Differentiable Modelling of Systematic Uncertainties in ATLAS Object Corrections

IRIS-HEP Fellowship Proposal

Fellow: Cody Tanner — **Institution:** University of Washington
Mentor: Gordon Watts — **Period:** June – September 2025

Motivation and Background

Modern collider analyses at the LHC must contend with two compounding challenges: the extreme rarity of signal processes and the systematic uncertainties introduced by detector response and reconstruction algorithms. In the ATLAS experiment, the associated production of a Z boson with two hadronic jets ($Z \to q\bar{q}$) occurs at a rate of roughly ten events per ten thousand recorded collisions, making precise signal–background separation essential. At the same time, mis-modelling of jet energy scale, angular resolution, and related observables can bias any downstream classifier trained naïvely on uncalibrated inputs.

This project addresses both challenges within a unified, differentiable machine learning framework. A calibration network has been developed that maps reconstructed particle four-momenta to systematically calibrated counterparts, trained on ATLAS Monte Carlo simulations with detector and pile-up effects. Validation against the well-known Z boson mass peak ($m_Z \approx 91.2$ GeV) confirms that the calibrations preserve known physical structure while reducing systematic bias. A downstream classification network, trained on these corrected inputs to separate $Z \to q\bar{q}$ signal from multijet and other backgrounds, is near completion.

The principal outstanding obstacle is that the current best-pair selection procedure – identifying the two jets most consistent with $Z \to q\bar{q}$– relies on a discrete $\arg\max$ operation, which is non-differentiable and severs the gradient path between the calibration and classification stages. Replacing this with a differentiable version is one the technical goals of this summer.

Proposed Work

Differentiable Jet-Pair Selection

The non-differentiable best-pair selection will be replaced by a smooth function, restoring end-to-end gradient flow through the full pipeline. Candidate approaches include soft attention mechanisms that produce a weighted combination of all jet pairs, and discrete relaxations such as Gumbel-softmax. The choice of method will be made in consultation with the mentor and informed by a survey of the literature on differentiable combinatorial operations in jet physics. Performance will be benchmarked against the hard-selection baseline using receiver operating characteristic (ROC) area and signal significance metrics.

End-to-End Pipeline Training

With a differentiable pair selection in place, the calibration and classification networks will be trained jointly, with freezing the calibration network, and only changing the weights of the classification network. This allows us to be confident the physics is remaining unchanged, while the classification changes.

Systematic-Aware Evaluation

The trained pipeline will be evaluated under variations of the leading systematic uncertainties to quantify the robustness gained by the calibration stage relative to an uncalibrated baseline classifier. Results will be reported as a function of signal significance, and then by a cross section measurement.

Documentation and Publication

Pending discussion with the mentor, the target output is a paper suitable for submission to a journal such as, documenting the full framework, validation, and performance results.

Timeline and Deliverables

Period	Tasks	Deliverables
June, week 0	Literature survey on differentiable combinatorics; review of existing pipeline codebase; environment setup.	Written summary of candidate pair-selection methods with recommended approach.
June, weeks 1–2	Implement chosen differentiable pair-selection method; unit tests and gradient-flow verification.	Working differentiable pair-selection module; gradient-check notebook.
June, weeks 3–4	Integrate module into full pipeline; begin joint (end-to-end) training runs on ATLAS simulation. Hyperparameter optimisation; loss-weighting study; Z-peak validation of jointly trained model.	Baseline joint-training results; comparison with hard-selection Optimised model checkpoint; m_{jj} validation plots. performance.
July, weeks 1–2	Build uncertainty network, and acquire data to train on. Begin training.	Loss functions of new networks, along with ROC.
July, weeks 3–4	Full pipeline training. Freeze weights of all but classification, train. Figure out good loss function.	Optimised model checkpoint; m_{jj} validation plots. Significance comparisons.
August, weeks 1–2	Paper/note writing; code clean-up and documentation; IRIS-HEP midpoint check-in presentation.	Full performance tables and figures suitable for a paper.
August, weeks 3–4	Paper writing; more code clean-up, post to GitHub	Draft manuscript or internal note; public code repository.
September	Final IRIS-HEP fellowship presentation; submission if paper target is met.	Final manuscript submission (target); recorded fellowship talk; public software release.