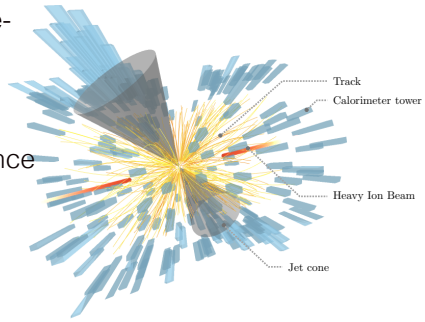
Robust and Generalizable Background Subtraction on Images of Calorimeter Jets using Unsupervised Generative Learning

Yeonju Go[1], Dmitrii Torbunov[2], Yi Huang[2], Shuhang Li[3], Timothy Rinn[1], Haiwang Yu[1], Brett Viren[1], Meifeng Lin[2], Yihui Ren[2], Dennis Perepelitsa[4], Jin Huang[1]
[1]Physics Department & [2]AI Department, Brookhaven National Laboratory, Upton, NY
[3]Physics Department, Columbia University, [4]Physics Department, University of Colorado Boulder

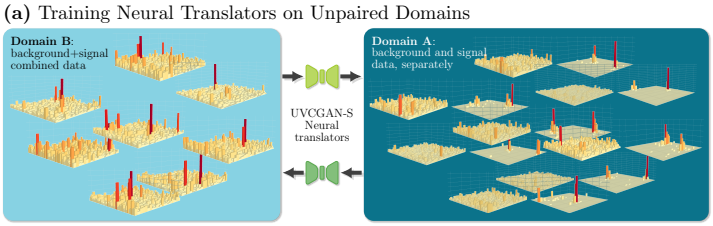
Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

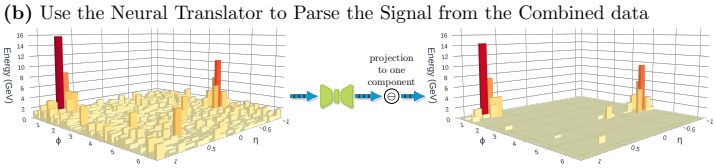
- **Jet background subtraction** using ***UVCGAN-S*** (Stratified U-Net Vision Transformer cycle-consistent Generative Adversarial Network): unsupervised and unpaired image-to-image (I2I) translation [1]
  - ➡ use **detector data** directly; preserves **full dimensional information**
  - ➡ can use **experimental data (without label) for training**: mitigate simulation-dependence



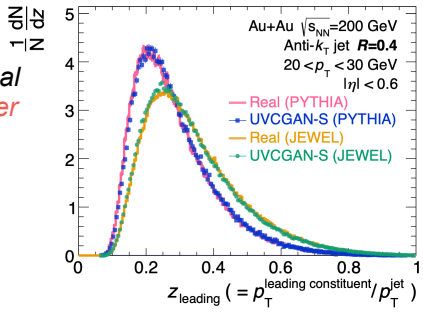
(a) Training Neural Translators on Unpaired Domains



(b) Use the Neural Translator to Parse the Signal from the Combined data



- ***UVCGAN-S***
  - ➡ *unsupervised, unpaired learning allows training directly on data*
  - ➡ *compared to conventional methods, up to ~2x better energy resolution, fake rate 80% → 7%, efficiency 74% → 92%*
  - ➡ *Robustness to out-of-distribution data*



Trained on PYTHIA and tested on JEWEL → successfully reproduce JEWEL's distributions including substructures

Paper: Y. Go et al. accepted by Scientific Reports, arXiv:2510.23717
Source Code: github.com/LS4GAN/uvcgan-s
Dataset: zenodo.org/records/17594612

Phase-I: Characterizing Jet Modification in the Quark Gluon Plasma Using Unsupervised, Cycle-Consistent Generative Learning

Baruch College, CUNY: Stefan Bathe (PI), **University of Colorado Boulder**: Jamie Nagle, Dennis Perepelitsa
Columbia University: William Zajc, **Brookhaven National Laboratory**: Yeonju Go, Dmitrii Torbunov

Abstract

This project will develop a new AI approach to measure **jet modification in Quark Gluon Plasma (QGP)** in a **more complete and less biased way** than is possible with conventional analysis techniques. Instead of focusing on a small number of predefined characteristics, the AI will learn directly from detector data to identify the full pattern of QGP-induced jet modifications. The approach uses recent advances in **unsupervised, unpaired cycle-consistent generative AI**, UVCGAN-S, allowing it to discover these patterns without requiring examples labeled by simulations. This makes the method well suited for application to **experimental data** and opens the possibility of revealing previously unrecognized signatures of the QGP.
During Phase I, the team will demonstrate the method using realistic simulations of the sPHENIX experiment at RHIC and quantify its performance relative to existing techniques. If successful, the resulting AI framework could become a powerful new tool for studying strongly interacting matter at RHIC, the Large Hadron Collider, and future facilities such as the Electron-Ion Collider.

- **Physics Goal**
  - ➡ Learn the **full QGP-induced jet modification** in high-dimensional calorimeter images using **cycle-consistent Generative AI** trained from experimental data
  - ➡ **No pre-selected jet observables**
  - ➡ Capture **correlated, multi-scale modification patterns**
  - ➡ Event-by-event mapping: **vacuum-like jets ↔ QGP-modified jets**
    - Ultimately enable **data-driven extraction from p+p and Au+Au data**

- **Phase-I Strategy**
  - ➡ Train on **PYTHIA (unquenched jets) ↔ JEWEL (quenched jets)** with full sPHENIX detector simulation
    - Demonstrate and validate the method using controlled simulations with known physics
    - Establish a pathway toward direct training on **p+p ↔ Au+Au experimental data**
  - ➡ Multi-channel calorimeter images: EMCal + IHCal + OHCal
  - ➡ Explore alternative architectures, e.g. **cycle-consistent diffusion models**

Acknowledgement

- We thank the sPHENIX collaboration for access to the simulated dataset, which was used in the training and validation of our algorithm. We also thank for valuable interaction with the sPHENIX collaboration on this work.
- The jet background subtraction work was supported by Brookhaven Science Associates, LLC under Contract No. 447 DE-SC0012704 with the U.S. Department of Energy.