JET CLASSIFICATION WITH PARTICLE TRANSFORMERS: A MULTICLASS LEARNING APPROACH

Andrés Duque Bran, Julien Donini, Samuel Calvet
LPCA, Clermont-Ferrand, France

December 1st, 2025

THE LHC AND THE ATLAS EXPERIMENT

- LHC is the largest particle accelerator and collides protons at 13.6 TeV.
- ATLAS is one of the 4 experiments at the LHC. It records particles from collisions to study fundamental physics and search for new phenomena.

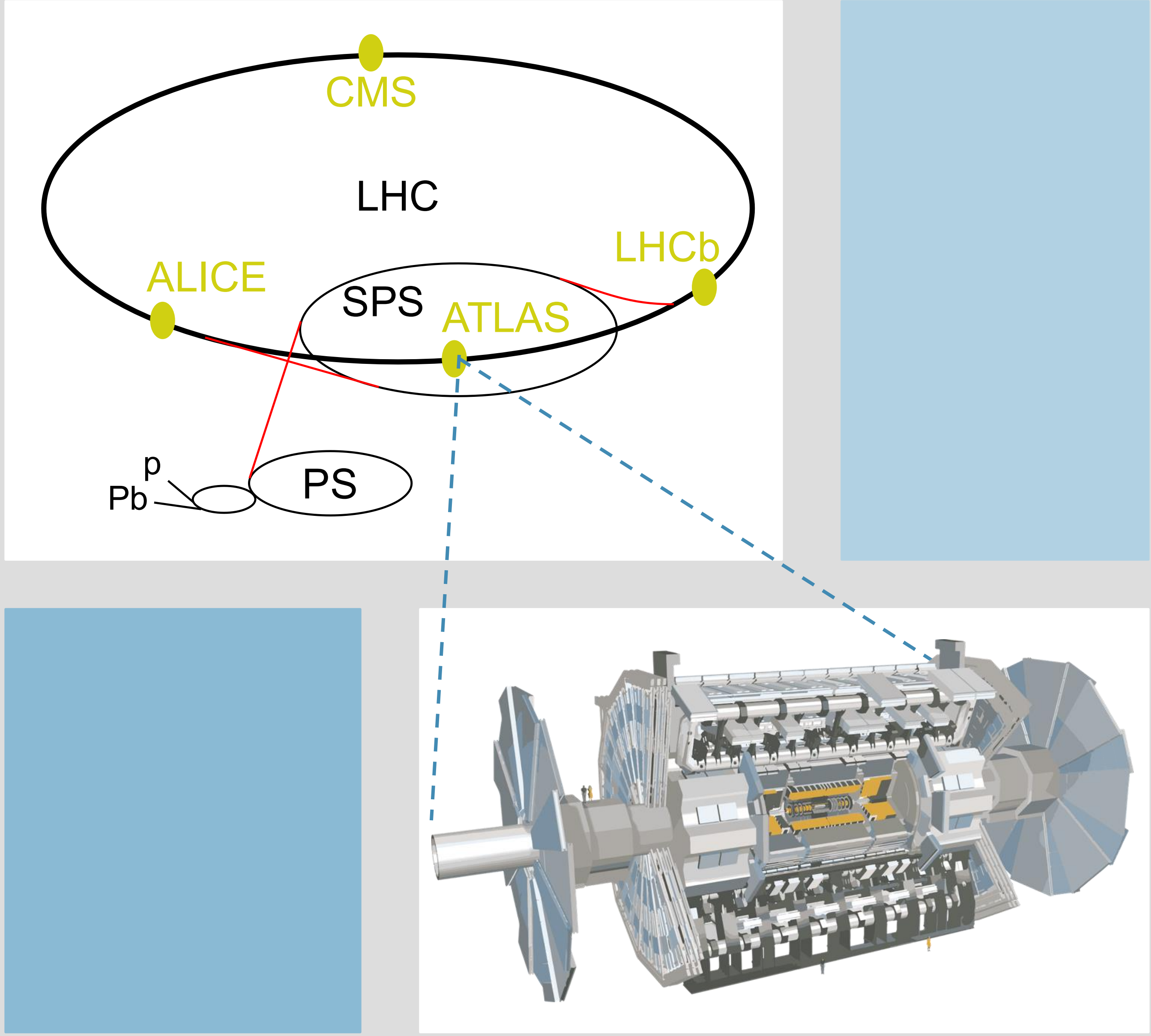


Fig. 1: Schematic view of the LHC complex and cut-away view of the ATLAS detector.

WHAT IS A JET?

- Quarks and gluons hadronize into sprays of particles after pp collisions.
- Jets are reconstructed from energy deposits and tracks in the detector.
- Bosons (W, Z, Higgs) can decay into quarks and at high energy their product form a jet.
- **Jet Tagging:** Identifying the origin of the jet.

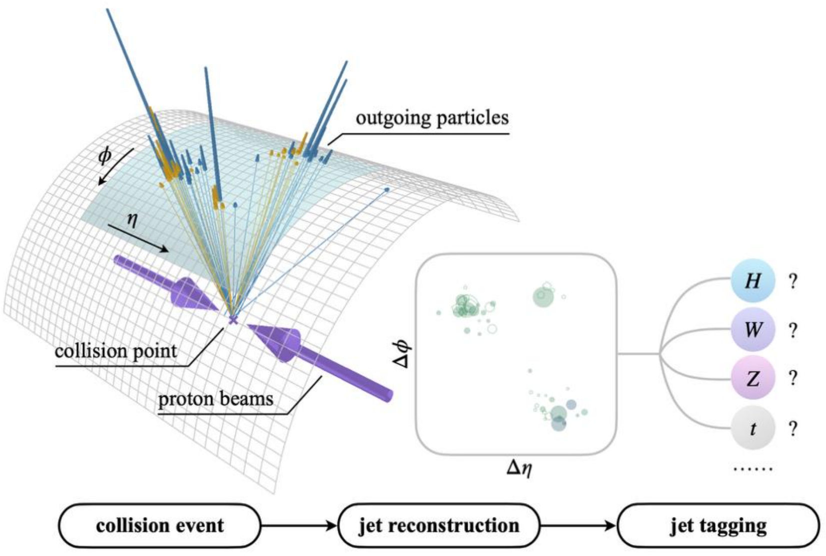


Fig. 2: Illustration of jet tagging at the LHC .

THE PART ARCHITECTURE

- Built on **Transformer attention blocks**.
- Takes **per-particle** and **interaction features** as inputs.
- Uses **Particle Multi-Head Attention**, where interaction features are embedded directly into the attention mechanism.
- Unlike NLP Transformers, jets are unordered sets.

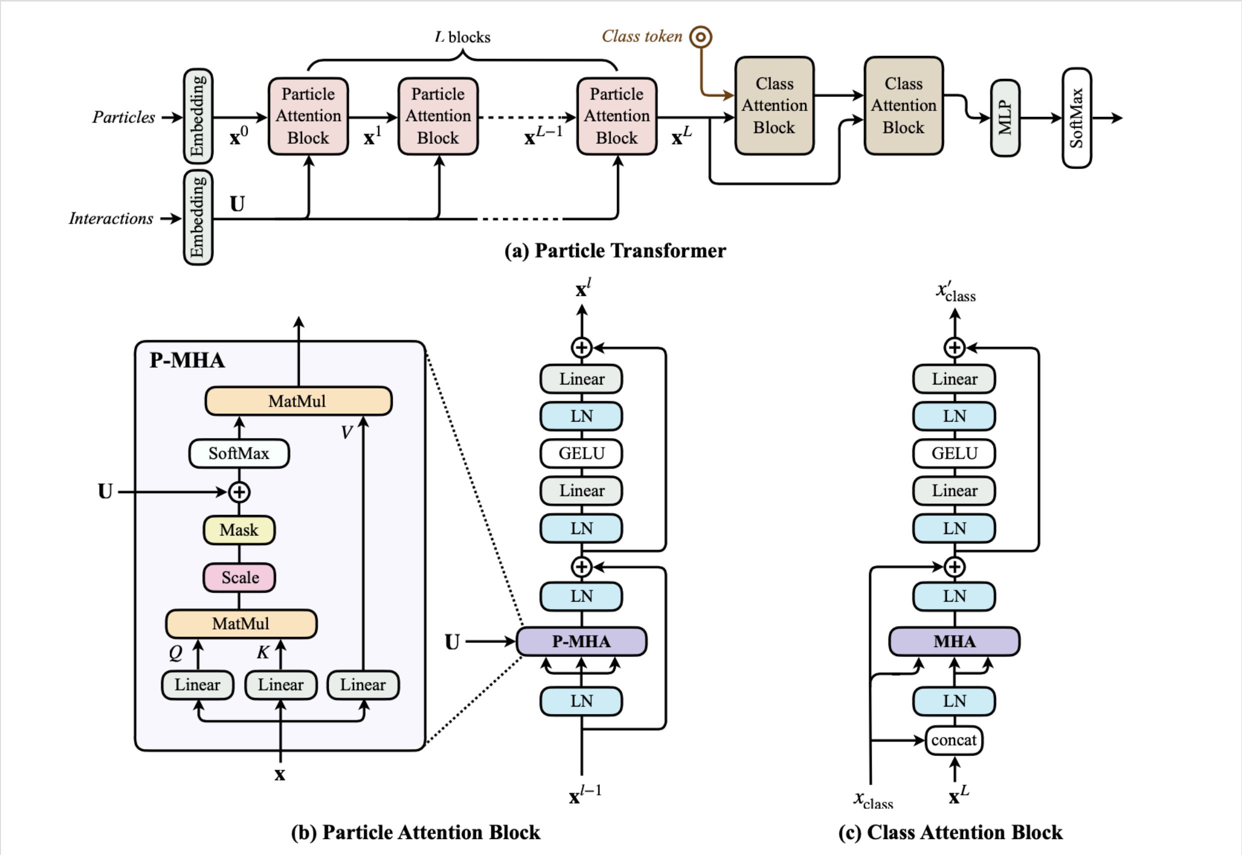


Fig. 3: The architecture of the **Par**ticle **T**ransformer **(ParT)** [arXiv:2202.03772].

DATASET

- Simulated jets from proton-proton collisions at 13 TeV:

  Events generated with **Pythia8** and ATLAS detector simulation.

- Large-R jets:
  - QCD
  - W → qq
  - Z → qq, bb, cc
  - top → qqb
  - Higgs → bb

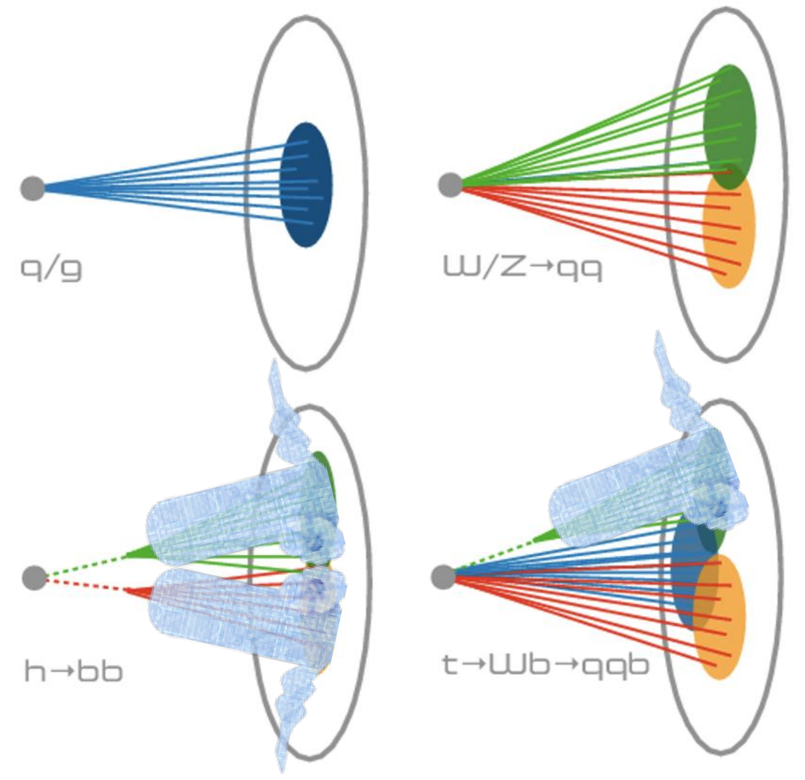


Fig. 4: Schematic representation of the types of jets present in the dataset.

MODEL SETUP

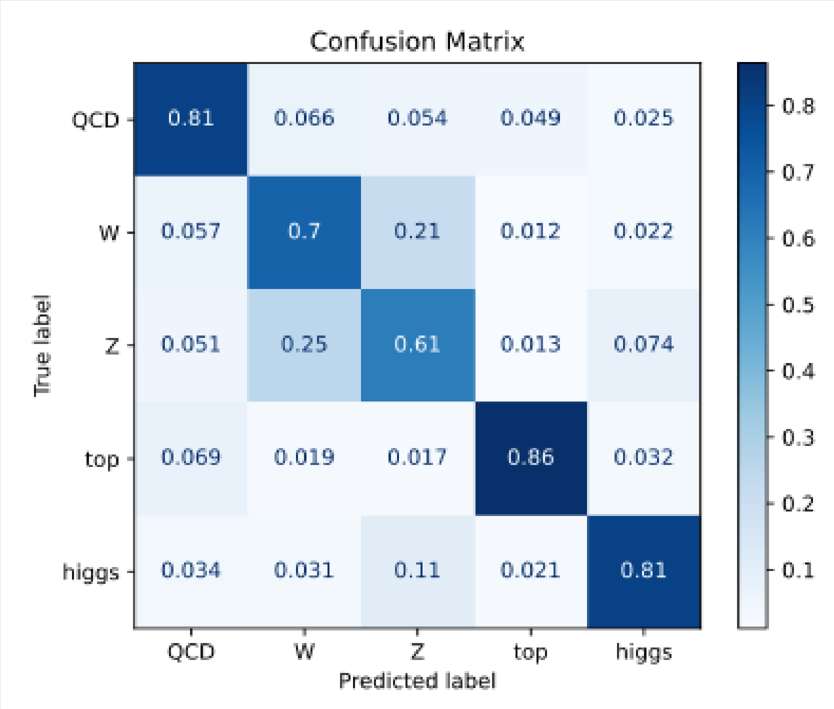
- Dataset:
  - 50M jets (90%/10% for train and validation)
  - 16M jets for prediction
- Features:
  - Per-particle: Kinematical and trajectory displacements
  - Pair-wise interaction: Angular differences.
- Max 80 constituents per jet.
- Events reweighted to flatten the p_T distribution per class

Class	Training+Validation	Prediction
QCD	15.5M	5.1M
W	8.6M	2.8M
Z	8.6M	2.8M
Top	15.5M	5.1M
Higgs	1.7M	0.6M

Table 1: Distribution of number of samples per class for training and inference.

PART PERFORMANCE

- Top and Higgs jets are well distinguished by the model.
- Noticeable confusion between Z and W jets, and a smaller amount between Z and Higgs.



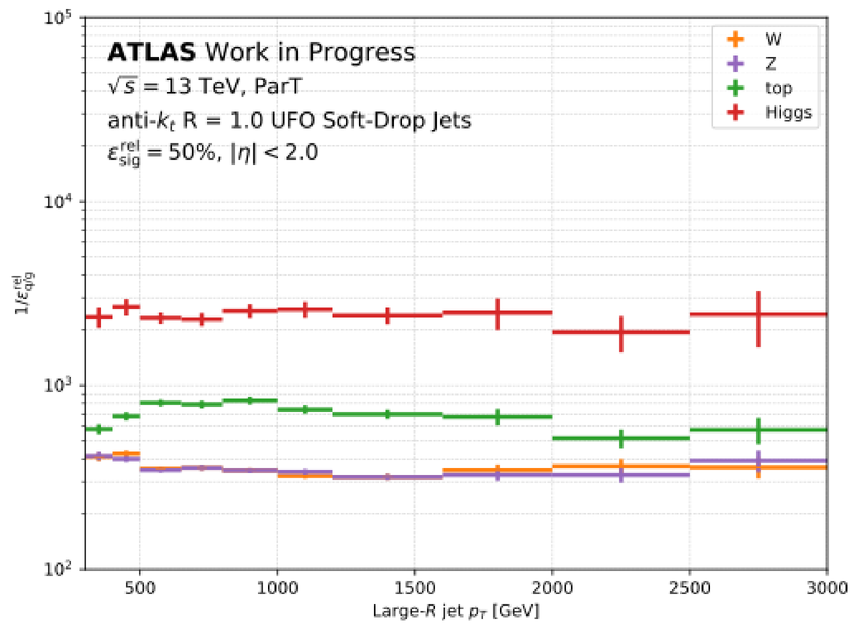


Fig. 5: Confusion matrix and QCD rejection vs p_T.

EFFECT OF IMPACT PARAMETERS

- Baseline features:
  - $p_T,\ \eta,\ \phi,\ E,\ \Delta R.$
- Additional features:
  - Track –Level: $d_0,\ z_0.$
- Addition of impact parameters **significantly improves tagging performance** for Higgs and top.

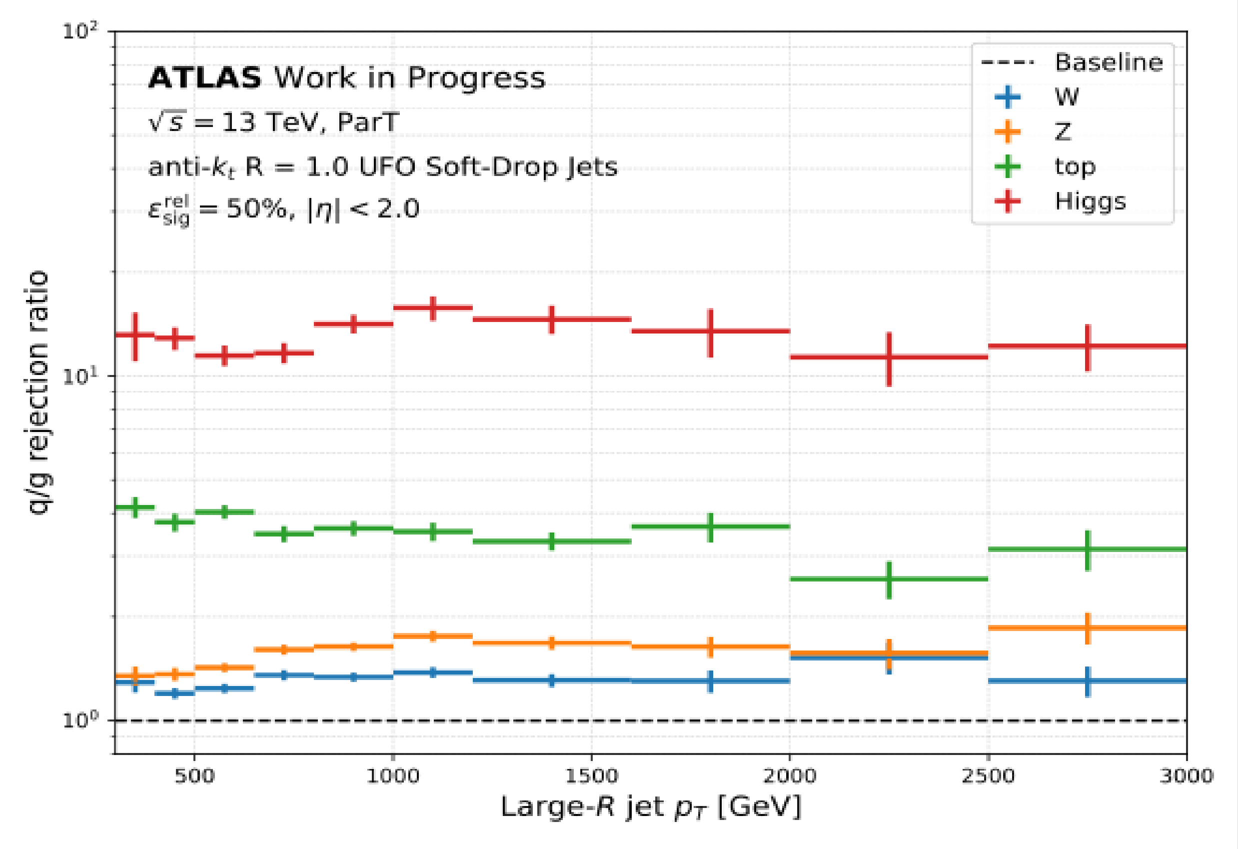


Fig. 6: QCD rejection ratio between baseline + additional and baseline features.

STATISTICS IMPACT

- **ParT** trained with 1M, 10M, and 50M jets.
- **Class proportions** are kept for all samples.
- More statistics lead to **lower losses** and **smoother convergence**.

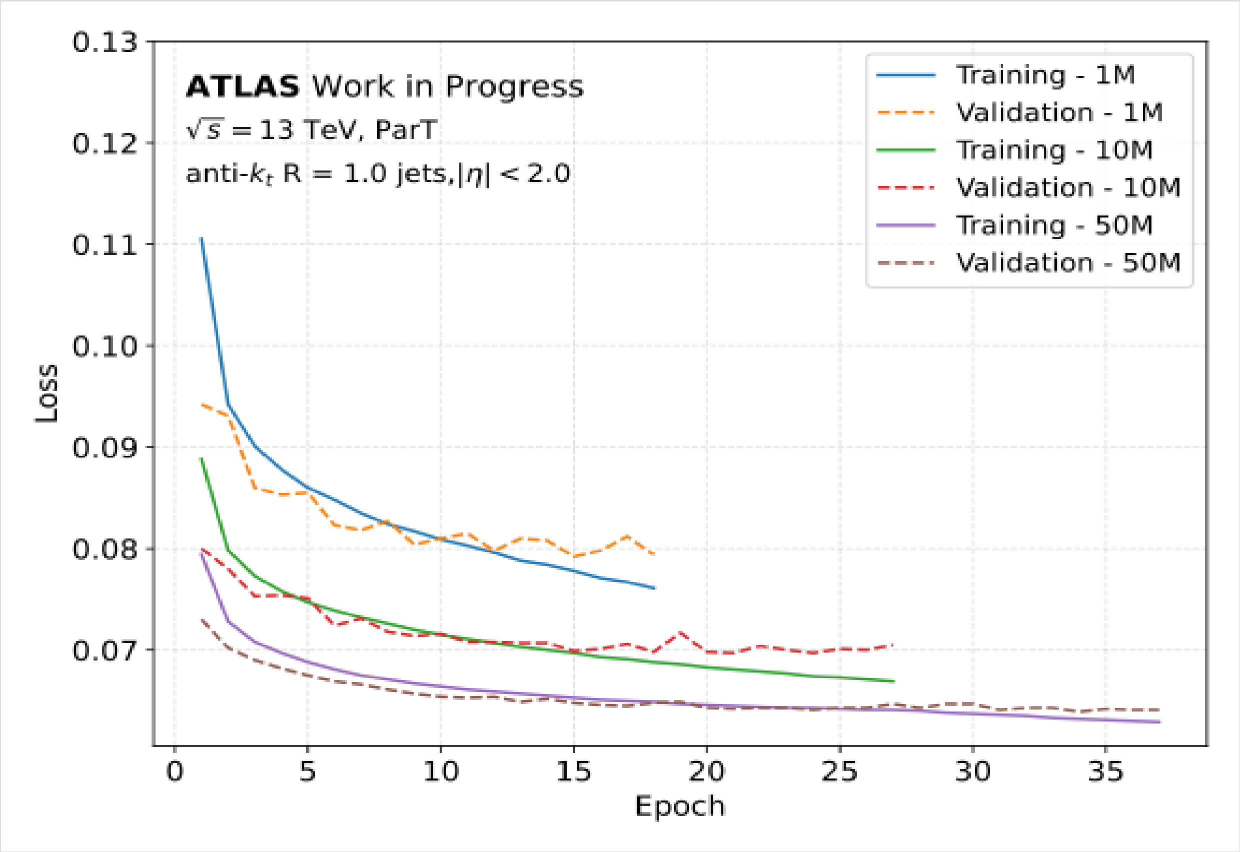
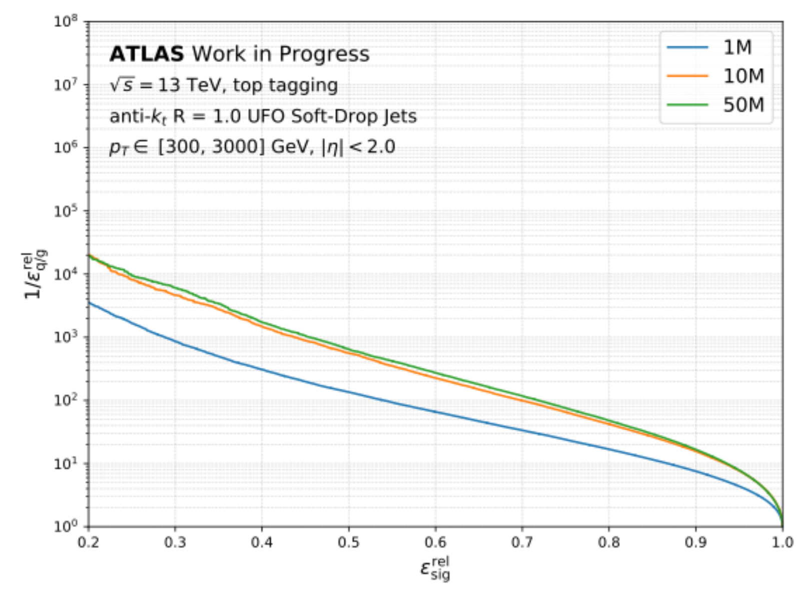


Fig. 7: Training and validation loss for ParT with different statistics.

STATISTICS IMPACT

More statistics lead to **better background rejection** for **top** and **Higgs.**



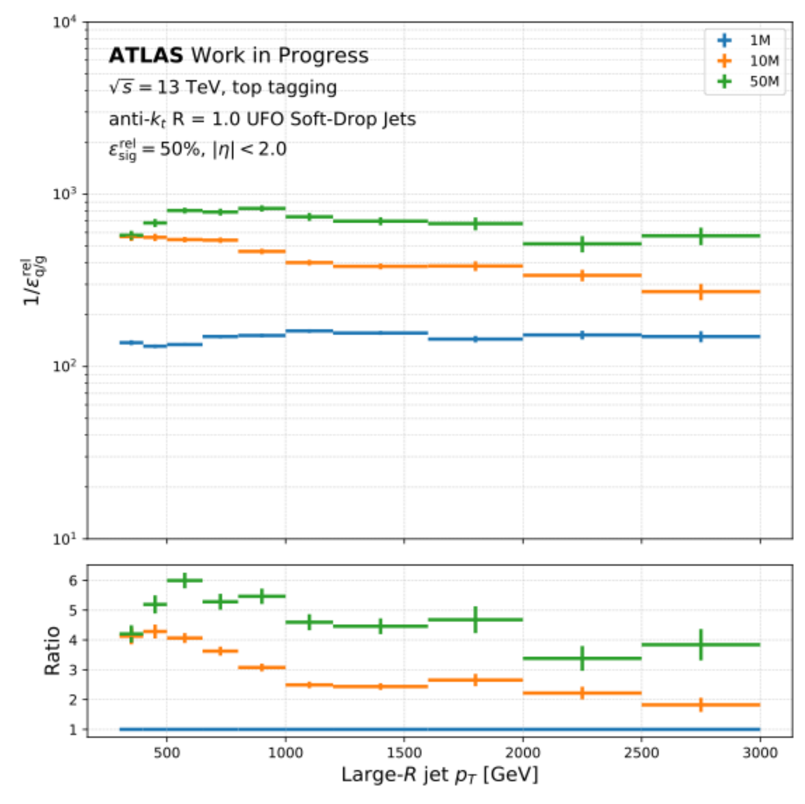
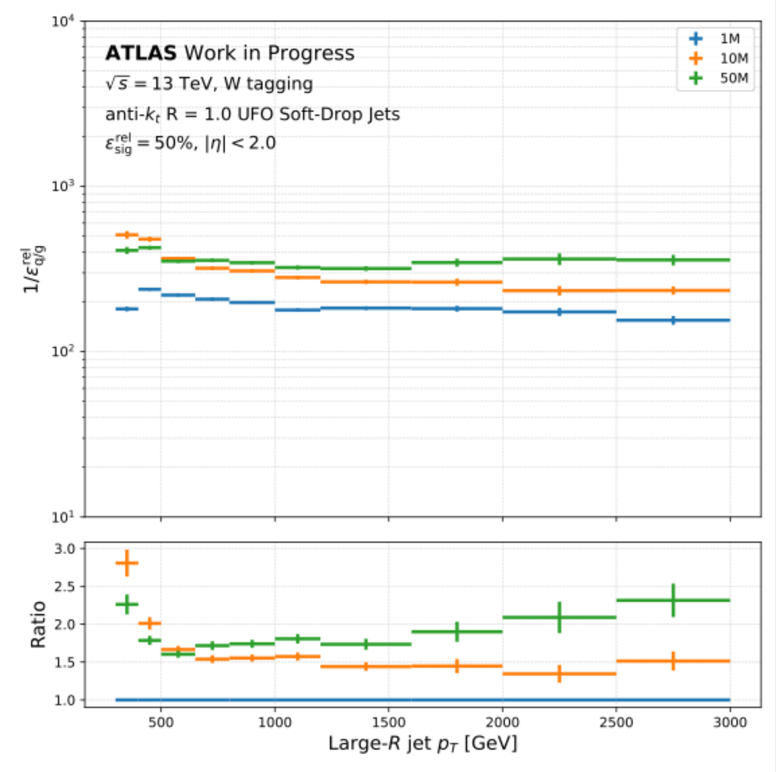


Fig. 8: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **top jets**.

STATISTICS IMPACT

- Better rejection at **low signal efficiency** for 10M training.
- Performance stabilizes between 10M and 50M for **W and Z jets** at **low p_T**.



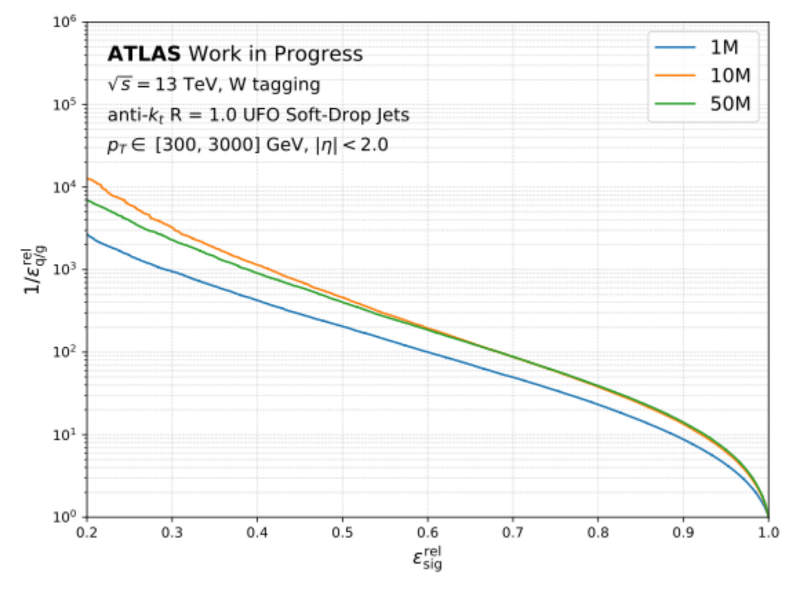


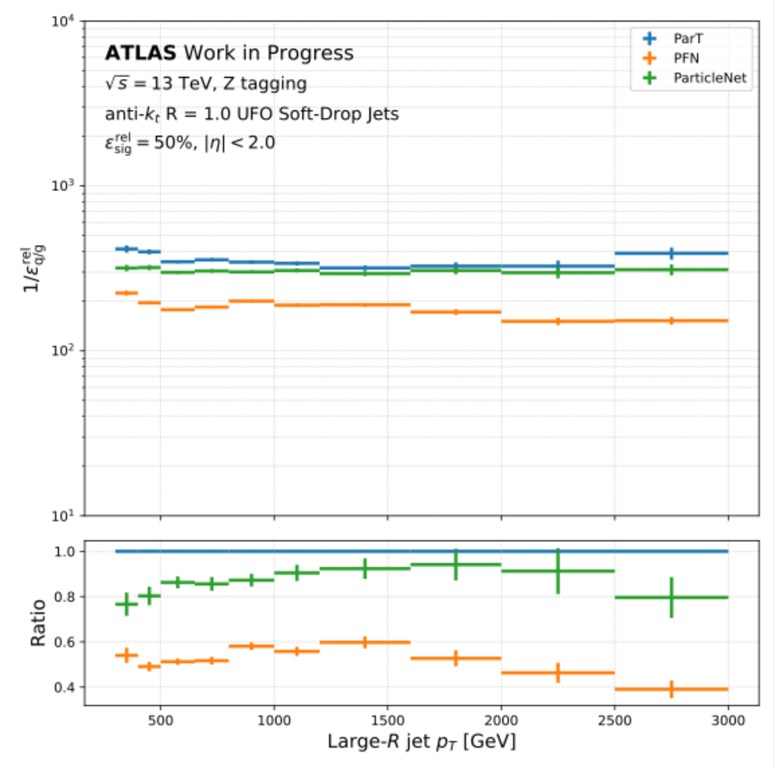
Fig. 9: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **W jets**.

COMPARISON TO OTHER ARCHITECTURES

- Additional architectures evaluated:
  - **PFN:** Per-particle MLP followed by per-jet aggregation. Uses **only per-particle features**.
  - **ParticleNet:** GNN with dynamic k-NN, combining **pairwise interaction features** with per-particle inputs.
- **Training stats are identical for all architectures.**

COMPARISON TO OTHER ARCHITECTURES

- Same dataset and statistics for **training**.
- PFN: per-particle features
- ParT, ParticleNet: per-particle features + pairwise interaction information.
- Clear performance gain with pairwise interaction variables.



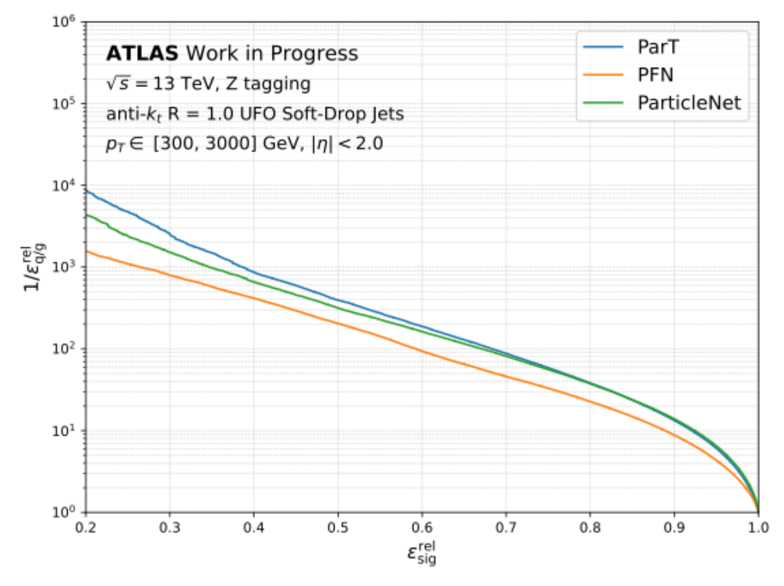


Fig. 10: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Z jets**.

COMPARISON VS OTHER MONTE CARLO GENERATORS

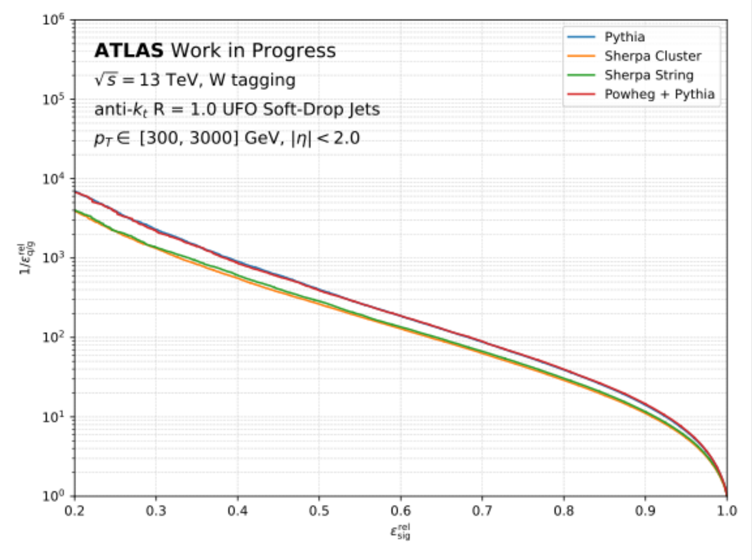
- The QCD background samples were replaced by different MC generators.
- Training statistics were kept identical across all generators for a fair comparison.

Generator	Hadronization
Training	
Pythia8	String
Testing	
Sherpa 2.2.5	Cluster
Sherpa 2.2.5	String
Powheg + Pythia8	String

Table 2: Monte Carlo generators used for training and testing.

COMPARISON TO OTHER MC GENERATORS

- Better rejection for samples including Pythia8.
- Comparable performance within 20% at higher p_T for all classes.



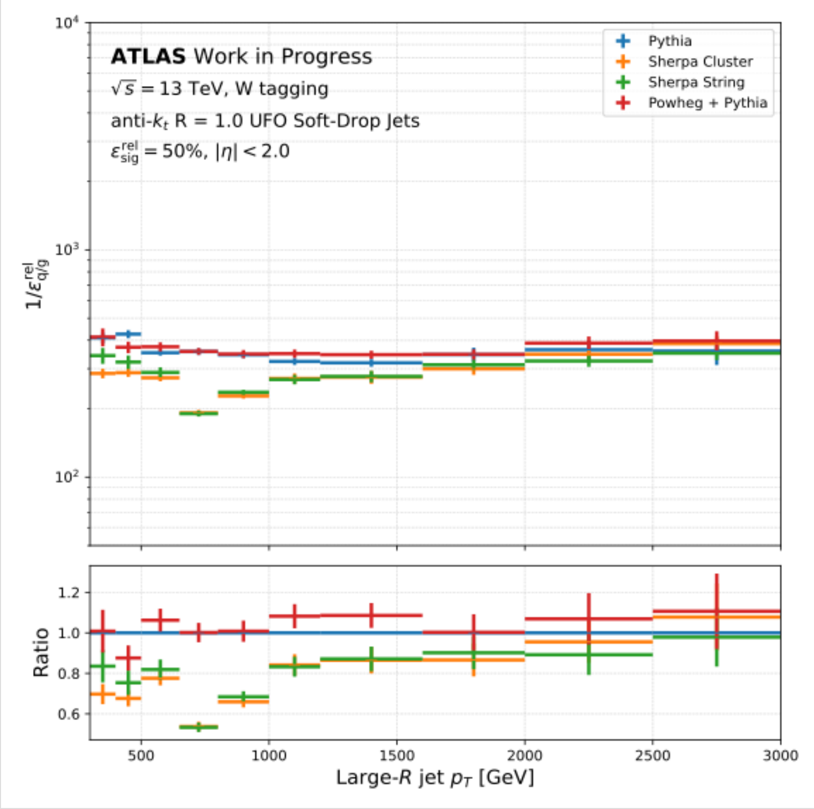
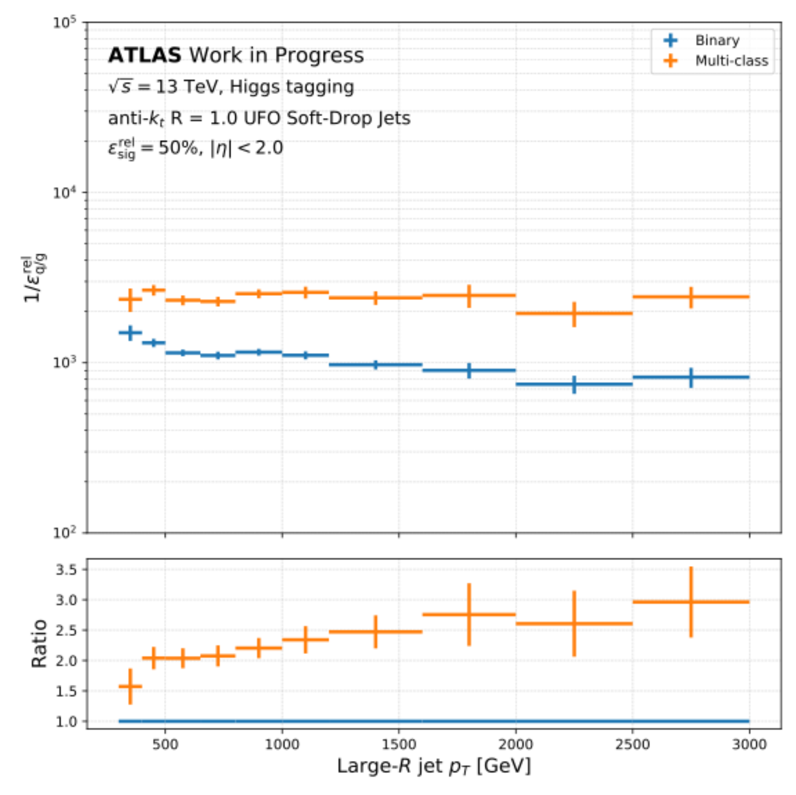


Fig. 11: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **W jets**.

COMPARISON TO BINARY PART

- A separate **binary ParT model** was trained for each signal:
  - QCD vs W, QCD vs Z, QCD vs top, QCD vs Higgs.
- Training and test statistics comparable.
- **Multiclass** slightly **outperforms** the **binary** model.



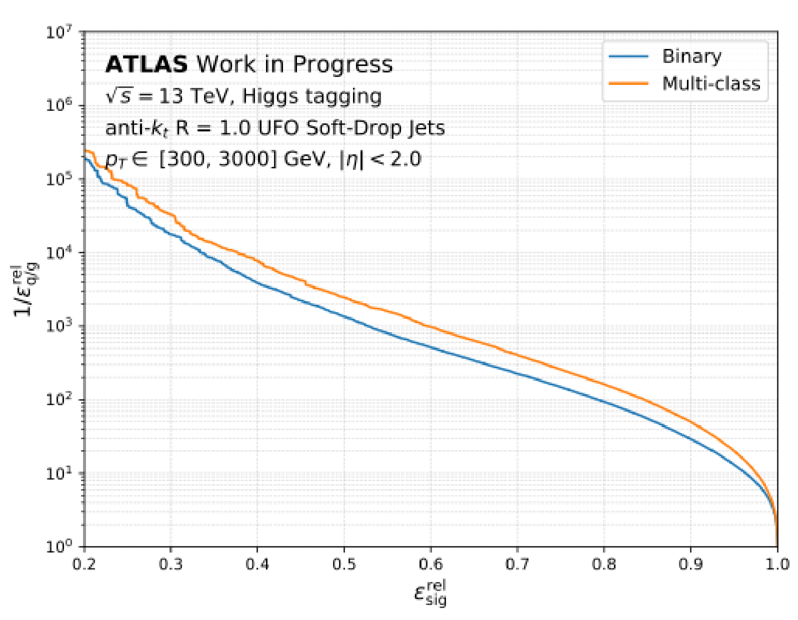


Fig. 12: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Higgs jets**.

CONCLUSION

- **ParT** evaluated across inputs, stats, architectures, MC generators, and binary vs multiclass.
- **Impact parameters** → **clear gain**, strongest for Higgs.
- **More training stats** → **better performance**, especially for top and Higgs.
- **ParT > PFN**, and com**parable to ParticleNet** thanks to relational features.
- **Stable across MC generators with Pythia8**.
- **Multiclass > binary** → one unified model works best.

THANK YOU

BACKUP

THE STANDARD MODEL

- The Standard Model describes elementary particles and forces.
- Fundamental particles:
  - Quarks
  - Leptons
- Force carriers
  - Vector bosons
  - Scalar bosons

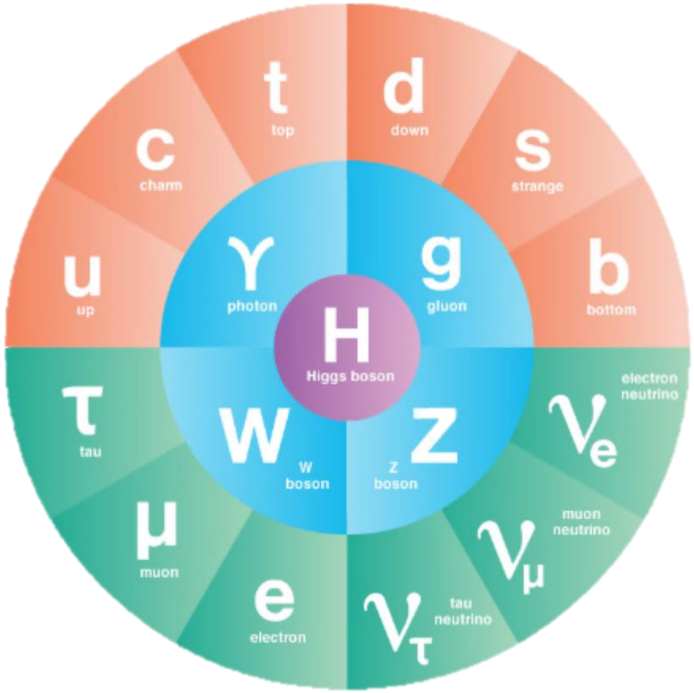


Fig. 13: Elementary particles of the Standard Model.

INPUT FEATURES

- Kinematical:
  - $p_T,\ \eta,\ \phi,\ E,\ \Delta R.$
- Impact parameters:
  - Track –Level: $d_0,\ z_0.$

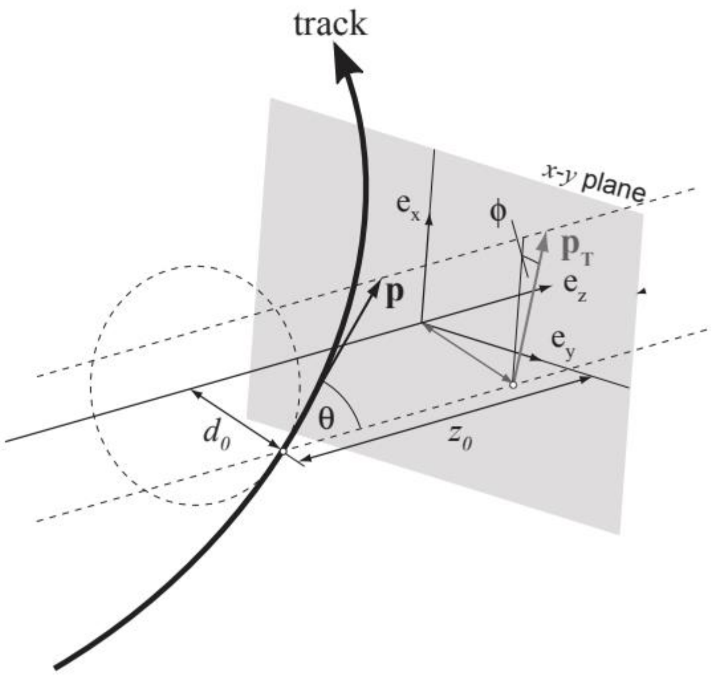


Fig. 14: Track coordinates Diagram. [ATL-PHYS-PUB-2017-014]

EFFECT OF IMPACT PARAMETERS

Baseline features:

$p_T, \ \eta, \ \phi, \ E, \ \Delta R.$

New features:

Track –Level: $d_0, \ z_0.$

- **Charge of constituents not used for training.**

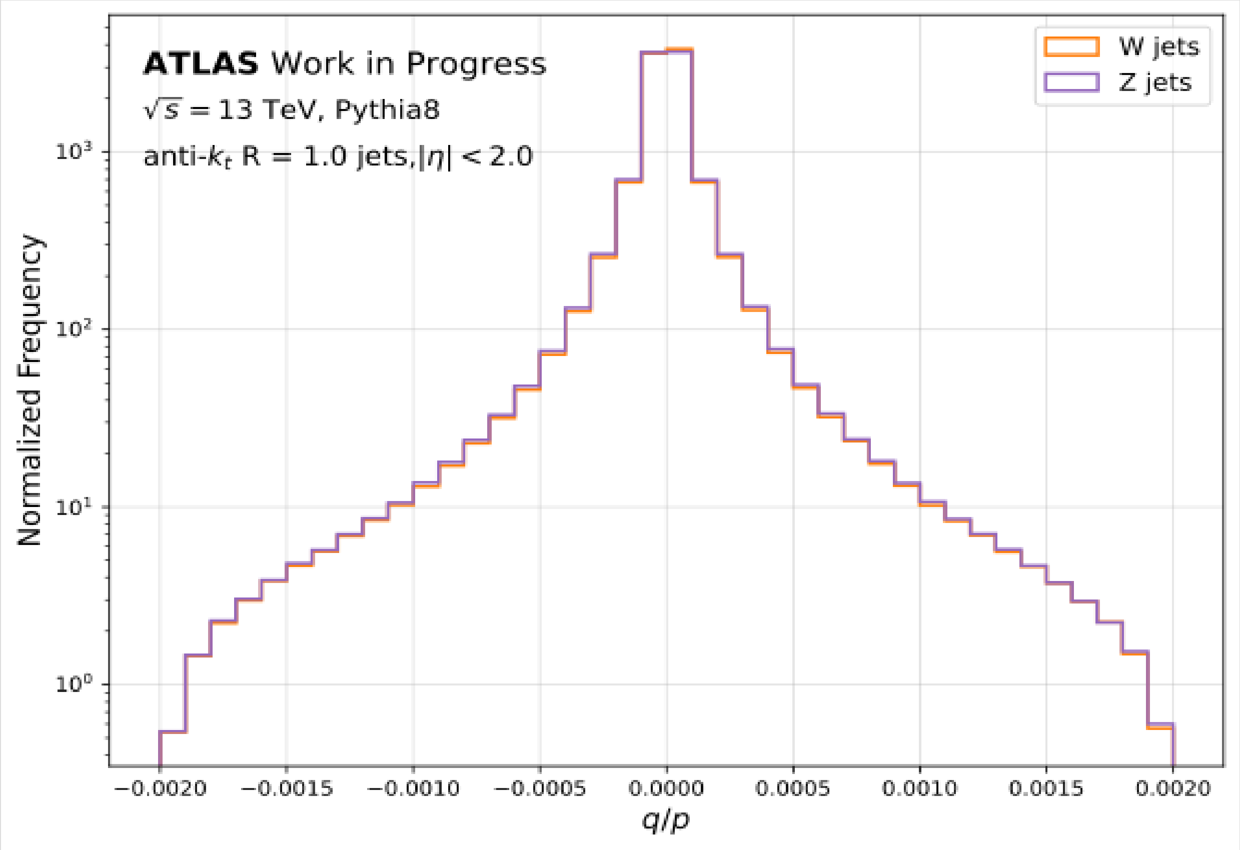


Fig. 15: Charge per constituent distribution from W and Z jets.

TRAINING DETAILS

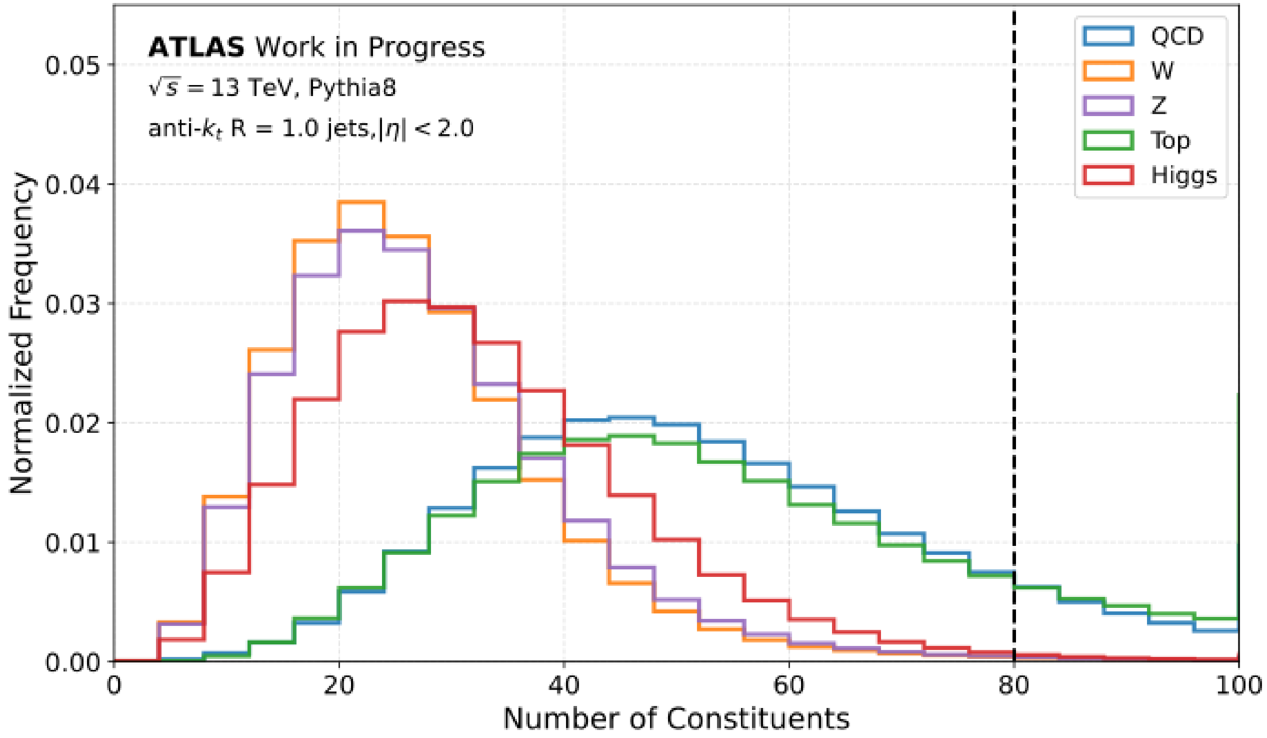


Fig. 16: Distribution of number of constituents per class.

p_T DISTRIBUTION REWEIGHTING

Jet p_T is reweighted before training to ensure that all classes share a similar p_T distribution, preventing the model from learning kinematic biases instead of physics features.

- **This reweighting stabilizes training and improves generalization,** ensuring that performance differences across classes are not driven by mismatched p_T spectra.

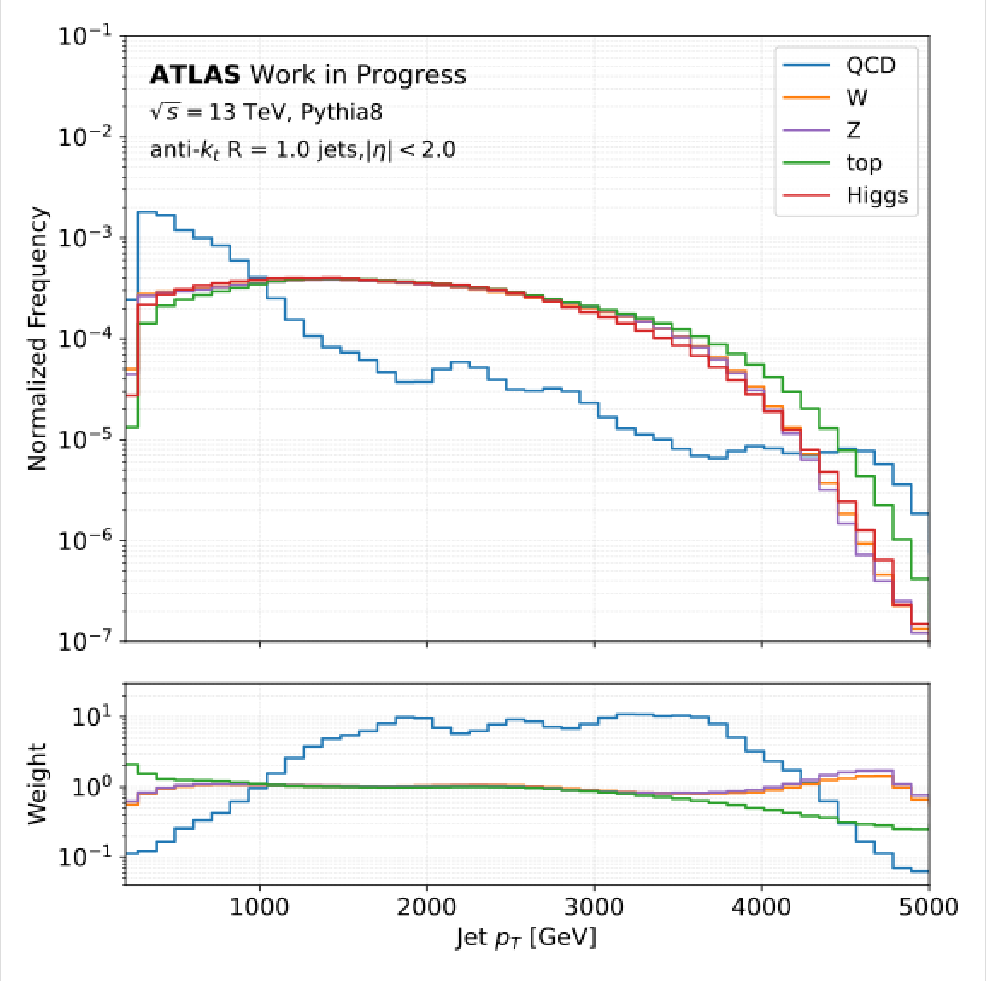


Fig. 17: p_T distribution per jet and weights applied during training.

THE PFN ARCHITECTURE

- **Permutation-invariant** architecture that processes jets as unordered sets of particles.
- Uses a **per-particle embedding network** followed by a sum to aggregate information.
- A final classifier network learns global jet features from the aggregated representation for tagging tasks.

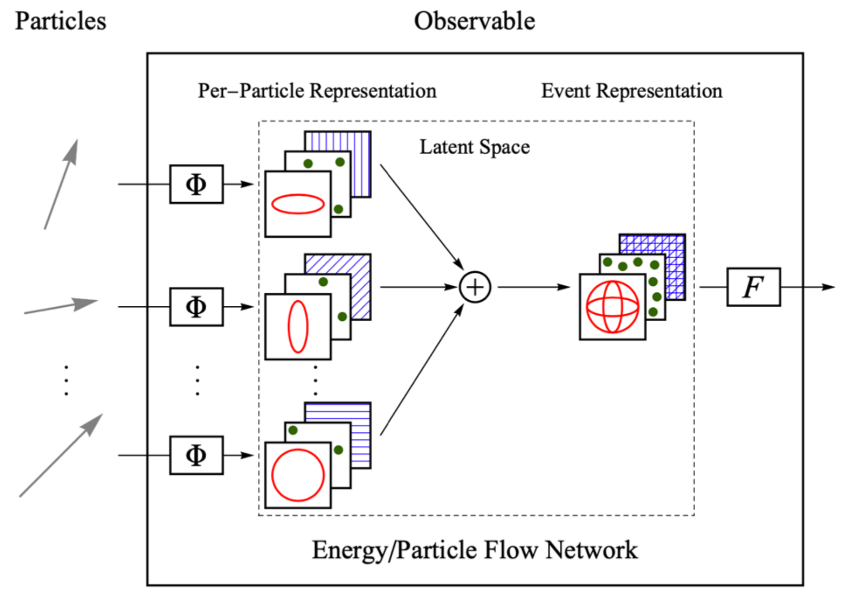
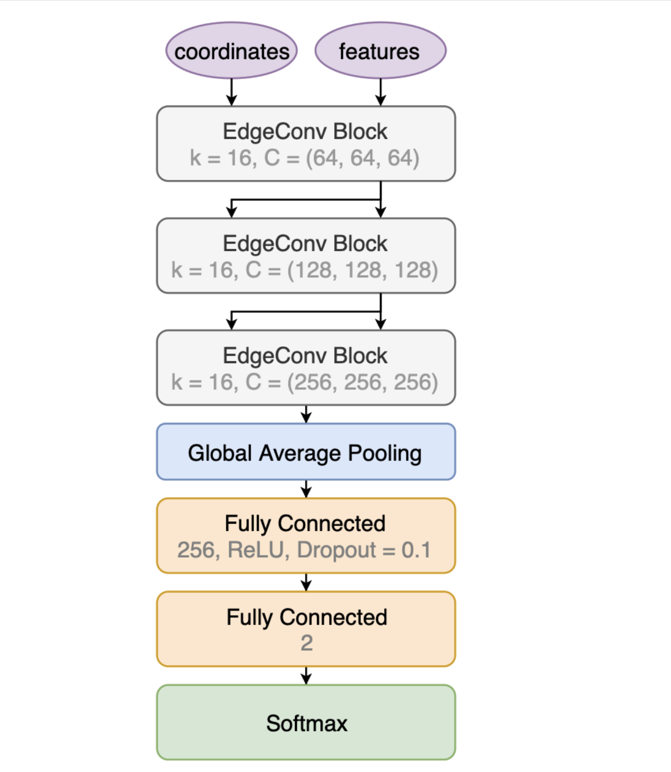


Fig. 18: The architecture of Particle **F**low **N**etwork **(PFN)** [arXiv:1810.05165].

THE PARTICLENET ARCHITECTURE

- **Dynamic Graph CNN:** builds and updates a particle graph using nearest-neighbor relations.
- **EdgeConv blocks:** learn geometric and relational features between constituents.
- **Hierarchical aggregation:** combines local and global information into a jet-level representation.



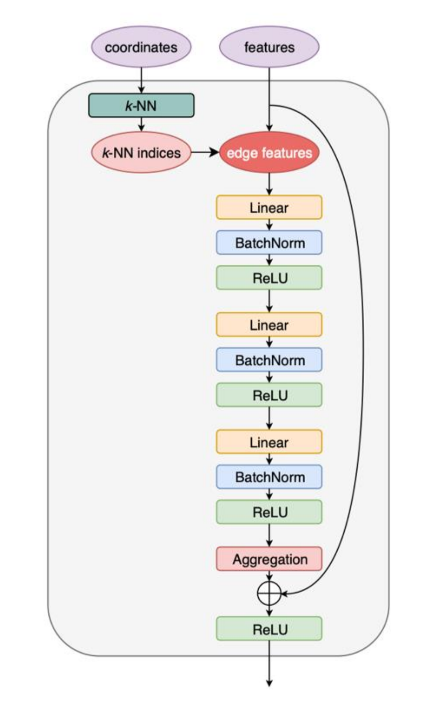
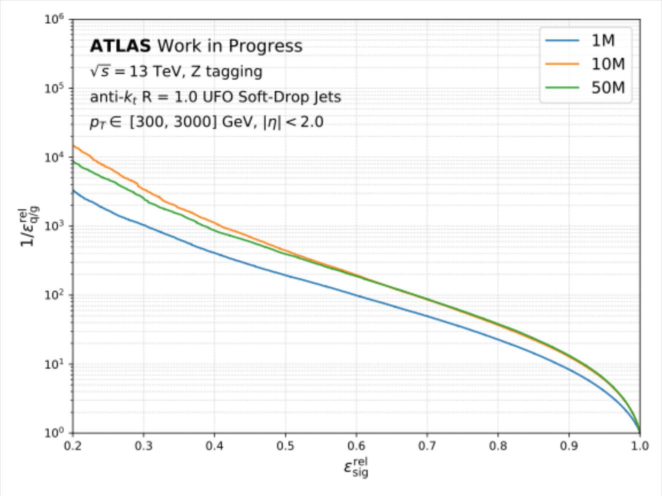


Fig. 19: The architecture of ParticleNet.

STATISTICS IMPACT

- Performance stabilizes between 10M and 50M for **W and Z** at **low p_T**.
- Better rejection at **low signal efficiency** for 10M training.



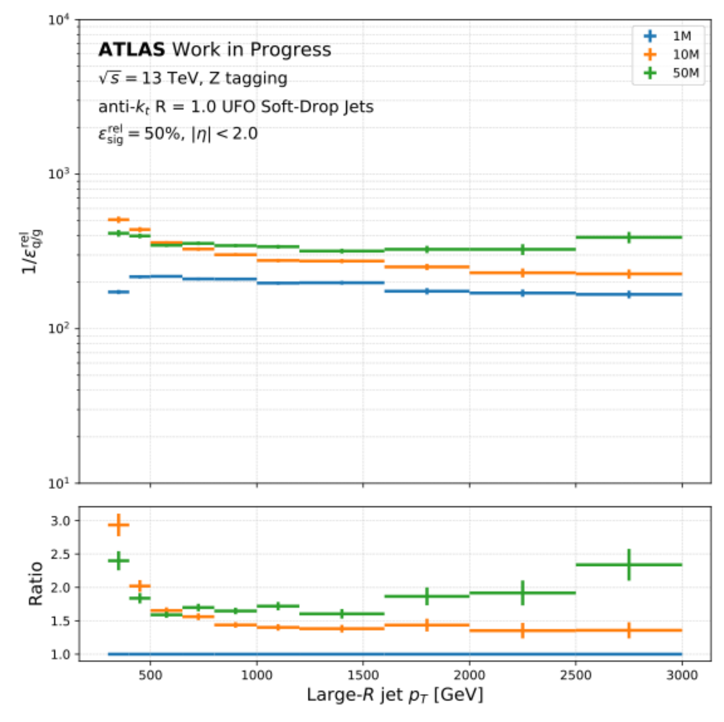
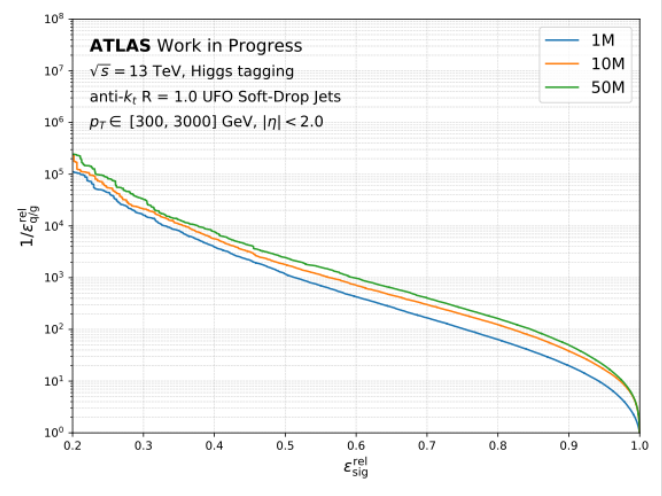


Fig. 20: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Z jets**.

STATISTICS IMPACT

More statistics lead to **better background rejection** for **top** and **Higgs.**



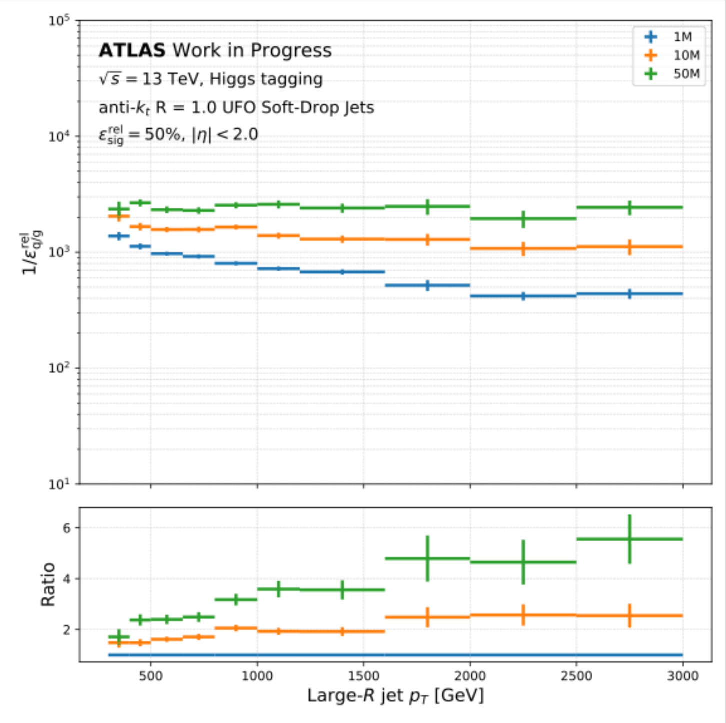
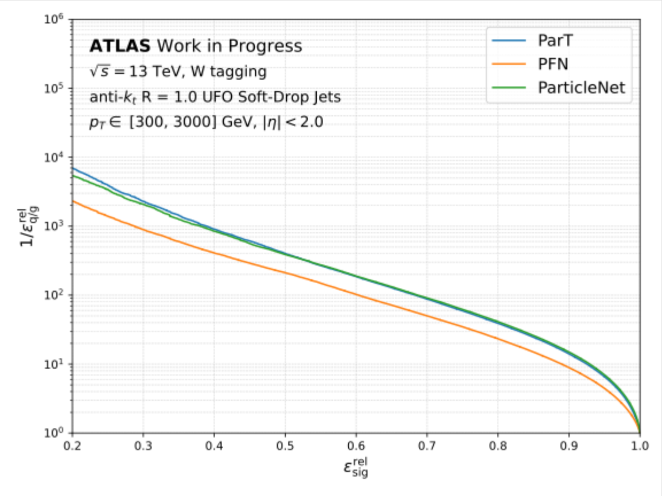


Fig. 21: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Higgs jets**.

COMPARISON TO OTHER ARCHITECTURES

- All models trained on the **same dataset and statistics.**
- PFN uses only **per-particle features**, while ParT and ParticleNet also use **pairwise interaction information**.
- Adding pairwise relations leads to a **big performance gain**: ParT and ParticleNet clearly outperform PFN.



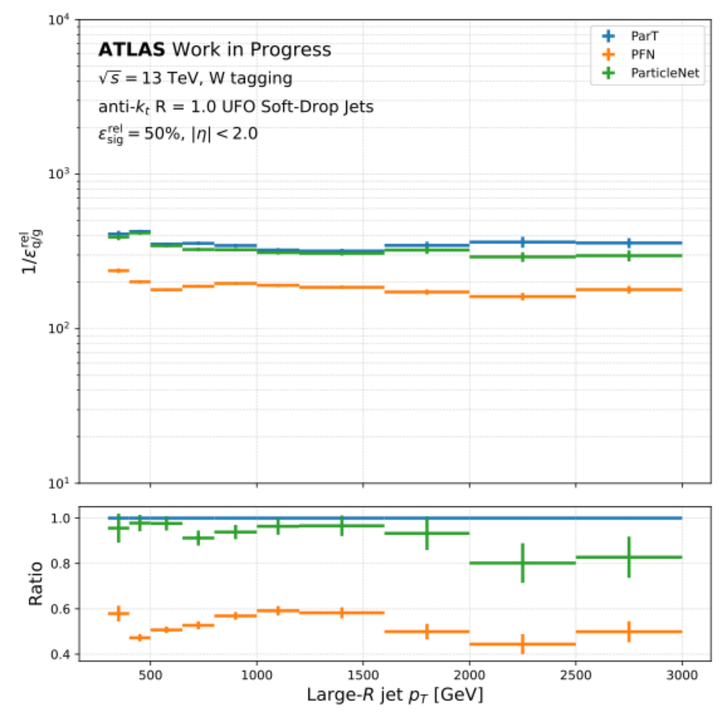
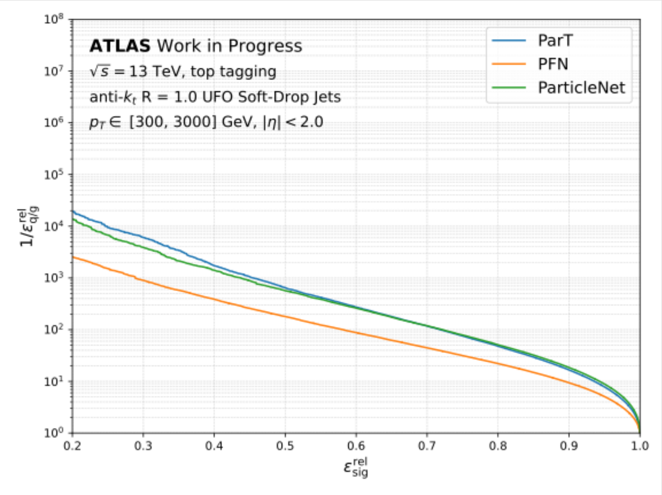


Fig. 22: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **W jets**.

COMPARISON TO OTHER ARCHITECTURES

- All models trained on the **same dataset and statistics.**
- PFN uses only **per-particle features**, while ParT and ParticleNet also use **pairwise interaction information**.
- Adding pairwise relations leads to a **big performance gain**: ParT and ParticleNet clearly outperform PFN.



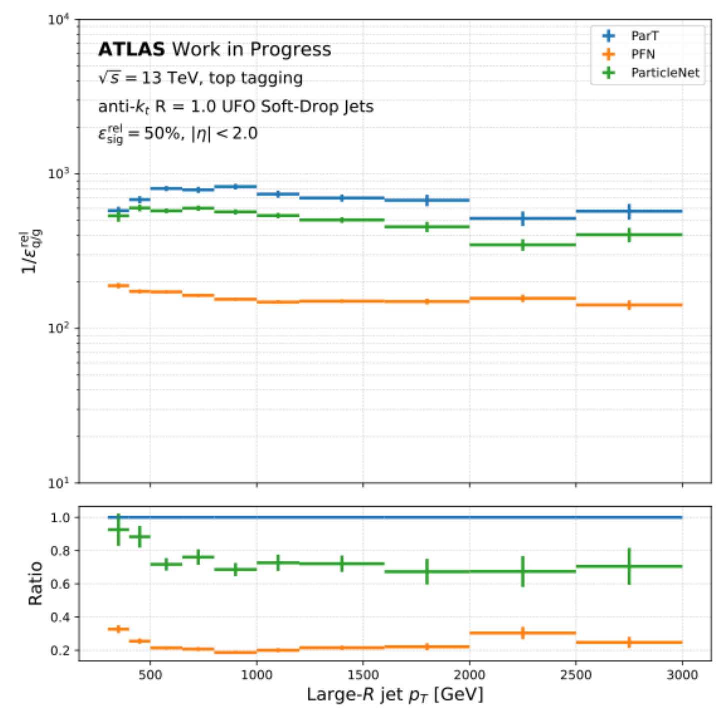
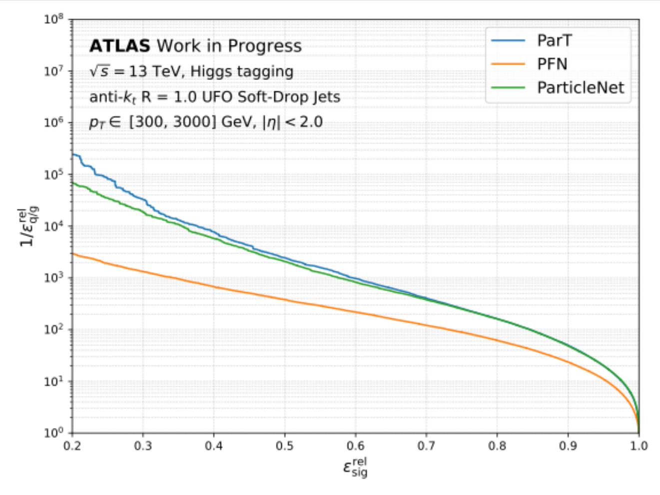


Fig. 23: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **top jets**.

COMPARISON TO OTHER ARCHITECTURES

- All models trained on the **same dataset and statistics.**
- PFN uses only **per-particle features**, while ParT and ParticleNet also use **pairwise interaction information**.
- Adding pairwise relations leads to a **big performance gain**: ParT and ParticleNet clearly outperform PFN.



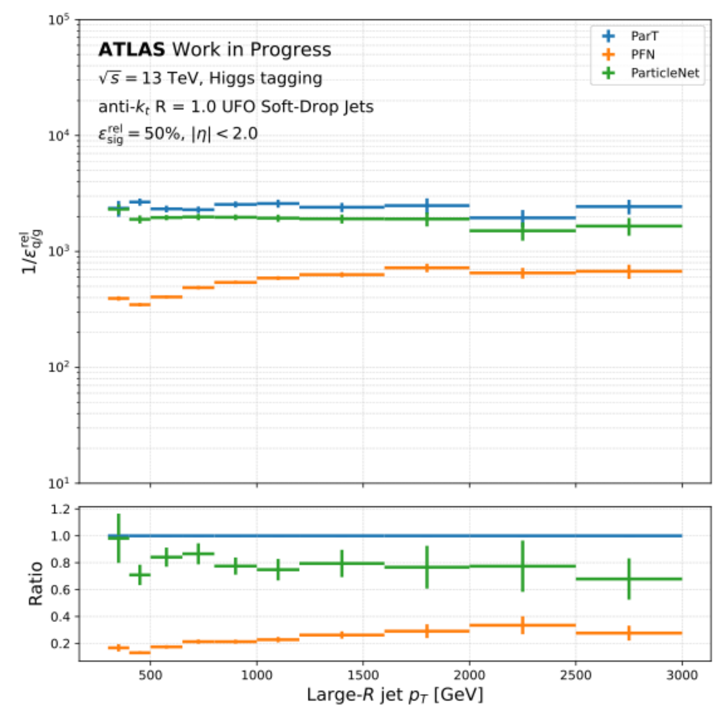
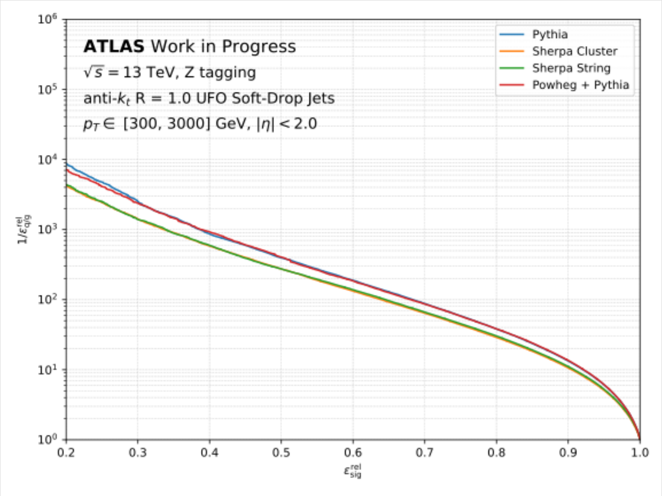


Fig. 24: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Higgs jets**.

COMPARISON TO OTHER MC GENERATORS

- Better rejection for samples including Pythia8.
- Comparable performance at higher p_T.



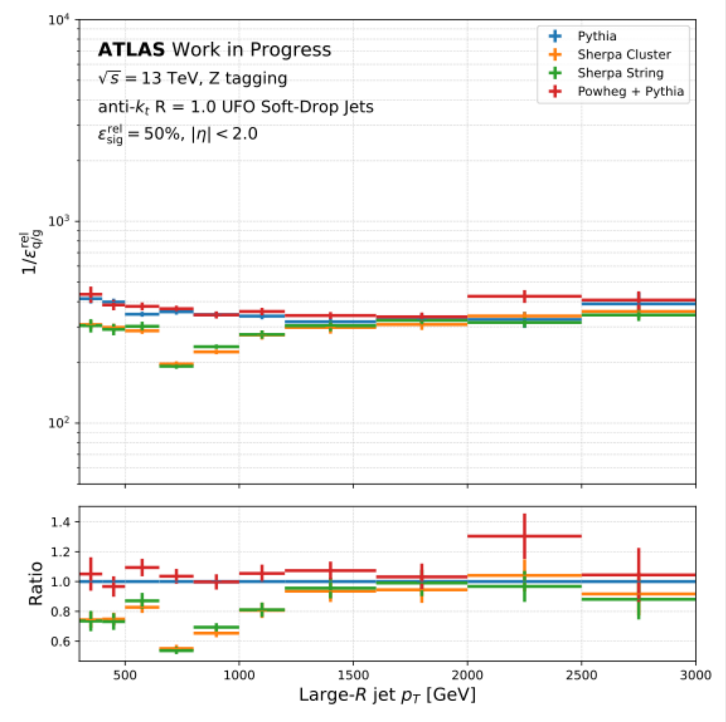
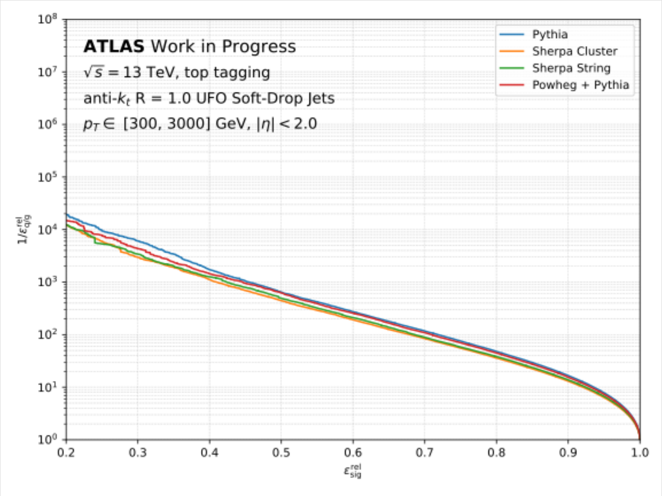
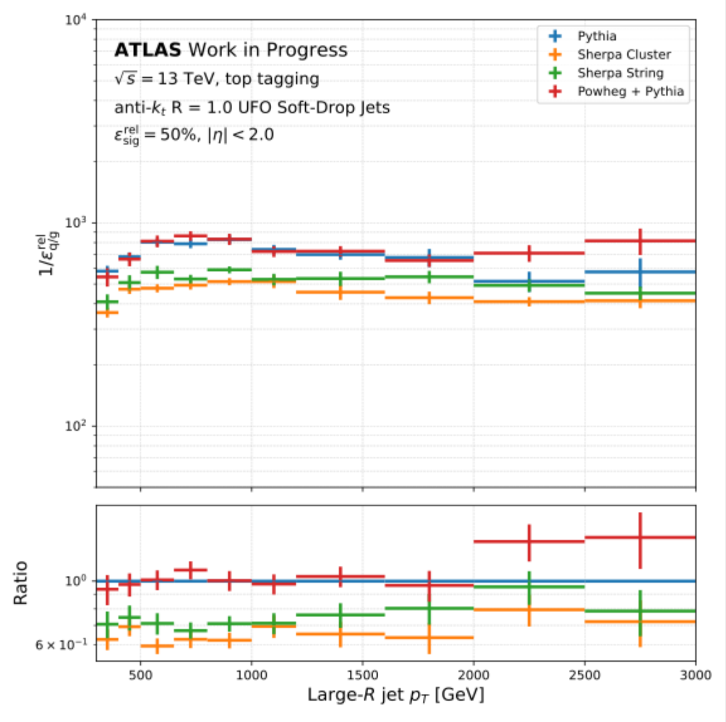


Fig. 25: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Z jets**.

COMPARISON TO OTHER MC GENERATORS



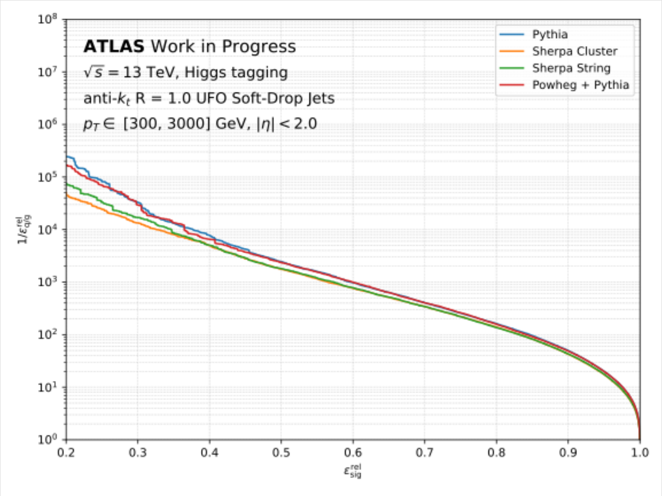


- Better rejection for samples including Pythia8.
- Comparable performance at higher p_T.

Fig. 26: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **top jets**.

COMPARISON TO OTHER MC GENERATORS

- Better rejection for samples including Pythia8.



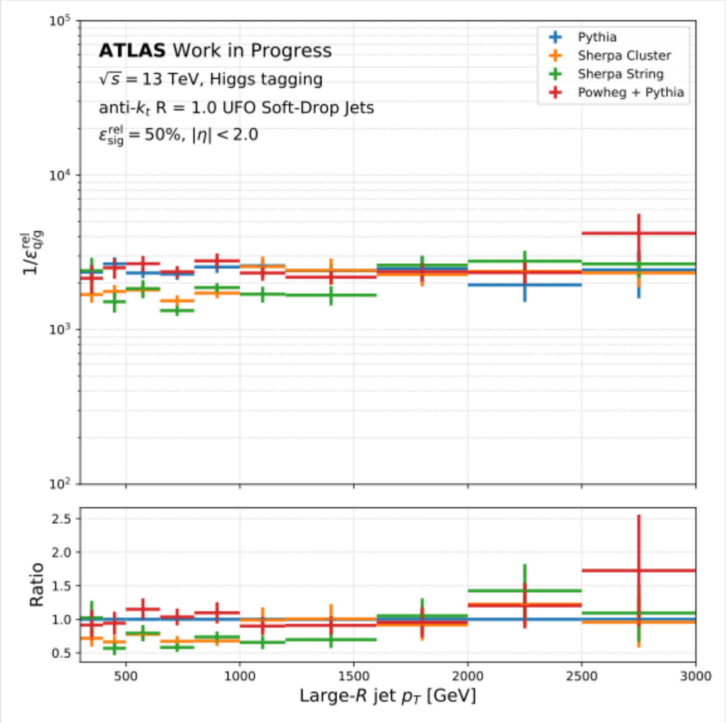
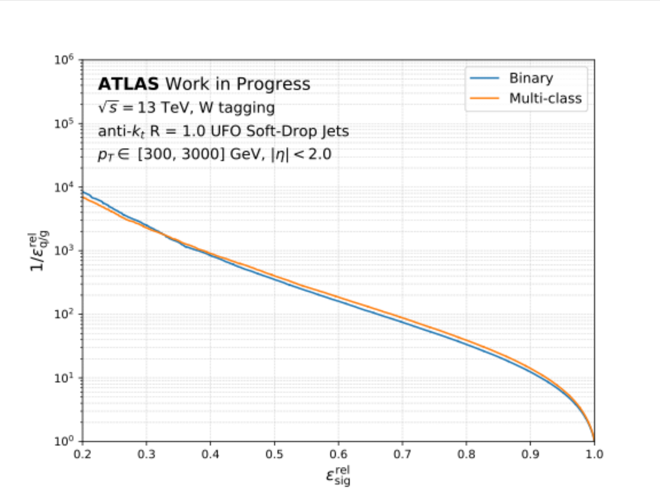


Fig. 27: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Higgs jets**.

COMPARISON TO BINARY PART

- A separate **binary ParT model** was trained for each signal:
  - QCD vs W, QCD vs Z, QCD vs top, QCD vs Higgs.
- Training and test statistics were kept the same as in the multiclass setup → results are directly comparable.
- Background rejection curves show that **binary and multiclass have similar performance.**
- The **multiclass model** is consistently **slightly better** across all jet types.



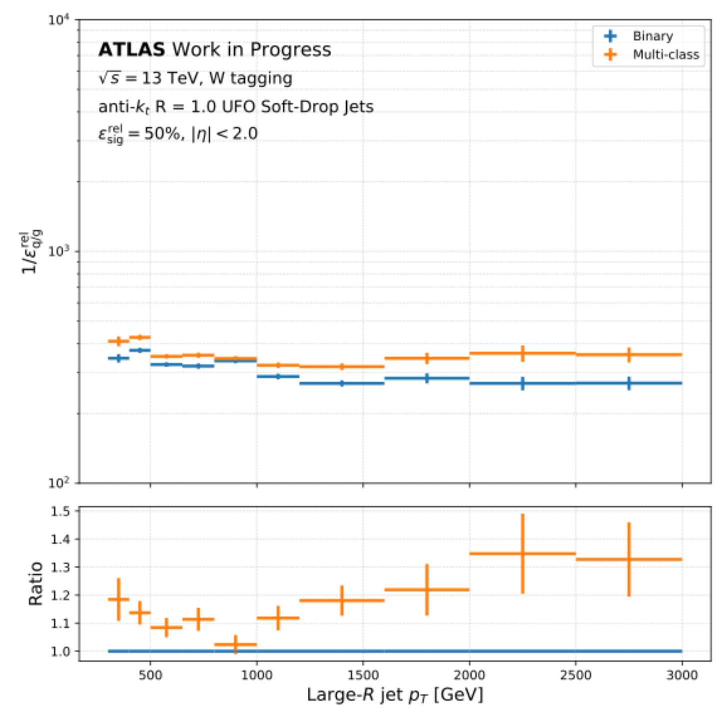
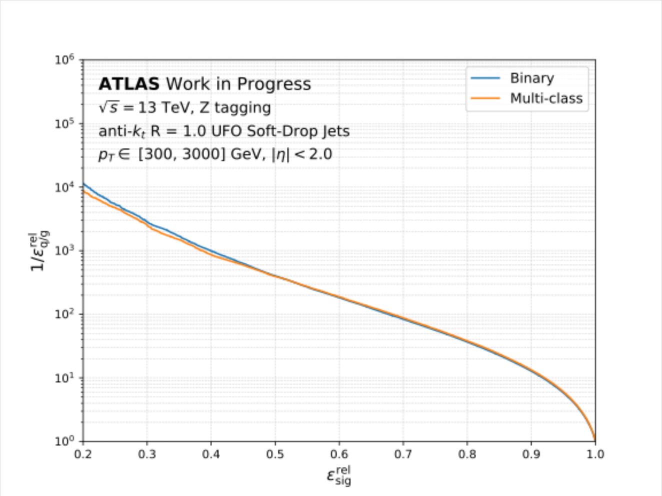


Fig. 28: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **W jets**.

COMPARISON TO BINARY PART

- A separate **binary ParT model** was trained for each signal:
  - QCD vs W, QCD vs Z, QCD vs top, QCD vs Higgs.
- Training and test statistics were kept the same as in the multiclass setup → results are directly comparable.
- Background rejection curves show that **binary and multiclass have similar performance.**
- The **multiclass model** is consistently **slightly better** across all jet types.



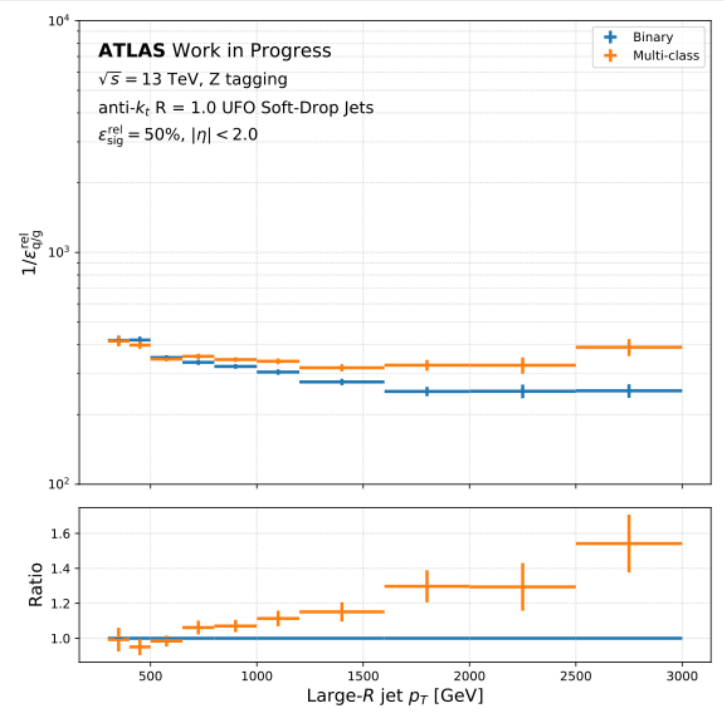
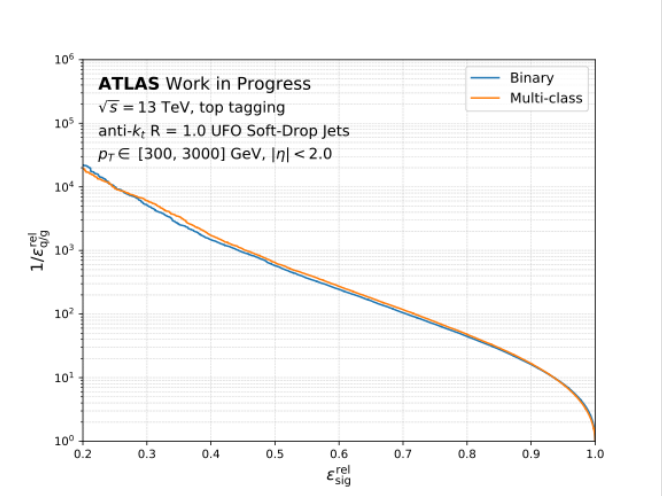


Fig. 29: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **Z jets**.

COMPARISON TO BINARY PART

- A separate **binary ParT model** was trained for each signal:
    - QCD vs W, QCD vs Z, QCD vs top, QCD vs Higgs.
- Training and test statistics were kept the same as in the multiclass setup → results are directly comparable.
- Background rejection curves show that **binary and multiclass have similar performance.**
- The **multiclass model** is consistently **slightly better** across all jet types.



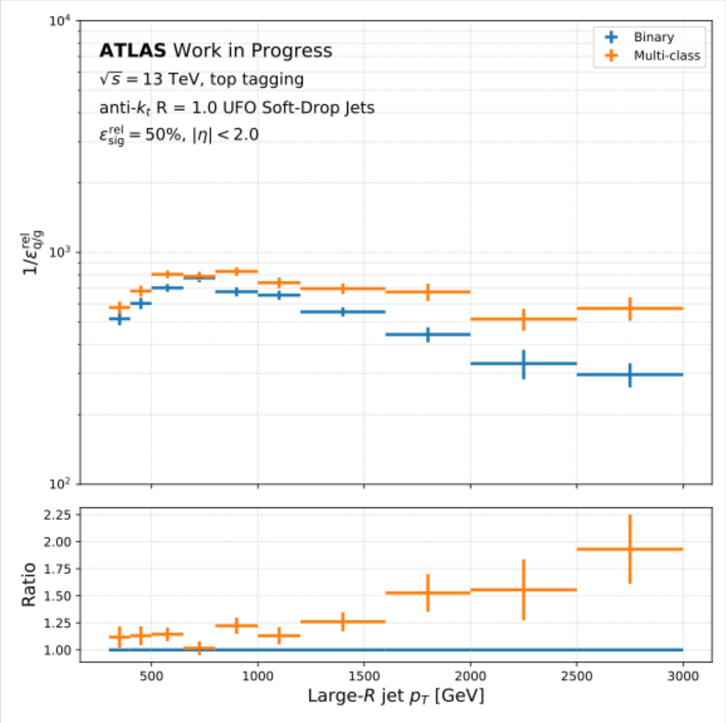


Fig. 30: QCD rejection vs signal efficiency and for all p_T spectra at 50% sig eff for **top jets**.