IRIS-HEP Project Proposal

# Conditional Generation of High-Energy Particle Collisions with Graph Networks

Mentor: Javier Duarte
Anni Li

## Introduction

The high-energy particle collisions generated by the Large Hadron Collider (LHC) at CERN provide rich data for high-energy physics research. Through accelerating and colliding two protons in LHC, elementary particles like quarks and gluons are generated and continuously radiate into sets of particles. After they cool down, a process called hadronization would happen, and the particles then combine to form some stable hadrons, and the final set of hadrons produced is referred to as a jet (R. Kansal et al., 2021). Jets are sparse distributions of particles, and they have complex patterns, so we can identify the generation of some rare particles, such as the Higgs Bosons, by analyzing jets (R. K. Ellis, 2003). Traditional Monte Carlo simulations of such collisions and phenomena consume a significant and increasingly unsustainable amount of computational resources at CERN, hence new machine learning methods are being explored to both improve and speed up these simulations. In particular, graph neural networks (GNNs) have been found to be most successful for a number of computational tasks in high energy physics (J. Shlomi et al., 2020), and recently our group has developed a GNN for simulating jets (R. Kansal et al., 2021). In this project, I will be extending this work and exploring applications of Auxiliary Classifier Generative Adversarial Networks (ACGAN) (A. Odena et al., 2016) in GNNs for conditional generation of jets.

## Generative Adversarial Networks

The Generative Adversarial Network (GANs) is a popular generative model in machine learning, where a generator, which produces the simulations, and a discriminator, which tries to distinguish between real and generated simulations, are trained together adversarially. By improving the discriminator's ability to distinguish, we can simultaneously improve the generator's ability to produce models closer to real jets. Recently, an advanced training method for GANs has been developed, called Auxiliary Classifier GAN (ACGAN), in which the discriminator outputs both the probability of the jet to be real and predictions of the jet's features. This allows us to generate jets of specific types and momenta, as would be required in practical applications at LHC. Some scientists have applied ACGAN to image-based data, but such ACGAN has not been developed yet for graph neural networks (GNNs).

## Graph Neural Networks

A standard way to represent jets for machine learning applications is simply to map particle constituents in a three-dimensional image. Since jets are sparse and irregular, this method is complex, taking a large amount of storage and computations. Comparatively, a graphical representation describes jets as nodes & edges, which represents geometrical features and relations of the jets in a more efficient and faithful way, and can be operated with graph neural networks (GNNs). Recently, the Message Passing GAN (MPGAN) was developed, which is a GAN using such GNNs for generating jets.

**Conditional Graph GAN**

In this proposed project, my goal is to extend our work and explore applications of Auxiliary Classifier Generative Adversarial Networks (ACGAN, A. Odena et al., 2016) in GNNs for conditional generation of jets. This network combines features of both MPGAN and ACGAN, developing a conditional classifier GAN using graph neural networks. The benefits of this method are: 1. compared with traditional GANs, it will be able to model jets conditionally (e.g. jets of specific types or momenta); 2. It's more efficient and performant for jet simulations due to the use of GNNs instead of image-based models.**The expected deliverable a strong ACGAN+MPGAN algorithm using a GNN to conditionally generate jets with high fidelity.** If successful, this model will be able to generate jets in a more efficient way with controllable features of the jets, which can serve as a strong tool for future applications in high energy particle physics at CERN. Overall, the development of such method of conditional generation could have an improvement on computation at CERN and provide more useful data for furthering high-energy-particle physics research.

**Proposed Timeline**

Week 1-2: Read relevant papers and talks on the application of ACGAN and MPGAN in high-energy physics; Go through tutorials on algorithms of GANs in python and other training resources from Duarte Lab and IRIS-HEP.

Week 3-6: Implement and test a baseline MPGAN+ACGAN model, become familiarized with UCSD/SDSC GPU resources and LHC data.

Week 7-9: Training, testing, and tuning parameters of the model using LHC data; Validate and compare conditional generation with previous MPGAN models.

Week 10-12: Expected outcome: a strong ACGAN+MPGAN algorithm using GNN to generate conditional models of jets with high fidelity; write a research report and poster; prepare the presentation; consult mentors for suggestions.