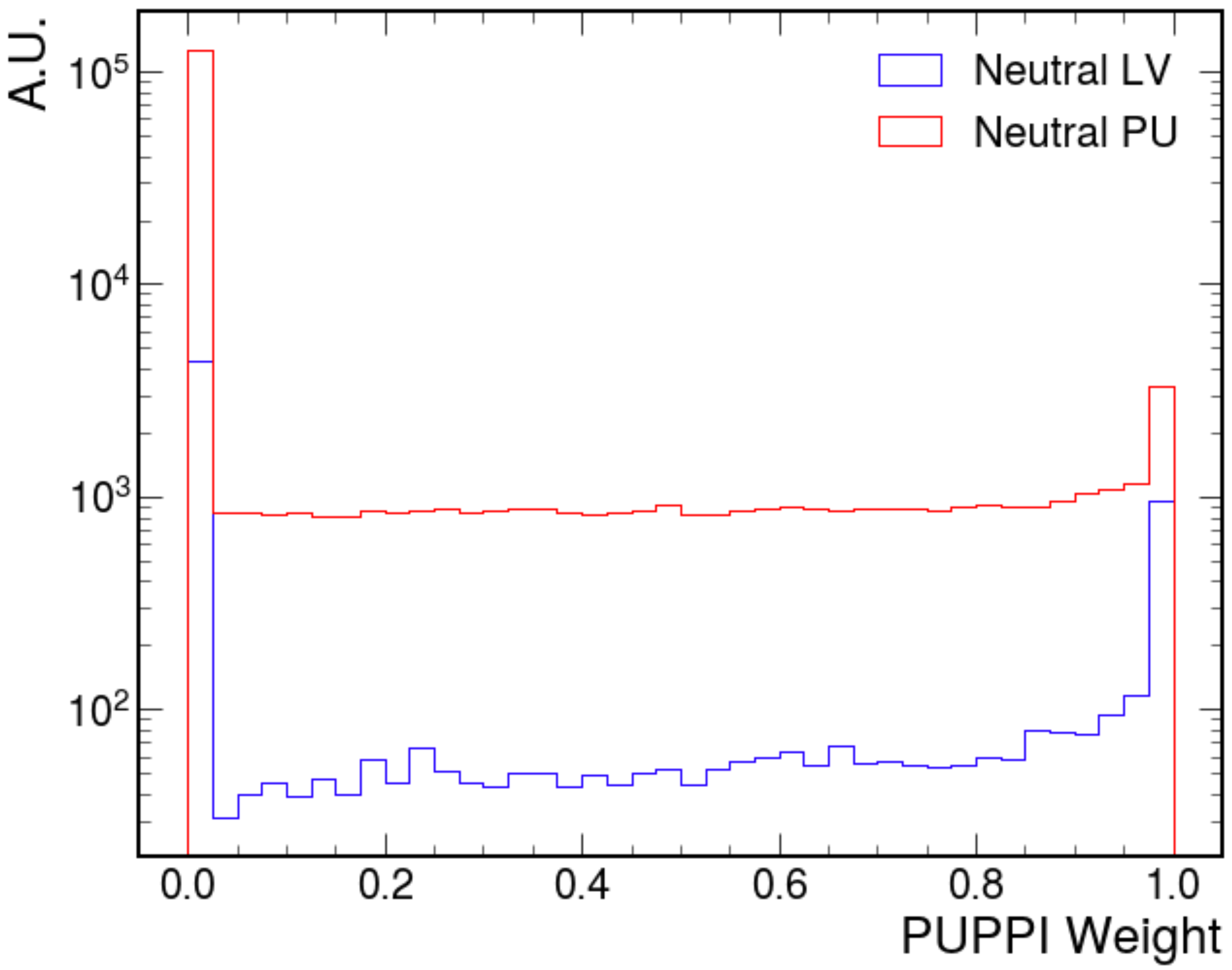
Event Display



- One example of the particle distribution in the $\eta - \phi$ space. Left is at the truth level. Middle is with PUPPI cleaning. Right is after semi-supervised GNN cleaning
- Comparing middle with right, GNN manages to remove more PU neutral particles close to the LV particles.

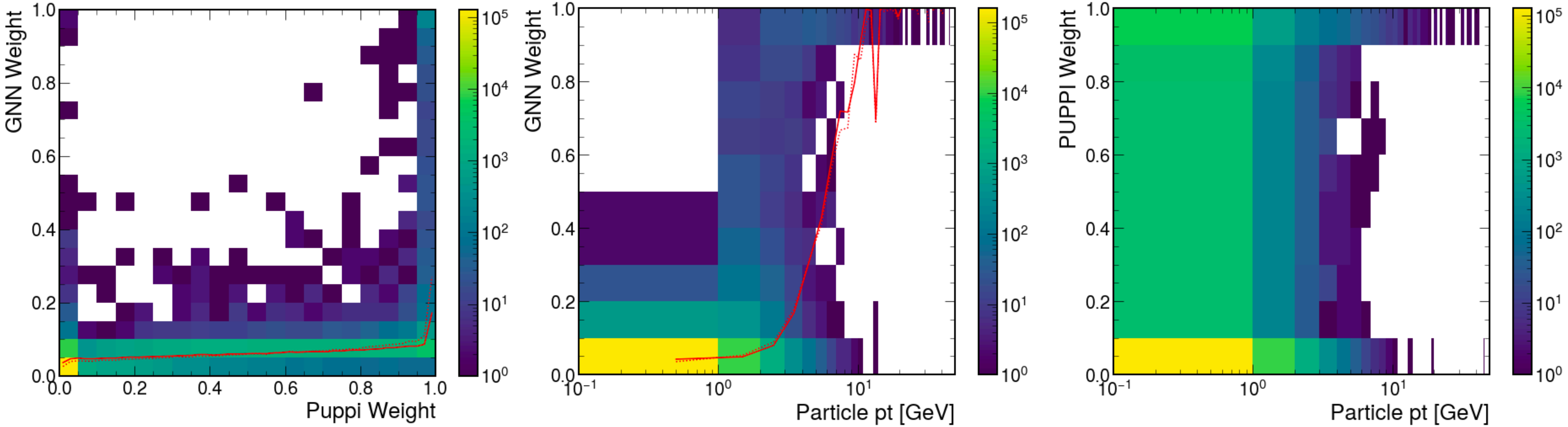
GNN Weights on Neutral Particles





- GNN Weights of neutral particles from the LV (blue) and pileup (red) on the left; right plot is the PUPPI weight distribution as a reference
- Much smaller fraction of particles get a weight around 1.
- Compared with Supervised training, the semi-supervised training seems to tend to have fewer particles in the middle weight range

GNN Weights on Neutral Particles



- GNN Weights vs PUPPI weights on the left. In the leftmost region, GNN assigns some particle higher weights while PUPPI weight is zero
- GNN weights vs Particle pt in the middle. For pT<1GeV, the GNN weight is lower; no particle gets assigned weight above 0.5 in this region

Summary

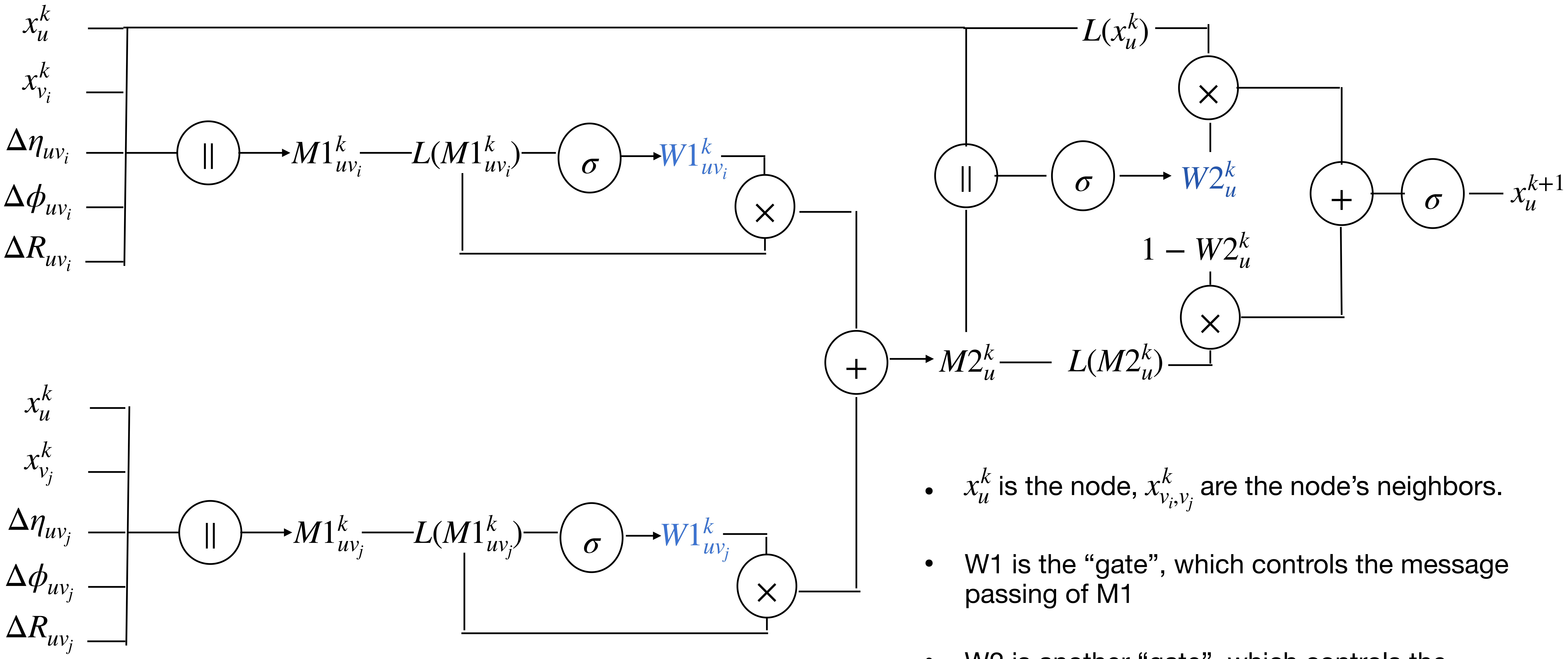
- Presented the study of applying **semi-supervised training** for pileup mitigation with GraphNN, where the training is done on charged particles, and the inference is on neutral particles

- Preliminary results look very promising:
  - Better ROC curve and resolutions on jet mass and MET for both supervised training and semi-supervised training
  - No significant performance drop going from supervised to semi-supervised

- Plan to train and test the performance of this technique on real data/full simulation. More work to do to fix problems such as:
  - Transfer learning to the forward region when there is no track;
  - Mixed final states with isolated leptons and photons, mixed samples.

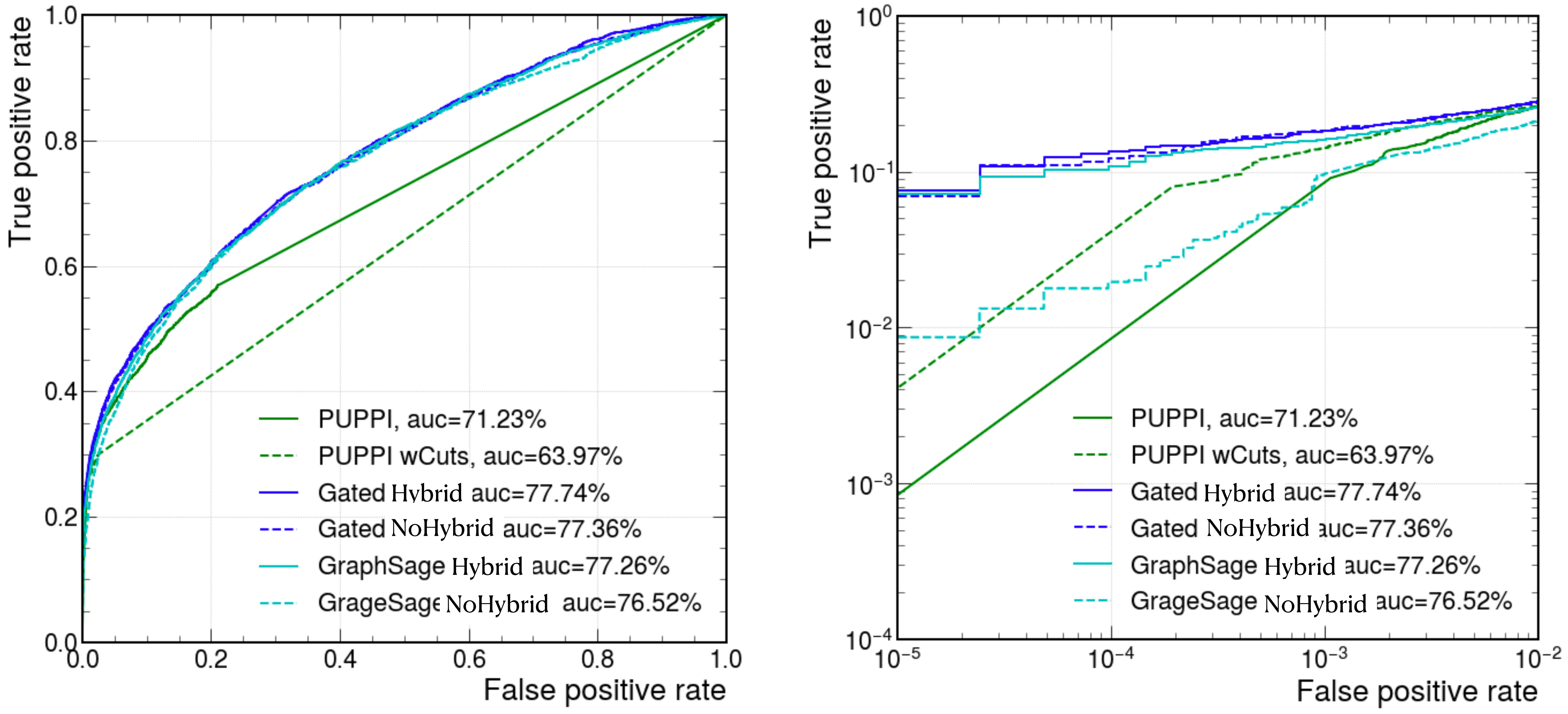
Back Up

Gated model Architecture



- x_u^k is the node, x_{v_i,v_j}^k are the node's neighbors.
- W1 is the "gate", which controls the message passing of M1
- W2 is another "gate", which controls the information updated from original feature $L(x_u^k)$ and neighbor representation $L(M2_u^k)$

RoC Curve



- PUPPI wCuts has lower AUC score, but the performance is much better than PUPPI, because of large improvements in the bottom left corner
  - Cuts depend on the global feature, i.e., number of pileup vertices
- GraphSage with and without Hybrid have similar AUC score, but the performance of Hybrid is much better than the one without Boost.
- Gated Hybrid have a bit improvements compared with NoHybrid

More Event Display

- Only includes particles with pT>1GeV in the plots
- For the improvements compared with PUPPI:
  - GNN can make use of particle self-features (more checks on isolated photon/leptons in the future), and GNN can find a better metric, instead of the alpha
  - GNN can make better use of neighboring features, instead of a ‘simple’ alpha equation in PUPPI, so it can remove more PU particles that are close to the soft LV particles, and more PU particles close to the LV jets.
  - GNN does not need many tunings (from our generalization test from PU80 to PU140)

More Event Display