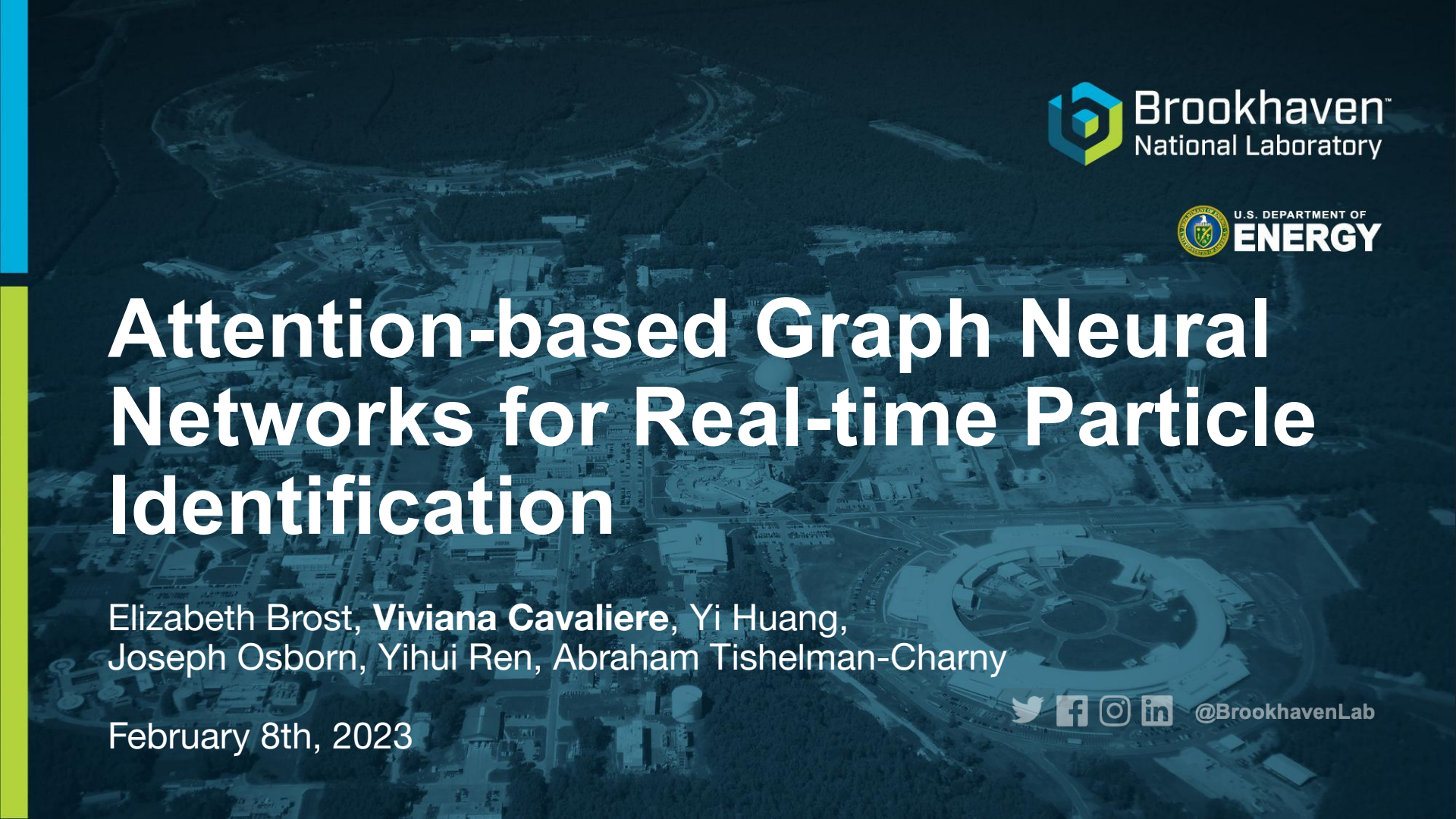


Attention-based Graph Neural Networks for Real-time Particle Identification

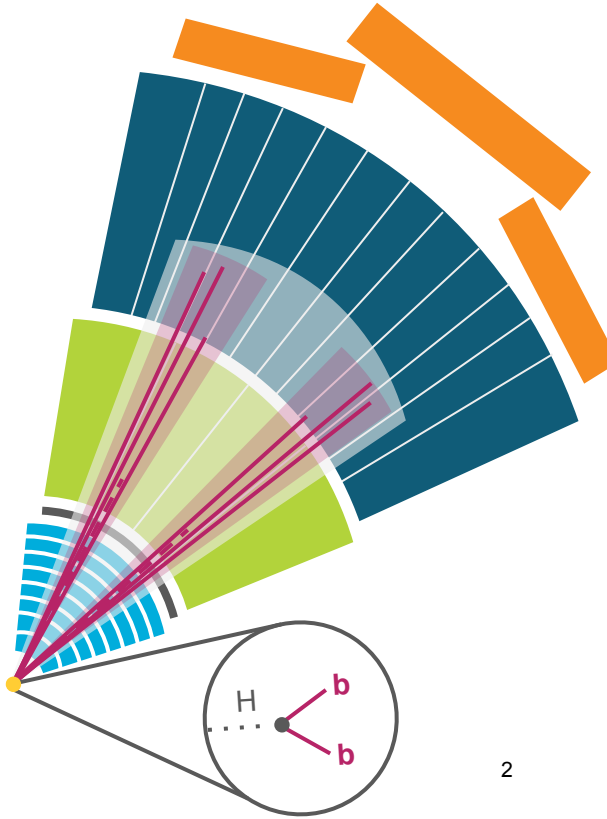
Elizabeth Brost, **Viviana Cavaliere**, Yi Huang,
Joseph Osborn, Yihui Ren, Abraham Tishelman-Charny

February 8th, 2023

@BrookhavenLab

FY2024 NPP LDRD Type B Pre-Proposal

- **Proposal Title:** Attention-based Graph Neural Networks for Real-time Particle Identification
- **Primary Investigator:** Viviana Cavaliere (PO/NPP)
- **Other Investigators:** Elizabeth Brost (PO/NPP), Yi Huang* (CSI), Joseph Osborn* (PO/NPP), Yihui Ren* (CSI), Abraham Tishelman-Charny (PO/NPP)
- **Indicate if this is a cross-directorate proposal:** Yes (NPP 🤝 CSI)
- **Proposal Term:** October 2023 - September 2025



Brookhaven National Laboratory

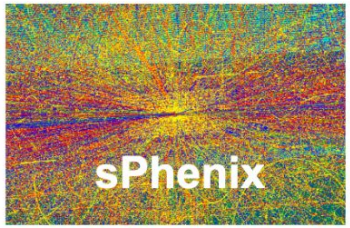
*ECA eligible

Attention-based Graph Neural Networks for Real-time Particle Identification

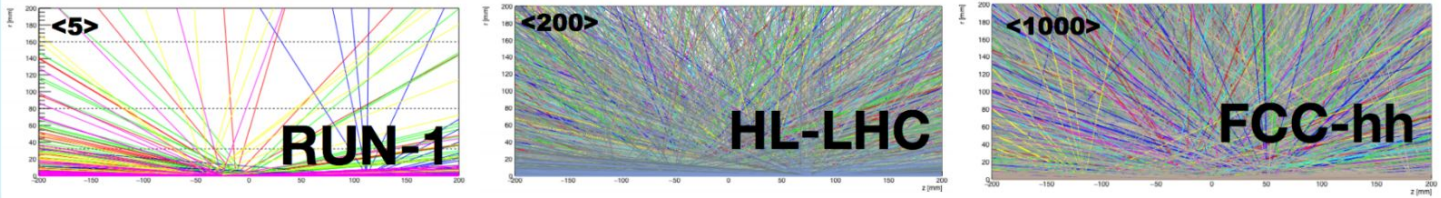
- **DOE Program:** HEP
- **Return on Investment:**
  - Brost, Cavaliere, Huang, and Ren, with university collaborators, prepared a proposal for DOE HEP “FY2022 Artificial Intelligence Research for High Energy Physics” (DE-FOA-0002705) in May 2022 (total proposed budget $2,405k / 3 years)
    - This LDRD-B project is a stepping stone to our next proposal for similar FOAs
  - This project aligns with topics in planned/future DOE Early Career Award applications
- **Broader impact on the activities at the laboratory:**
  - Optimization of ML models for use in hardware will shed light on future applications at CFN, CryoEM, NSLS-II, RHIC and other key user facilities at BNL
- **Total planned funding per year in FY24 and FY25:** $200k

The Problem We Aim to Solve

- **Problem:** real-time (microseconds) particle identification is needed to decide which events to keep for analysis
- Particle identification in dense environments (HL-LHC, sPHENIX, etc) is very hard



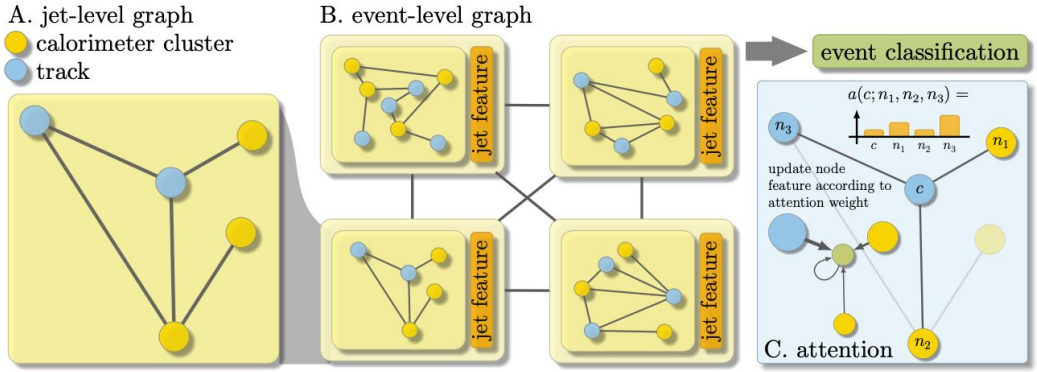
Au+Au collisions at RHIC



Increasing pileup at proton-proton colliders: <μ> = 5 (LHC Run 1), <μ> = 200 (HL-LHC), <μ> = 1000 (FCC-hh)

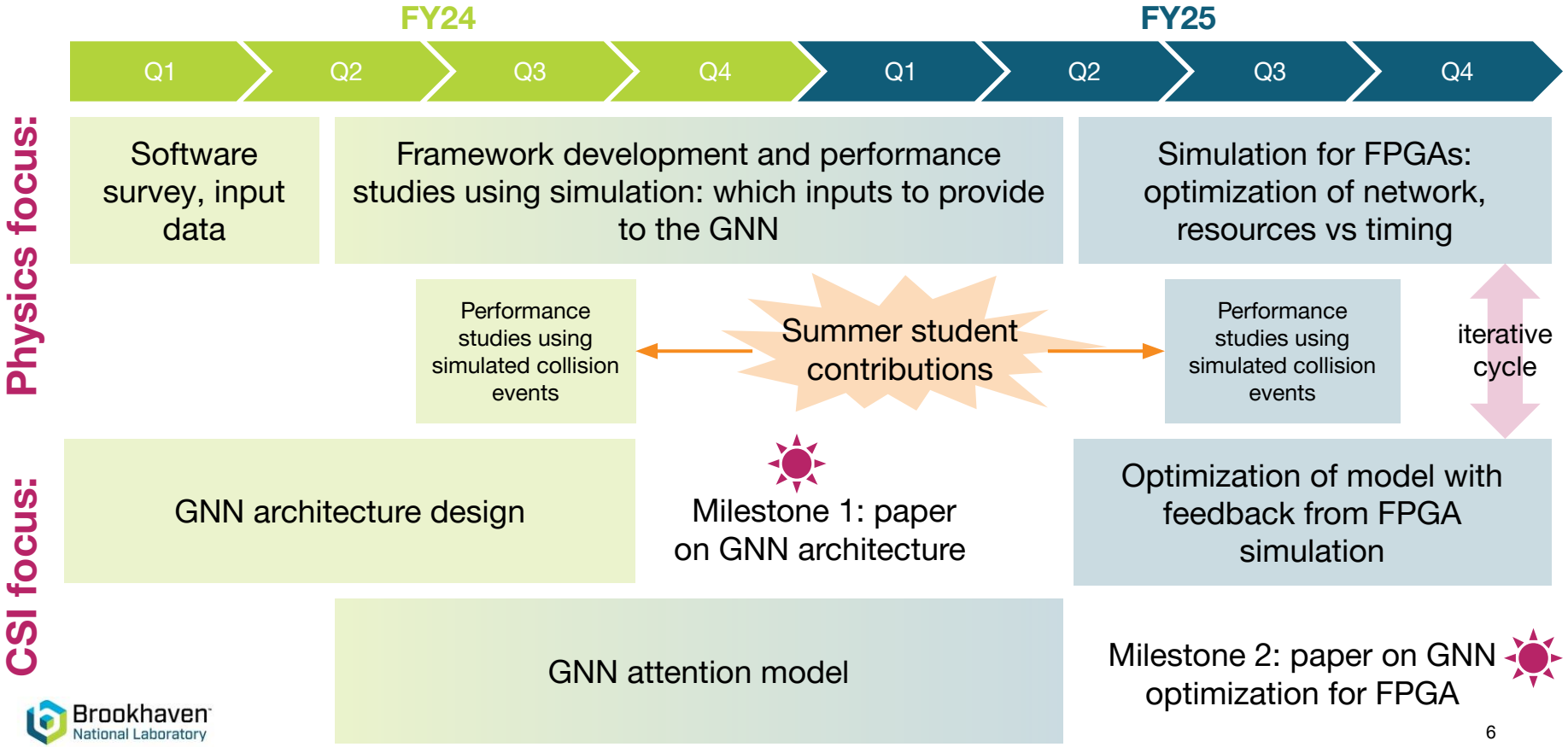
Attention-based Graph Neural Networks for Real-time Particle Identification

- We will design a Graph Neural Network that uses low-level detector information to identify target events (Higgs → bb at ATLAS, b and c decays from the hard scattering at sPHENIX) in real time
  - Expertise in CSI (see recent publications, e.g. https://arxiv.org/pdf/2102.09844.pdf)
- We will optimize the GNN for deployment in trigger hardware (FPGA)
  - This field is just developing and the success of this project will put BNL on the map for GNN deployment on FPGAs for other applications
- We will recruit undergraduate students from California State University (an MSI; we have ongoing collaborations) to come to BNL for the summer and contribute to performance studies for this project



Brookhaven National Laboratory

Research Plan & Milestones

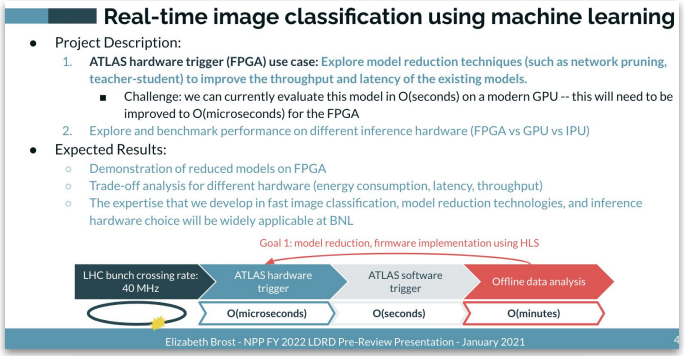


Brookhaven National Laboratory

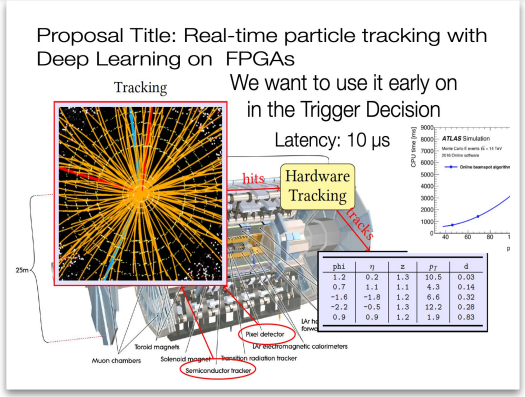
Summary

- **Project**: Real-time identification of target hadronic decays using GNNs
- **Physics outcomes**: first attention-based GNN optimized for FPGAs
- **Return on Investment**: many applications, stepping stone for other funding lines
- **Personnel:**
  - **Physics:** three staff from Omega and NPPS (Brost, Cavaliere, Osborn), one postdoc @10-20% (Tishelman-Charny), 2 undergraduates (CSU students, 8-10 weeks per summer)
  - **CSI:** one staff (Ren), one postdoc @30-50% (Huang)
- **Budget:** $200k/year in FY24 and FY25

Builds on previous successful NPP/CSI collaboration:



LDRD FY22, E. Brost, Y. Ren



LDRD FY18, V. Cavaliere

Brookhaven National Laboratory

Backup

FY24 LDRD-B Schedule:

✅ February 6 - Inform NPP about your intent to submit LDRD B proposal. Send E-mail to Fran Capasso with the proposal title, list of PI(s) and a brief abstract.

February 8 and February 10 - Meetings to pre-review proposals based on completed templates (template attached).

February 15 - Feedback to PIs about development of full proposals.

February 28 and March 1 - Full proposal presentations by PIs. Template of presentations to be provided.

March 3 - NPP decision on which proposals to be submitted to the laboratory.

March 8 - All LDRD PIQ proposals, complete with BOM and Department Chair signatures to be uploaded to a designated Indico web page, information to follow.

March 9 - Final review and signature by ALD.

March 10 – Fran Capasso submits NPP proposals to the laboratory.

Attention-based Graph Neural Networks for Real-time Particle Identification

Abstract:

One of the largest challenges at particle colliders is the real-time selection of interesting collision events for later analysis, since events that do not have a positive trigger decision are lost forever. Field Programmable Gate Arrays (FPGAs) are the preferred technology for these low-latency situations -- O(microseconds) at the Large Hadron Collider (LHC) -- as current GPU and CPU devices are too slow. The flexibility offered by FPGAs is particularly important, as they can be updated to meet new experimental needs and can incorporate the latest technological advances. This flexibility will be even more important at the next generation of colliders, such as the High Luminosity LHC (HL-LHC), Electron Ion Collider (EIC), and the proposed Future Circular Collider (FCC), which will operate with much higher data rates, posing additional challenges for their trigger systems.

To improve the real-time selection of interesting events, we will develop a machine learning algorithm for particle identification, leveraging the geometric awareness of Graph Neural Networks (GNN). We will start with the standard message-passing approach for network optimization, and then proceed to explore attention mechanisms as feature aggregation functions, which help when some parts of the input data are more discriminative than others, as in particle identification. The GNN will be optimized to fit into an FPGA for the example use-case of hadronically decaying particles at current experiments such as ATLAS and sPHENIX. The competencies that we develop in this project can be extended to other use-cases at colliders, as well as other applications of GNNs in low-latency situations at other user facilities at BNL.