Early Career Pre-proposal:
Enabling Calibrations of Machine Learning Approaches (ECMLA)

Walter Hopkins, Physicist
Argonne National Laboratory
(630) 252 7551, whopkins@anl.gov
Year Doctorate Awarded: 2013
Number of Times Previously Applied: 2
Topic Area: Experimental Research at the
Energy Frontier in High Energy Physics
Eligibility Extension Requested: No
FOA Number: DE-FOA-0003176

Results from the Large Hadron Collider (LHC) experiments have verified the predictions of the highly successful Standard Model (SM), culminating with the discovery of the Higgs boson. However, the SM lacks an explanation for several observed phenomena (e.g., dark matter, the matter-antimatter asymmetry, etc) motivating the search Beyond the Standard Model (BSM) physics. The LHC Run 1 and Run 2 extended the energy reach (first from ~ 2 TeV at the Tevatron to 7 TeV and then to 8 TeV and finally to 13 TeV) of searches for direct production of BSM, significantly. Future LHC upgrades will no longer include substantial increases in energy and move High-Energy Physics into the precision era, with a doubling (Run 3) and then a tenfold increase (High Luminosity-LHC, HL-LHC) of the Run 2 data set. The lack of evidence for BSM physics after Run 2 and the unprecedented data volume and precision that the HL-LHC will deliver motivates a change in strategy for finding evidence of BSM effects.

Precision measurements of SM processes, especially interaction involving Higgs bosons, can probe for BSM effects resulting from particles which may have large masses that prevent them from being directly produced at the LHC. An essential aspect to improving the precision of measurements, which will maximize sensitivity to BSM physics, is the calibration of physics objects (i.e., photons, jets, etc). This calibration involves evaluating identification (ID) efficiencies and resolutions in Monte Carlo (MC) simulations and correcting these quantities to what is observed in data. This proposal presents **the development of machine-learning-based physics object ID algorithms and resolution evaluation that include calibration considerations and that are robust against mismodelling. This will be achieved by employing decorrelation techniques to make the ML algorithms independent from quantities that are needed for the calibration procedure and invariant to quantities that are mismodelled.** The proposed techniques could also improve physics object resolution and ID at the trigger level for the HL-LHC upgrade, especially given the recent work on fast inference on field-programmable gate arrays (FPGAs).

The Higgs boson self coupling and the coupling to the top quark are sensitive to BSM effects. The associative production of a top quark pair with a Higgs boson ($t\bar{t}H$) is sensitive to both of these couplings [1] and thus the $t\bar{t}H$ total and especially differential cross section measurements, e.g., as function of Higgs p_{T}, can serve as probes for BSM physics. The latest ATLAS $t\bar{t}H(\to \gamma\gamma)$ differential cross section results [2] remain limited by the statistical uncertainty in each kinematic

observable bin. However, the photon ID and resolution uncertainties are the dominant detector-based systematic uncertainties and are currently on par with the other dominant source of systematic uncertainty, theory uncertainties. Reducing these systematic uncertainties will become significantly more important, not only for $t\bar{t}H$ but for all Higgs measurements, as ATLAS drastically increases the size of its data set near the end of the HL-LHC data taking. Thus, the unprecedented data volume of the HL-LHC offers an opportunity to **improve the precision cross section measurements as a function of kinematic variables in channels with high $t\bar{t}H$ purity, i.e., where the Higgs decays to two photons ($t\bar{t}H(\to \gamma\gamma)$).**

The proposed techniques will first be applied to photon ID and resolution because these are key ingredients to maximizing the sensitivity of Higgs precision measurements. Currently, the ATLAS photon ID is one of the few remaining physics object ID algorithms that does not employ machine learning (ML). Studies have been performed within ATLAS that showed a potential 20% improvement in both ID efficiency and resolution. However, as soon as the calibration of the ID efficiency and resolution was taken into account, the gain in efficiency and enhancement in resolution were lost. For the ID, ML approaches were found to be correlated with quantities that needed to be independent of the ID for the calibration while for photon resolution it was found that ML methods were sensitive MC mismodelling of shower shapes. Recent developments in decorrelating ML algorithms from particular quantities [3, 4] yield an opportunity to address these challenges when calibrating photons and other objects. This proposal presents **the development ML-based photon ID and resolution evaluation that can be calibrated and are robust against mismodeling. This would not only improve the $t\bar{t}H(\to \gamma\gamma)$ measurement but all ATLAS measurements involving $H \to \gamma\gamma$.**

The PI's experience in ML, e.g., as one of the ATLAS ML Forum conveners, as well as in the LAr calorimeter will aid the success of this proposal. The PI will also draw from his work within the ATLAS SUSY group, both as a leader of flagship $t\bar{t} + E_{\mathrm{T}}^{\mathrm{miss}}$ searches [5–8] and as a subgroup convener of the ATLAS SUSY strong production group. Additionally, the PI will leverage the computing resources and ML expertise at the Argonne Leadership Computing Facility (ALCF).

Project Objectives

The objective of the proposed work is to maximize the power of precision measurements by improving physics object ID and resolution evaluation by incorporating calibration consideration into ML algorithms. This approach will first be applied to the ATLAS photon ID efficiency and energy resolution which will improve the $t\bar{t}H(\to \gamma\gamma)$ differential cross section measurements.

- Develop method to incorporate calibration considerations into ML physics object ID

  - Using recent decorrelation schemes that have already been deployed for large-R jet ID

- Apply decorrelation technique to calibrate photon ID efficiency

- Apply decorrelation approach to calibrate photon energy resolution

- Perform updated $t\bar{t}H(\to \gamma\gamma)$ differential cross section measurement with improved photon ID and resolution

- Improve photon ID and resolution calibration by using k-means clustering to bin in more variables than η and p_{T} which can affect the calibration.

- Deploy tool that can calibrate other physics objects.

References

[1] F. Maltoni, D. Pagani, A. Shivaji, and X. Zhao, *Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC*, The European Physical Journal C **77** (2017).

[2] ATLAS Collaboration, *Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \to \gamma\gamma$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment*, Journal of High Energy Physics **2023** (2023), ISSN: 1029-8479, URL: http://dx.doi.org/10.1007/JHEP07(2023)088.

[3] G. Kasieczka and D. Shih, *Robust Jet Classifiers through Distance Correlation*, Phys. Rev. Lett. **125** (12 2020) 122001, URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.122001.

[4] G. Louppe, M. Kagan, and K. Cranmer, *Learning to Pivot with Adversarial Networks*, 2017, arXiv: 1611.01046 [stat.ML].

[5] ATLAS Collaboration, *Search for direct pair production of the top squark in all-hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **09** (2014) 015, arXiv: 1406.1122 [hep-ex].

[6] ATLAS Collaboration, *ATLAS Run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider*, Eur. Phys. J. C **75** (2015) 510, [Erratum: Eur.Phys.J.C 76, 153 (2016)], arXiv: 1506.08616 [hep-ex].

[7] ATLAS Collaboration, *Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at $\sqrt{s}$=13 TeV with the ATLAS detector*, JHEP **12** (2017) 085, arXiv: 1709.04183 [hep-ex].

[8] ATLAS Collaboration, *Search for a scalar partner of the top quark in the all-hadronic $t\bar{t}$ plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. C **80** (2020) 737, arXiv: 2004.14060 [hep-ex].