Machine Learning-Based Identification of b-Jet Origin (Top decay vs Gluon splitting) and Measurement of $t\bar{t}$+b-jets Using Run 2 + Run 3 Data

Abderahmane Maiza
Universität Würzburg

Abstract

The production of top-quark pairs in association with additional b-jets provides essential tests of quantum chromodynamics (QCD) predictions. This process constitutes a large irreducible background to $t\bar{t}H$ production which probes the top-Higgs Yukawa coupling.

In my PhD project, I aim to improve the measurement of $t\bar{t}$+ b-jets in the dilepton $e\mu$ channel using the full Run 2 dataset together with early Run 3 data. The analysis focuses on improving b-jet origin identification through machine learning methods building upon previous Run 2 approaches based on likelihood methods. I employ the GN2 model, a Transformer-based architecture originally developed for flavour tagging to classify b-jets according to their origin: top decay versus gluon splitting. This framework allows correlations between jets and event-level observables to be exploited effectively.

The project involves training the model on dedicated Monte Carlo samples, validating its performance and applying the improved classifier on simulated data. The expected outcome is a more accurate separation of b-jets from top decays and those from additional production, leading to a more precise cross-section measurement using observables enriched with improved origin-related information such as the p_T of b-jets assigned to top quarks. These improvements will directly benefit future analyses and other studies involving multi-b-jet final states.