PhD Annual Report

PhD Student: Lucrezia Rambelli (XXXVIII Cycle)
Supervisors: Andrea Coccaro, Francesco Armando Di Bello

15th September 2023

The work i've done in this first phd year can be divided into three separate research projects:

1. The ATLAS Qualification Project
2. Studies of fast inference for the ATLAS trigger system
3. Differentiable programming techniques for including physics information in ML-based model training

1 The ATLAS Qualification Project (AQP)

In the ATLAS experiment the Qualification Project is considered as mandatory to become part of the collaboration's author list, and each new researcher wishing to enter in the collaboration is asked to work on such a project for one year for at least the 50% of its time. These AQPs serve a dual purpose: they contribute to the specific service task assigned to each group within the collaboration, while also aiding qualifying students and newcomers in becoming familiar with the essential instruments for their future roles within the group. My AQP is chosen in the Flavour Tagging Group in the context of the release of the next-generation of the Xbb and Hbb-cc taggers. These taggers represent innovative tools designed to identify and classificate final states involving pairs of b- and c-quarks. In this scenario, the two jets produced by the hadronization of these quarks are treated as a single large-radius jet and analyzed using Graph Neural Network based models. When a new model is developed and trained it relies only on Monte Carlo simulations, and is crucial to develop a calibration technique for computing Scale Factors (SFs) for relating the performances in simulations to the performances observed in data.

In this setting, I have undertaken the task of performing calibrations using a novel approach that had never been applied to these types of taggers before. This innovative technique relies on a Monte Carlo-based calibration method, enabling the assessment of uncertainties across all flavor categories, particularly those that typically prove challenging to calibrate using real data. This technique, in essence, involves analyzing the performance of flavor tagging efficiency for these taggers under various Monte Carlo (MC) inputs. It entails comparing the performance achieved with the nominal MC data with the ones obtained when different systematic uncertainties, linked to variables that influence the tagger response, are introduced to the MC data prior to generating model predictions. Precisely what I have done until now is using the official software developed within the ATLAS Flavour Tagging Group [1], for understanding firstly how to add systematics to a certain sample, and then how to inspect the results. I have expanded the functionalities of the software framework being used for studying the flavor tagging performance in simulation, adding the possibility to compute the effect of a given systematic variation on the final flavor tagging discriminant, producing then a pipeline for obtaining, starting by these new modified outputs, the final SFs that will have to be considered when real data will be used.

2 Studies of fast inference for the ATLAS trigger system

Experimental particle physics demands a sophisticated trigger and acquisition system capable to efficiently retain the collisions of interest for further investigation. Heterogeneous computing with the employment of FPGA cards may emerge as a trending technology for the triggering strategy of the upcoming high- luminosity program of the Large Hadron Collider at CERN. In this context I've worked using new commercial tools for implementing novel algorithms for selecting Beyond Standard Model (BSM) processes in the ATLAS Muon Spectrometer (MS) High Level Trigger (the HLT is the last step in the ATLAS detector of data selection after which data are stored offline) on FPGA accelerators. Precisely I've defined and trained a Deep Neural Network (DNN) based trigger algorithm for predicting the radial decay length associated to events

[1] `https://gitlab.cern.ch/atlas-flavor-tagging-tools/training-dataset-dumper`

containing with final states produced by the decay of a neutral Long Lived Particle (LLP), for selecting their signatures with respect to SM ones. After the dataset simulation and the model definition and training, I have produced a pipeline for quantizing, compiling and deploying the model over FPGAs accelerators using a commercial tool provided by Xilinx, Vitis-AI, comparing at the end the processing times performances and the prediction ones with the ones obtained on CPU (currently used in the ATLAS HLT) and GPU boards. These studies, with their results, are summarized in an article submitted to the journal *Machine Learning: Science and Technology* and available on the arXiv [2]. Also with these studies with the Vitis-AI tool I have worked in the Machine Learning testbed ATLAS group investigating different and newer tools like Zebra, analyzing the same problem (and model) as before but with a different tool used for the conversion and deployment, comparing the obtained performances with the Vitis-AI ones, reporting the obtained results to the group and to the Zebra team.

3 Differentiable programming techniques for including physics information in ML-based model training

The last part of my 1st year work is always related to the deployment of novel ML-based models for trigger systems, but with a different approach: using specific novel libraries to incorporate physics information during the training of a model for pattern recognition in the MS.

For doing that I have used the the Google library JAX [3], which is a high-performance numerical computing library that offers several advantages with respect to the standard ones when running on GPU boards, the main ones are speed (as it's based on XLA and allows Just in Time compilation and execution), and the possibility of auto differentiate functions, i.e. the possibility of get the gradient of a function and find its derivatives also of higher orders. Also there are multiple libraries built on top of JAX that can be used to define and train complex ML models, like Haiku, Jraph for graph neural networks, Optax for the optimizers definition.

Using these libraries I've deployed (and I'm still working on it) a ML-based model for pattern recognition in the ATLAS MS, i.e. the recognition of the muon patterns inside the detector, that are related through the magnetic field in which the MS is enveloped to their momenta and charge. For this purpose I have used a Graph Attention Network (GAT) based model for working on the last chambers of the MS (that are the ones with less noise), followed by a Multi Layer Perceptron (LMP) one for propagating the outer layers information to the internal (and noisier) ones.

GATs are novel neural network architectures that operates with graph based data where, using custom functions for the graph nodes connections is possible to weight them with respect to their neighborhood, an MLP is on the other hand a model in which the input are vector-structured data and where the model neurons are fully connected between them and followed by a non-linear activation function.

The information about the physics of the problem is added during the training thanks to the JAX auto differentiation instruments. In fact, with the output of the first GAT a circular fit is performed with the aim of instructing the model to recognize that distinct input hits correspond in reality to more physically objects such as the circular tacks produced by a charged particle in a magnetic field, using then these results for defining the MLP input and accomplishing the complete pattern in the events.

Attended Courses and Exams Given:

- *Theoretical Physics*: waiting for the end of the course, exam not given yet
- *Machine Learning for Particle Physics*: for the exam I've defined and trained a Graph Neural Network for selecting (tagging) events with different jets in the final state
- *The double trouble of the missing matter in the Universe*: exam not given yet

Research Outputs:

- *Incontri di Fisica delle Alte Energie* [4]: 'Inferenza su FPGA di algoritmi di trigger basati su Deep Neural Network per la selezione di particelle a lunga vita media ai Colliders' (Poster Session)
- 'Fast Neural Network Inference on FPGAs for Triggering on Long-Lived Particles at Colliders'.

[2] `https://arxiv.org/abs/2307.05152`
[3] `https://jax.readthedocs.io/en/latest/index.html`
[4] `https://agenda.infn.it/event/34702/`