Identifying the origin of jets with machine learning

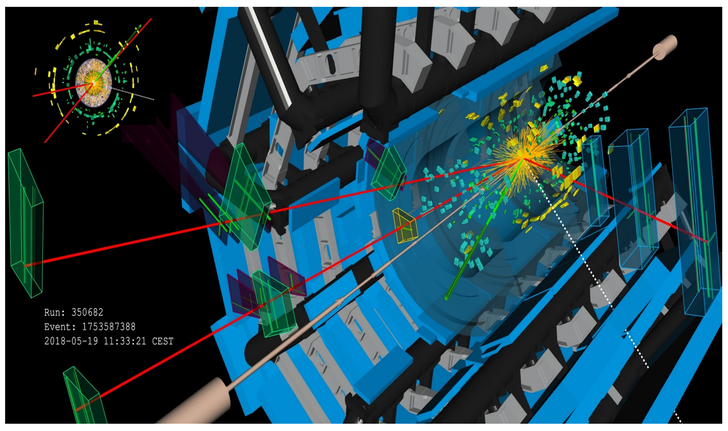
Zeyi Zhang, Leyi Dong, Ismael Sadiq, Ian Irwan Ambak and Qiren Zhao

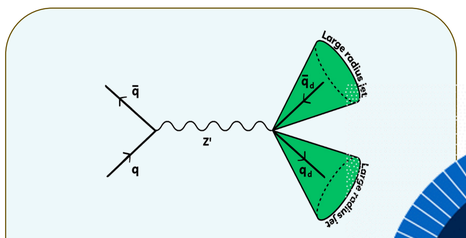
UCL Department of Physics and Astronomy & London Centre for Nanotechnology, 17-19 Gordon Street, London WC1H 0AH, UK

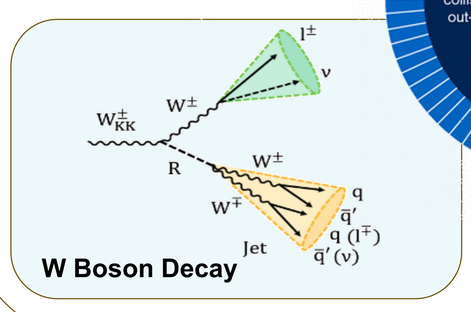
LHC Jets

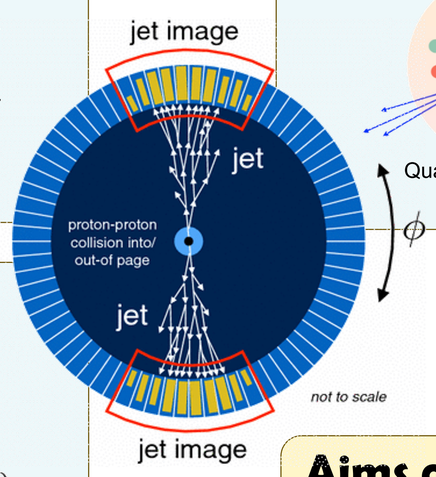
In the Large Hadron Collider (LHC), jets are produced through the high-energy collisions of protons. These jets arise when quarks and gluons—fundamental components of hadronic matter—are ejected, rapidly forming a stream of particles. These particle streams, known as jets, provide crucial insights into fundamental particle physics and help us explore the fundamental laws of the universe.

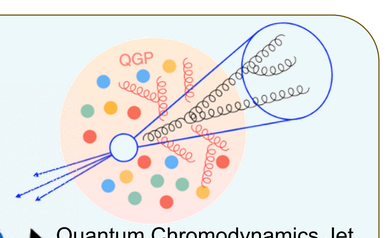
- **Collision Point**: Proton collisions result in the ejection of primary particles.
- **Detector Components**: Used to capture and measure particle properties. Includes tracking detectors, calorimeters, and particle identifiers.
- **Particle Trajectories**: Represent different types of particles or their energy levels.







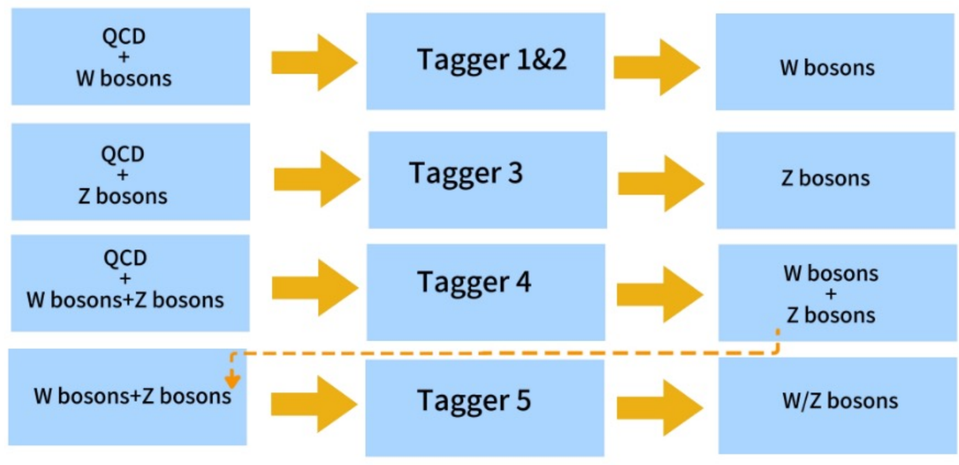




Aims of the project:

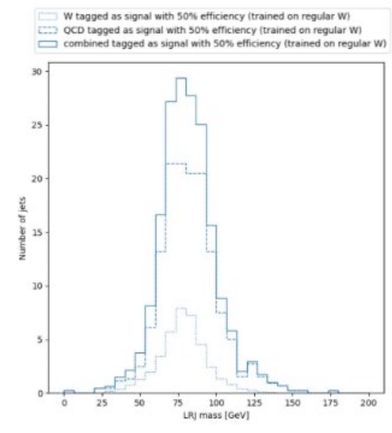
Distinguish between:
- QCD jets
- Hadronic decay of W and Z bosons

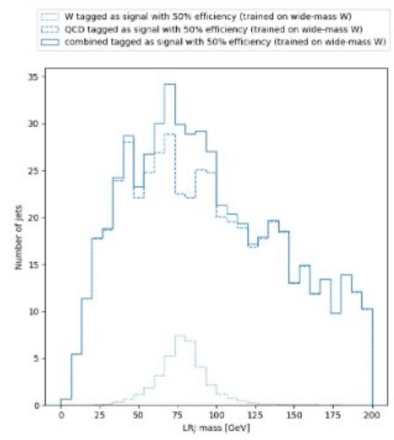
Tagger Training



- **Tagger 1 (Trained on Regular-Mass W Boson)**: Performs well but may suffer from limited generalization and accuracy due to a narrow mass window.
- **Tagger 2 (Trained on Wide-Mass W Boson)**: Shows better performance in practical applications, demonstrating greater adaptability.
- **W and Z Bosons Similar in Mass**: Difficult to distinguish
- **Uses Tagger 4's output as the input of Tagger 5,** treating one boson as the background and the other as the signal to differentiate between them.

Mass Sculpting

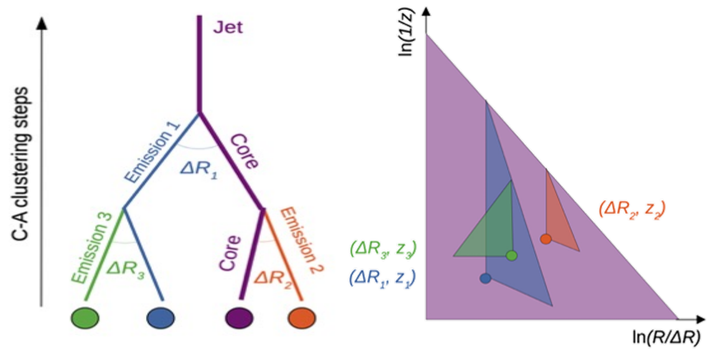




- After training with regular-mass W jets, tagger learns to **classify** jets by **mass** without considering **jet substructure**. Distribution clusters around W mass peak, showing mass sculpting - complicating signal extraction/background modelling (left)
- Mass sculpting is avoided by using a **wide mass distribution** training sample of W jets (right), although background rejection rate is lower. Overall mass distribution is no longer **distorted.**

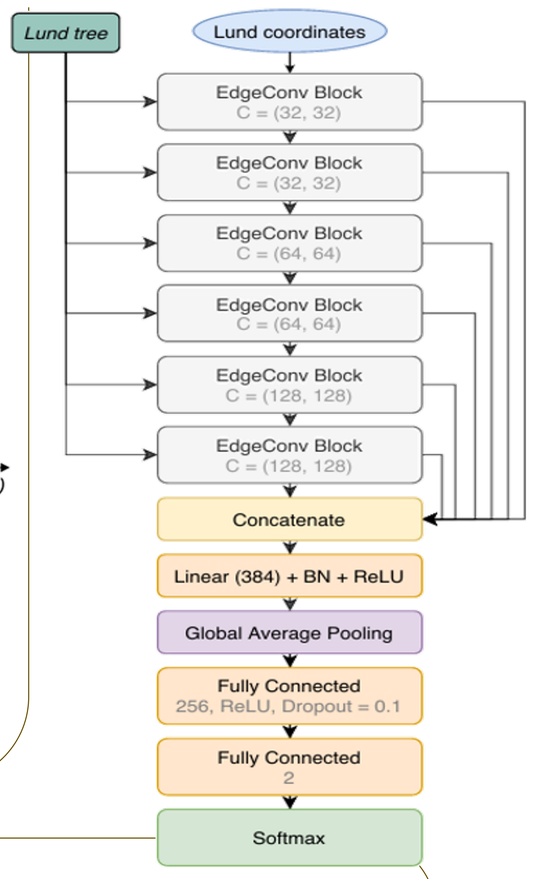
The Lund Jet Plane

A method for displaying the internal structure of jets by mapping the emission phase space of jets onto a two-dimensional plane. It effectively allows the study and identification of substructures and radiation patterns in jets within high-energy physics.



- Jet particles are sorted into a tree structure, at each branch: separation angle Δ, transverse momentum kt and momentum fraction z are stored.
- The branches are then plotted onto a graph with axes of ln kt and ln 1/Δ - the 2D Lund Plane
- Machine learning is used to classify origin of jets, using jet substructure and radiation signals.

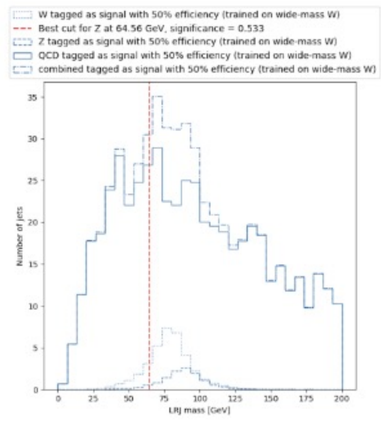
The Lund Net Model



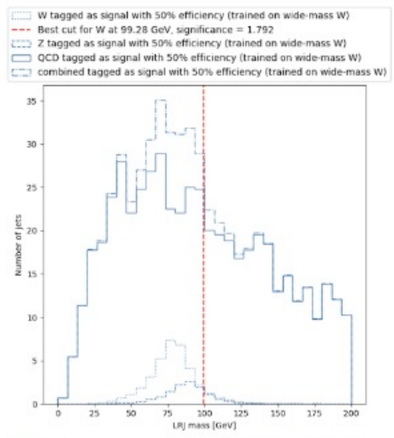
Mass Cuts

Establish a specific mass threshold used to differentiate types of jets.

- **Mass Cut for W Bosons (Figure b)**: Jets below this threshold (under the signal "bump") are identified as W bosons.
- **Mass Cut for Z Bosons (Figure a):** Jets above this threshold (under the signal "bump") are identified as Z bosons.



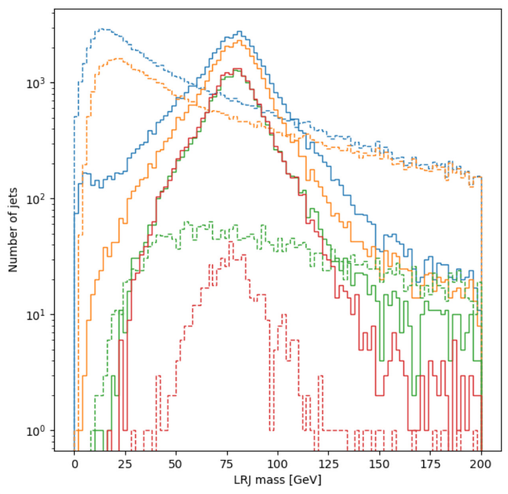
(a) The mass cut (54.56 GeV) with highest significance (0.533) for classifying Z jet region (on the right of the cut).

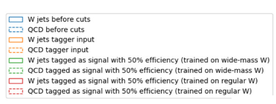


(b) The mass cut (99.28 GeV) with highest significance (1.792) for classifying W jet region (on the left of the cut).

Sample Normalization

- In practical scenarios, the number of QCD jets significantly exceeds that of W and Z jets, necessitating normalization.





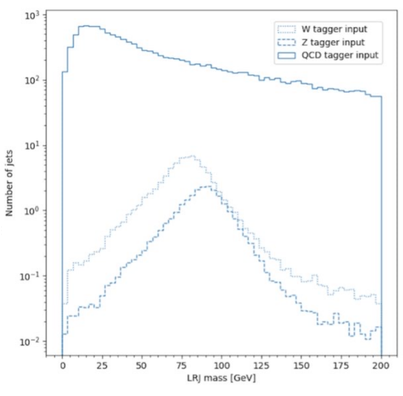
$$W_i = \frac{\sigma_i}{N_{gen_i}}$$

W_i: weights applied
σ_i: cross-section of the corresponding jet
N_{gen_i}: number of generated events

$$W_{QCD} = \frac{\sigma_{QCD}}{N_{gen_{QCD}}} \times \epsilon_{gen}$$

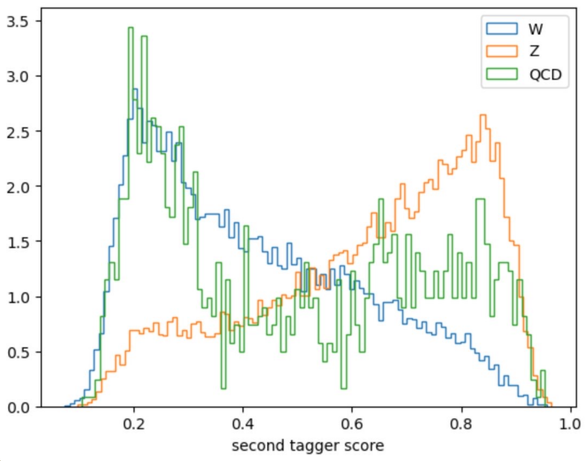
W_{QCD}: weights applied to the QCD jet sample
σ_{QCD}: cross sections
N_{gen}: numbers of generated events
ϵ_{gen}: generator filter efficiencies

After normalization, the W, Z, and QCD jet samples are now in realistic sizes

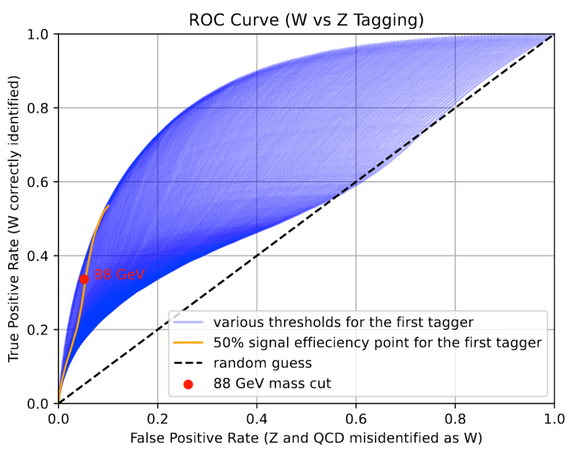


Second-Stage Classifier

- W and Z jets were labelled as signal and background, and a second classifier (Tagger 5) was trained.



- At 50% signal efficiency applied during training, there is some distinction between W and Z jets, but it is not significant



[1] B. Nachman, "Deep learning usage by large experiments," J. Phys. Conf. Ser., vol. 1085, p. 022002, Sep. 2018.

[2] F. A. Dreyer and H. Qu, "Jet tagging in the Lund plane with graph networks," J. High Energy Phys., vol. 2021, no. 3, pp. 1–23, Mar. 2021.

[3] ATLAS Collaboration, "Massive vector bosons also come in triplets," ATLAS, 2025. [Online]. Available: https://atlas.cern/Updates/Briefing/VVZ-Observation [Accessed: Mar. 17, 2025].

[4] ATLAS Collaboration, "When jets go dark: Identifying elusive 'dark jets' at ATLAS," ATLAS, 2025. [Online]. Available: https://atlas.cern/Updates/Briefing/Dark-Jets [Accessed: Mar. 18, 2025].