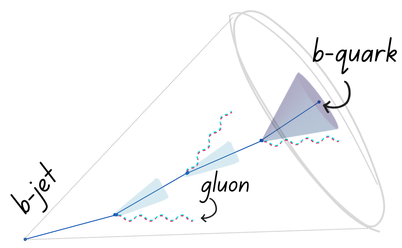
B-hadron identification in b-jets using novel deep learning technique in pp and PbPb collisions in CMS

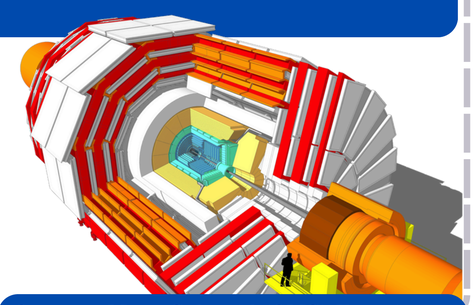
Prince Kumar (IIT Madras) on behalf of the CMS Collaboration
(kumar.prince@cern.ch)

Objectives

Understanding heavy-flavour jet substructure is key to QCD studies, especially for probing the dead-cone effect and jet quenching. However, b hadron decay kinematics complicate substructure measurements in inclusive b jets. We propose a geometric deep learning approach that learns from jets represented as point clouds of its charged decay products to identify b hadron tracks, with b jet tagging as an auxiliary task. The method is demonstrated using simulated pp and PbPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV with the CMS detector and compare it with existing boosted decision tree based methods.



Dead cone effect



Compact Muon Solenoid (CMS) Experiment

Dataset

We use jets clustered with the anti-k_t algorithm with R=0.4, Monte Carlo samples generated using PYTHIA8.

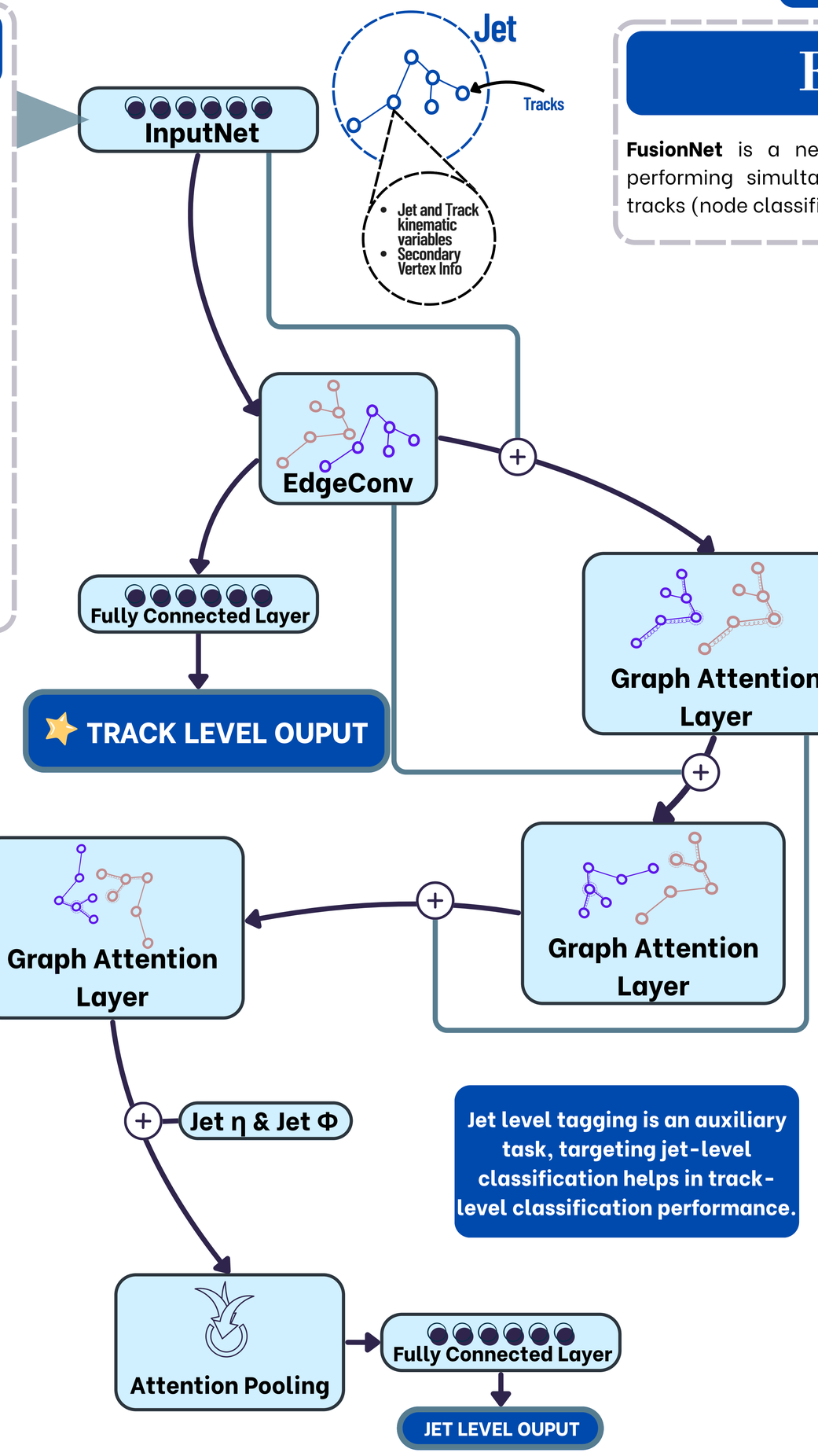
- **pp 5.02 TeV**: ~13M jets
  (Testing data: ~40K jets);
- **PbPb 5.02 TeV**: ~5M jets
  (Testing data: ~140K jets);

Tracks matched to generator-level charged particles from b hadron decays are considered signal. FusionNet uses 16 features as input, ranging from jet-level (assigned identically to all tracks) to secondary-vertex-level information. We train the model using categorical cross-entropy losses.

Jets are modelled as graphs with individual tracks as nodes, enabling the network to capture complex dependencies inherent in particle interactions.

FusionNet Model

FusionNet is a new Graph Neural Network architecture customised for performing simultaneous classification of jets (graph classification) and tracks (node classification) originating from b hadron decays in b-tagged jets.



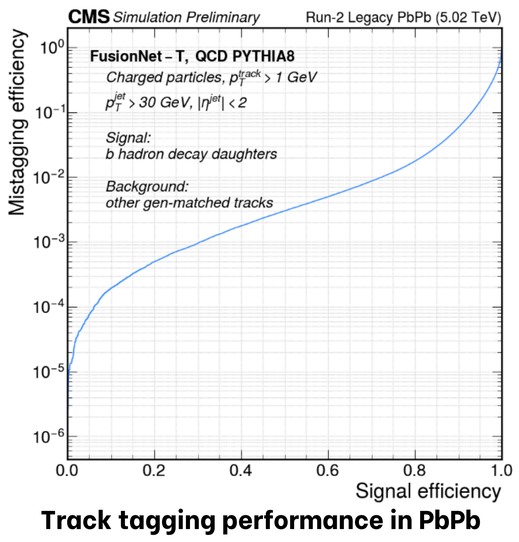
FusionNet-S

Trained on pp (FusionNet-S_pp) or PbPb (FusionNet-S_PbPb) Run2 simulations.

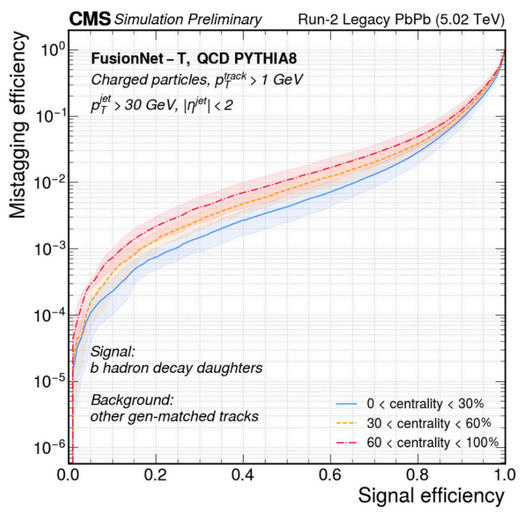
FusionNet-T

Trained on pp samples followed by fine-tuning on PbPb Run2 samples.

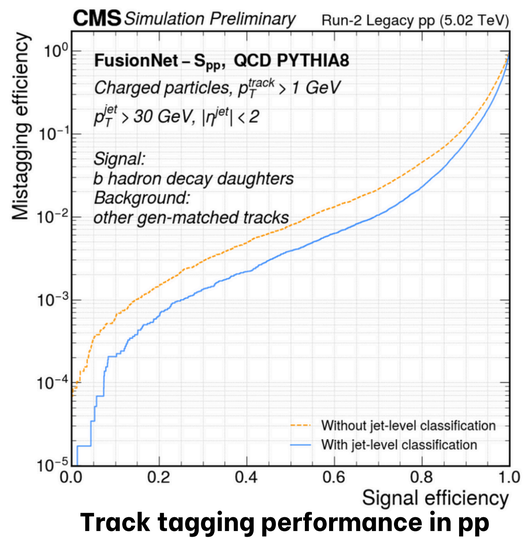
Results in PbPb



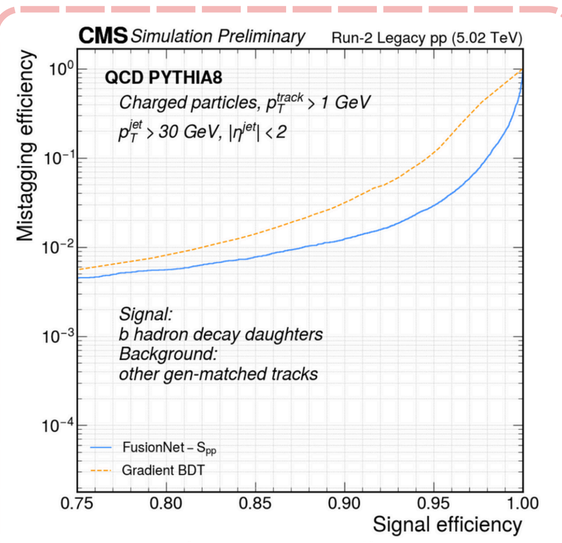
Track tagging performance in PbPb



Results in pp

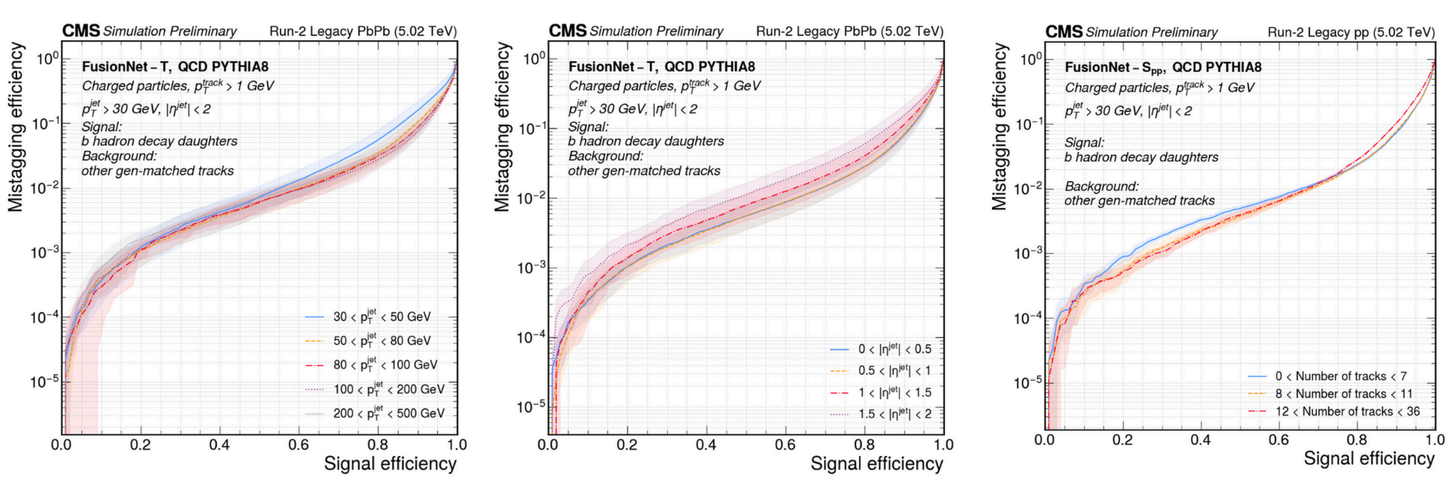


Track tagging performance in pp



FusionNet vs Gradient BDT

Track tagging performance in ParticleNet tagged b jets in pp



Performance of FusionNet in bins of jet transverse momentum, η, track multiplicity and centrality

Summary

- **FusionNet** is a novel **track-finding algorithm** that leverages advanced geometric deep learning techniques to identify b hadrons within b jets, while also performing jet tagging.
- This novel model performs well for **track tagging** in both proton-proton **(pp)** and lead-lead **(PbPb)**, and **outperforms GradientBDT** for track tagging in ParticleNet tagged b jets in pp.
- The performance of the model **remains largely consistent across jet transverse momentum, η, and track multiplicity**, showing robustness.

References

- CMS Collaboration. (2025). A novel track finding algorithm to identify b hadron in b jets using FusionNet: a geometric deep learning model. CMS-DP-2025-035
- Qu, H., & Gouskos, L. (2020). Jet tagging via particle clouds. Phys. Rev. D, 101, 056019.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks. In International Conference on Learning Representations (ICLR).
- CMS Collaboration. (2024). Run 3 commissioning results of heavy-flavor jet tagging at $\sqrt{s}$ = 13.6 TeV with CMS data using a modern framework for data processing. CMS-DP-2024-024.
- CMS Collaboration. (2024). Jet fragmentation function and groomed substructure of bottom quark jets in proton-proton collisions at $\sqrt{s}$ = 5.02 TeV. CMS-PAS-HIN-24-005.