Fermilab U.S. DEPARTMENT OF ENERGY | Office of Science

Intelligent experiments through real-time AI: Fast Data Processing and Autonomous Detector Control for sPHENIX and future EIC detectors

Micol Rigatti

Fast Machine Learning for Science Workshop 2022

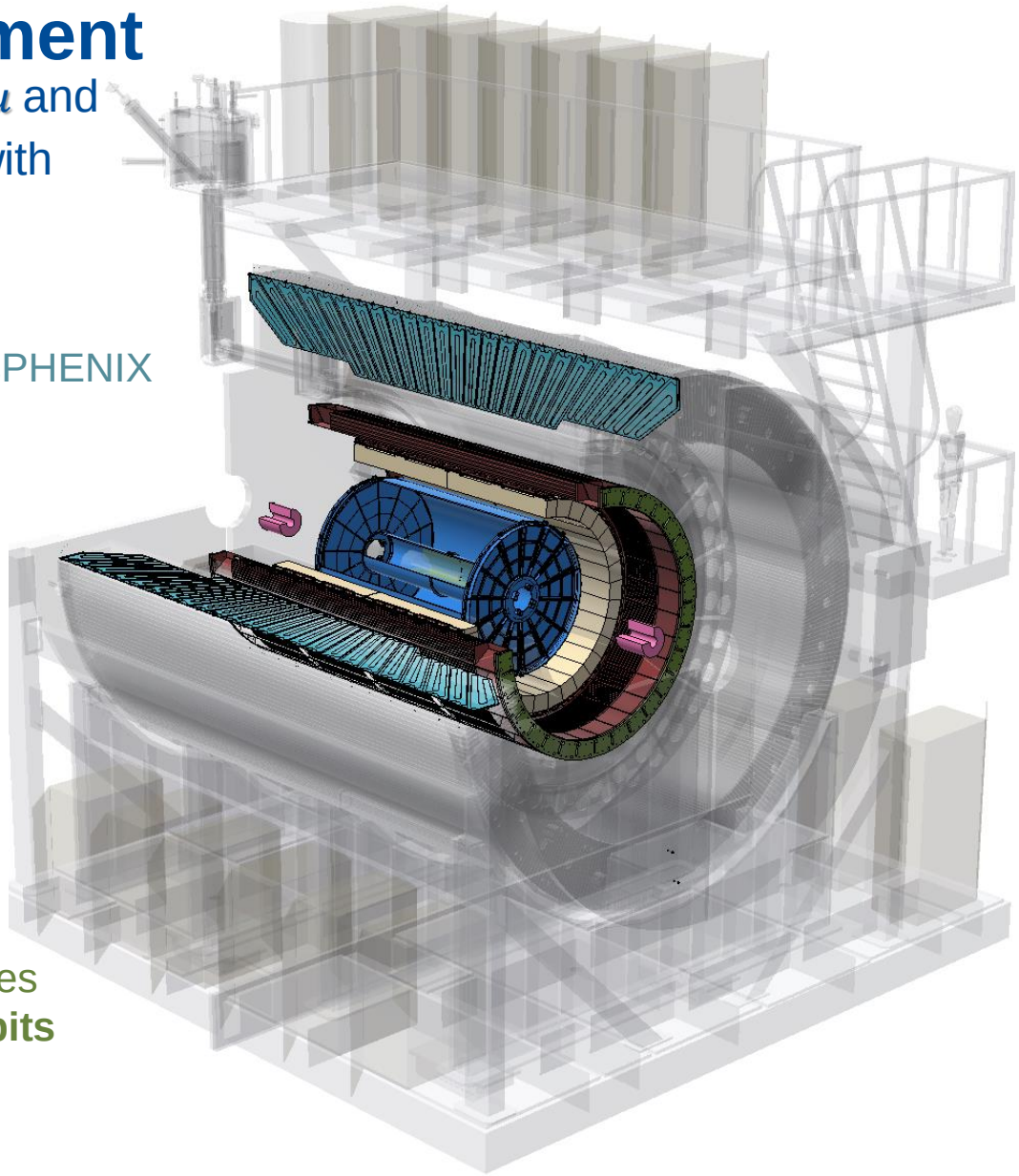
4 October 2022

The sPHENIX Experiment

at RHIC/BNL will take $p+p$, $p+Au$ and $Au+Au$ data at $\sqrt{S_{NN}} = 200\ GeV$ with **unprecedented collision rates**, approaching **10 MHz** for $p+p$.

At these collision rates, the maximum sPHENIX detector raw data rates **significantly exceed the throughput limits** of the current sPHENIX DAQ design (300 Gbit/s) and the capacity of the BNL/SDCC tape storage system and Computing resources.

Realizing the science potential of **modern nuclear physics (NP)** experiments at colliders relies on the **collection and processing of very large** datasets, with sustained data rates from future detectors **exceeding Terabits per second.**



Fermilab

We propose

to develop **real-time AI technologies** implemented **in the detector readout electronics** loop that address the challenges for the next generation of NP experiments at **RHIC** and **EIC**.

Brookhaven National Laboratory

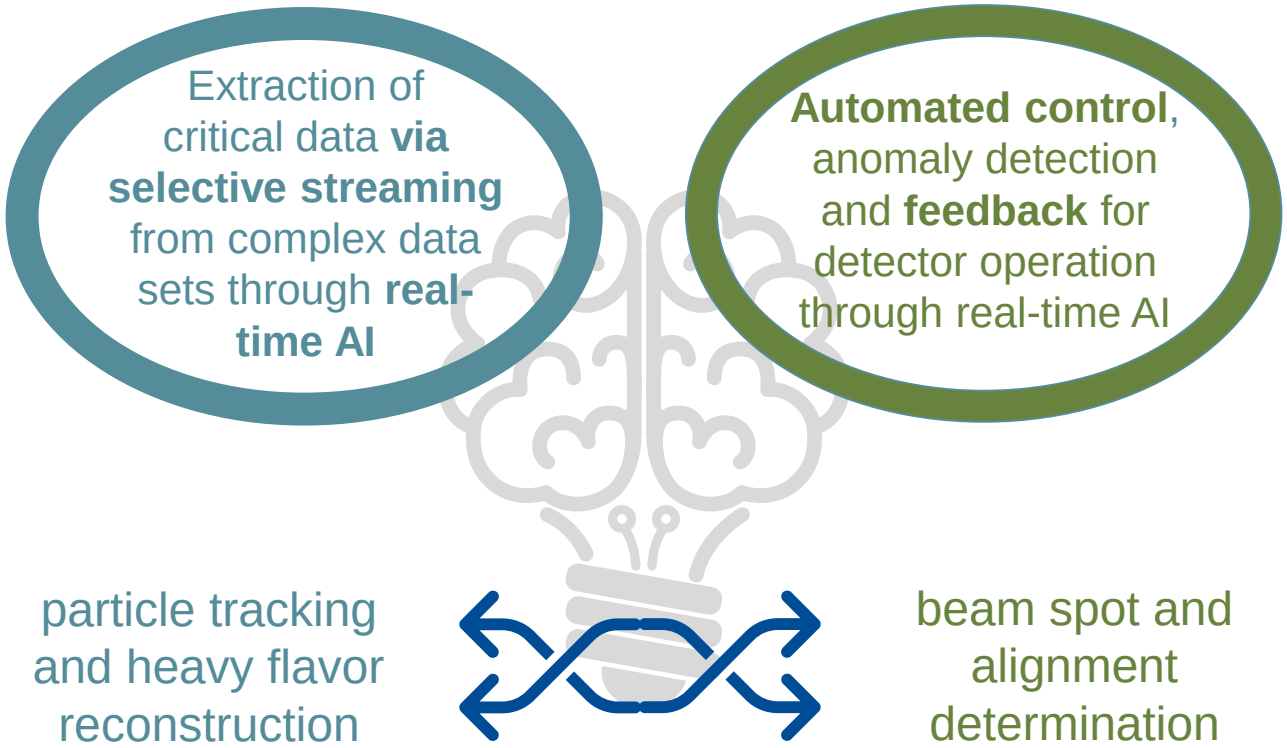
Massachusetts Institute of Technology

NJIT New Jersey Institute of Technology

Los Alamos NATIONAL LABORATORY

Technical Approach

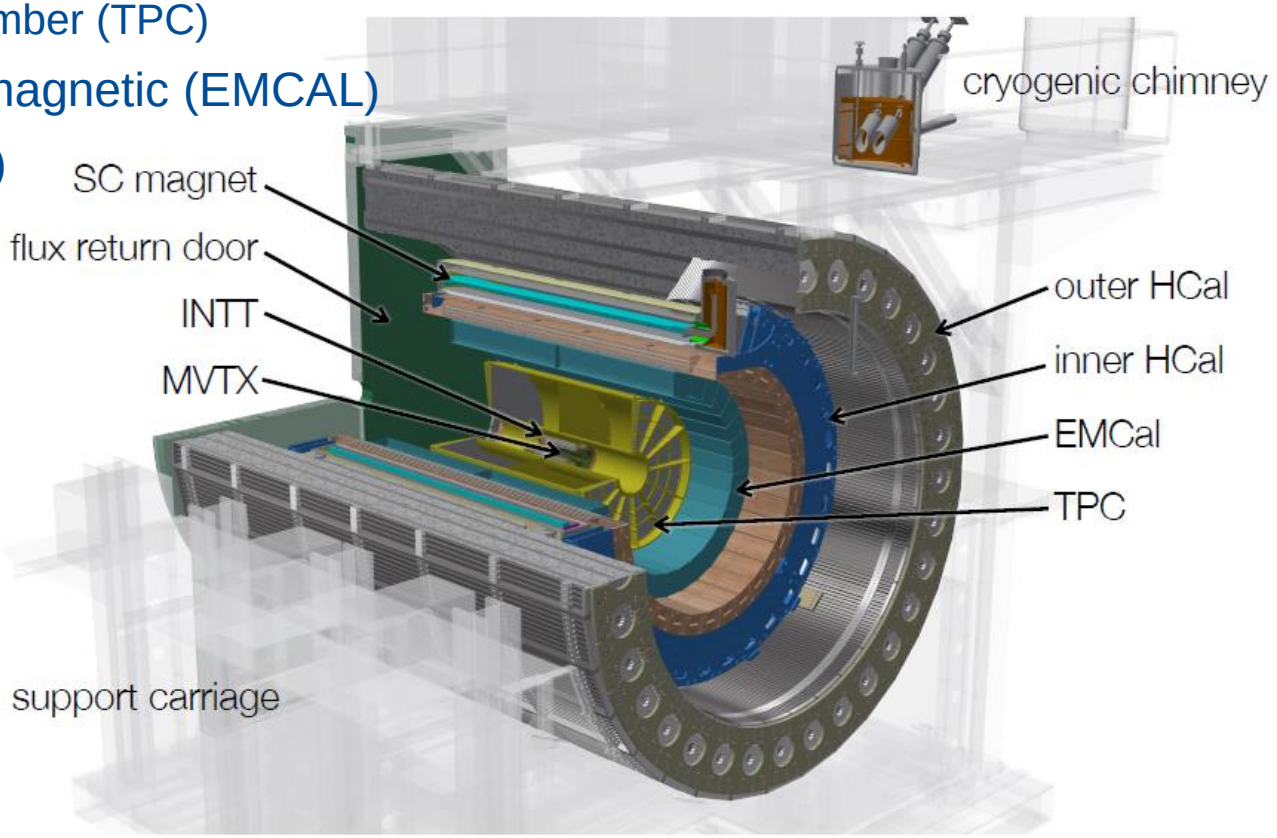
The proposal seeks to transform future flagship detectors of the US NP program into **intelligent experiments** and **smart data acquisition and control** through the integration of next-generation **AI/ML hardware**, **electronics** and **algorithms** into these detector systems.



The sPHENIX application

- Starting in 2023
- Tracking system:
  - MAPS-based vertex detector (MVTX)
  - Silicon strip tracker (INNT)
  - Time projection chamber (TPC)
- Calorimeters: electromagnetic (EMCAL) and hadronic (HCAL)
- 1.4 T super conductive solenoid

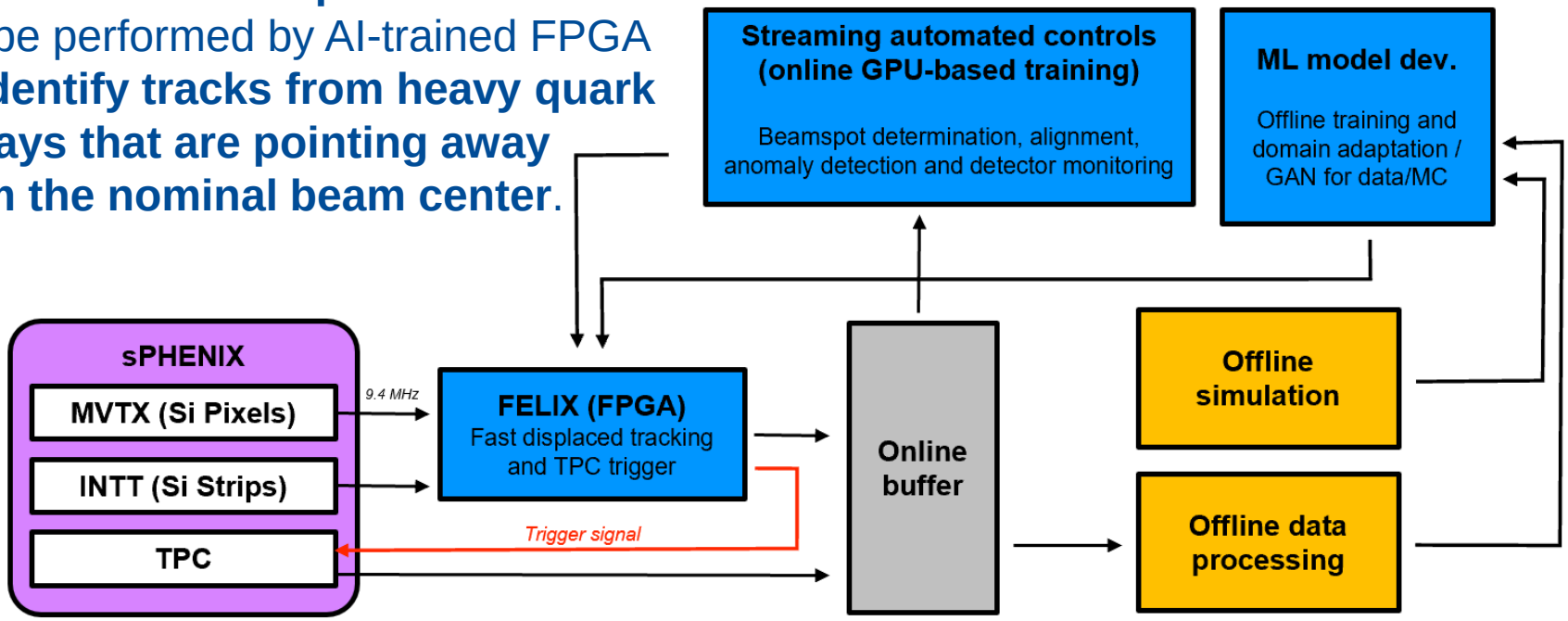
sPHENIX will study the **hot QCD medium, called Quark-Gluon Plasma (QGP)**, using scale-dependent probes to search for the existence of novel quasi-particles.



Technical Approach

The buffered data stream is **throttled** by selecting output data correlated to calorimeter triggers to **limit data volume**.

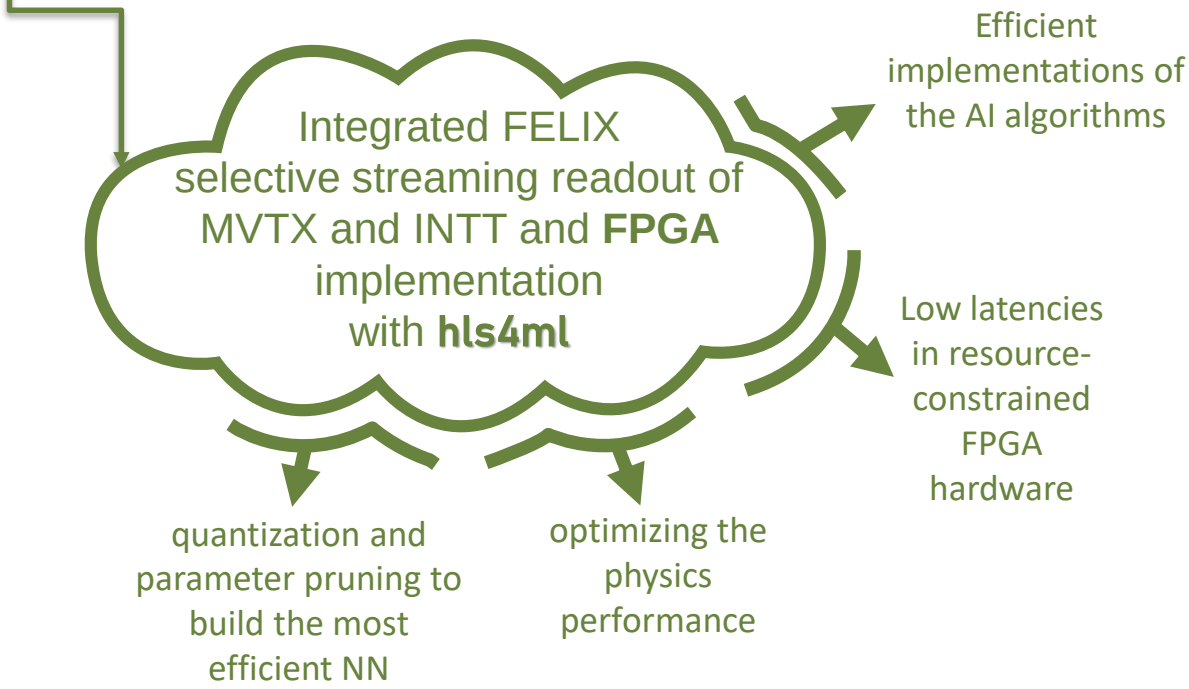
A fast search for **displaced tracks** will be performed by AI-trained FPGA **to identify tracks from heavy quark decays that are pointing away from the nominal beam center**.



A successful implementation requires the combination of several state-of-the-art AI approaches: including the **execution of ML algorithms on fast dedicated hardware** (e.g., FPGAs), with the **overall workflow itself controlled through autonomous ML processes**.

Methods

1. Physics simulation
2. AI model development
3. Integration of AI within the streaming systems
4. AI acceleration and autonomous closed loop feedback

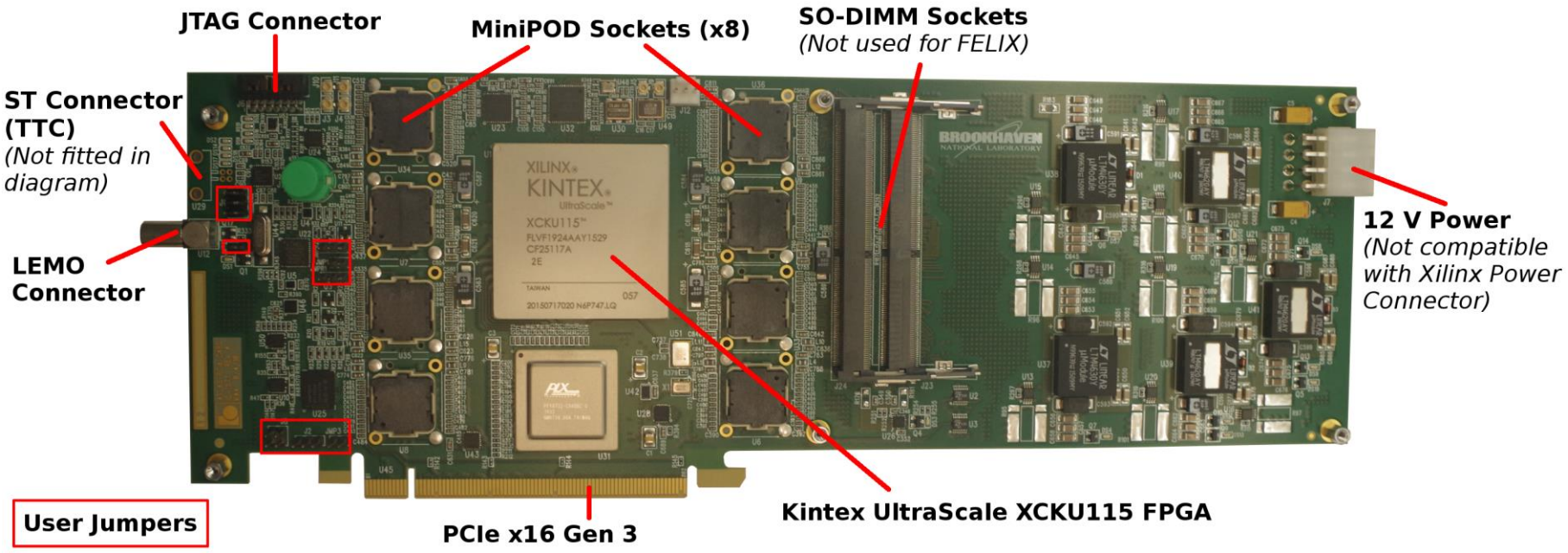


Our primary focus is on achieving low latency, real-time processing of scientific data. For algorithms to be successful, their deployment on hardware should be effective and efficient.

Fermilab

Hardware and firmware implementations

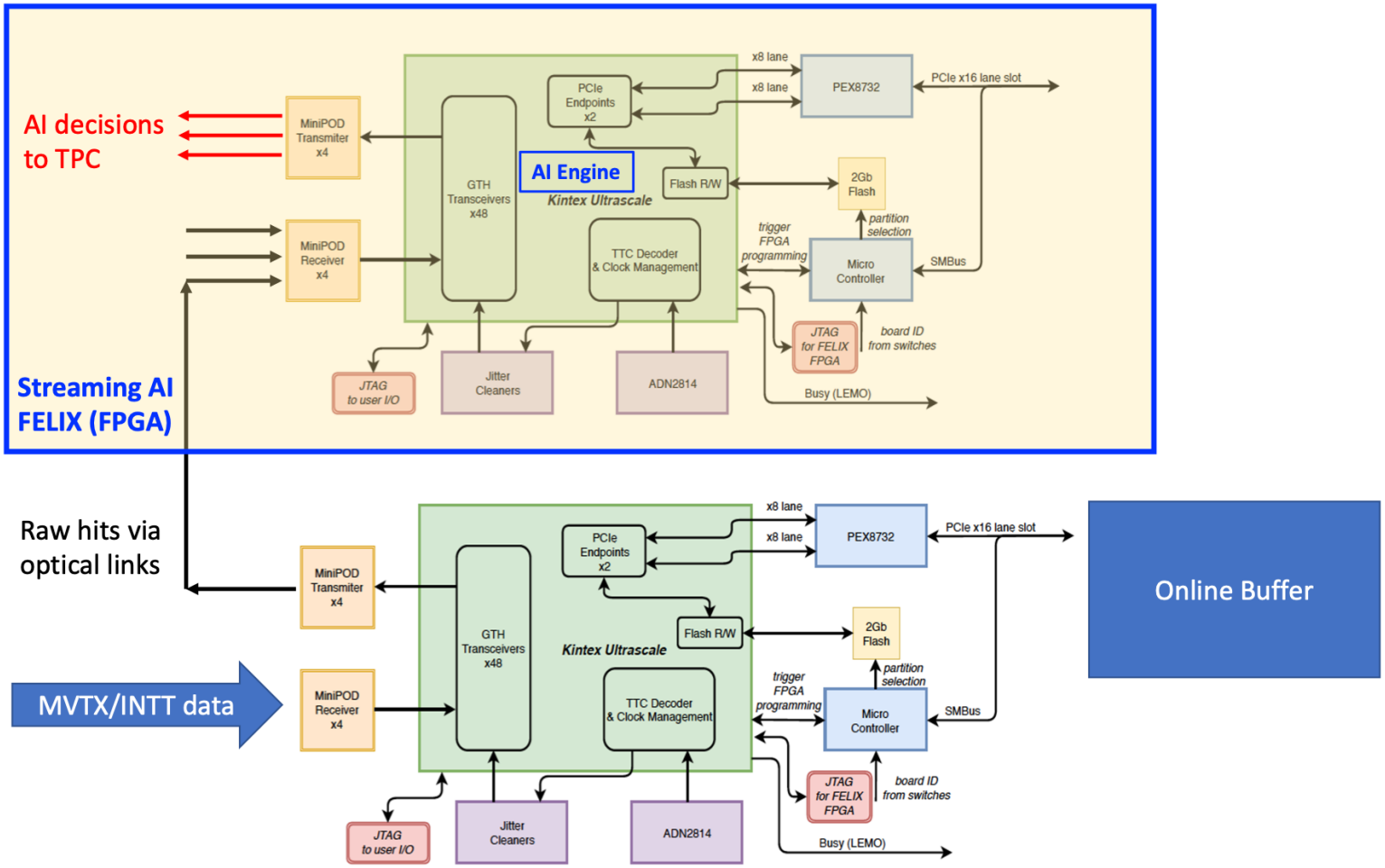
of selective data streaming in sPHENIX will use the **FELIX** board.



FELIX is a 16-lane Gen-3 **PCIe** card with 48 transmitters and receiver optical links. The on-board FPGA is a **Kintex Ultrascale** XCKU115FLVF1924-2E

Fermilab

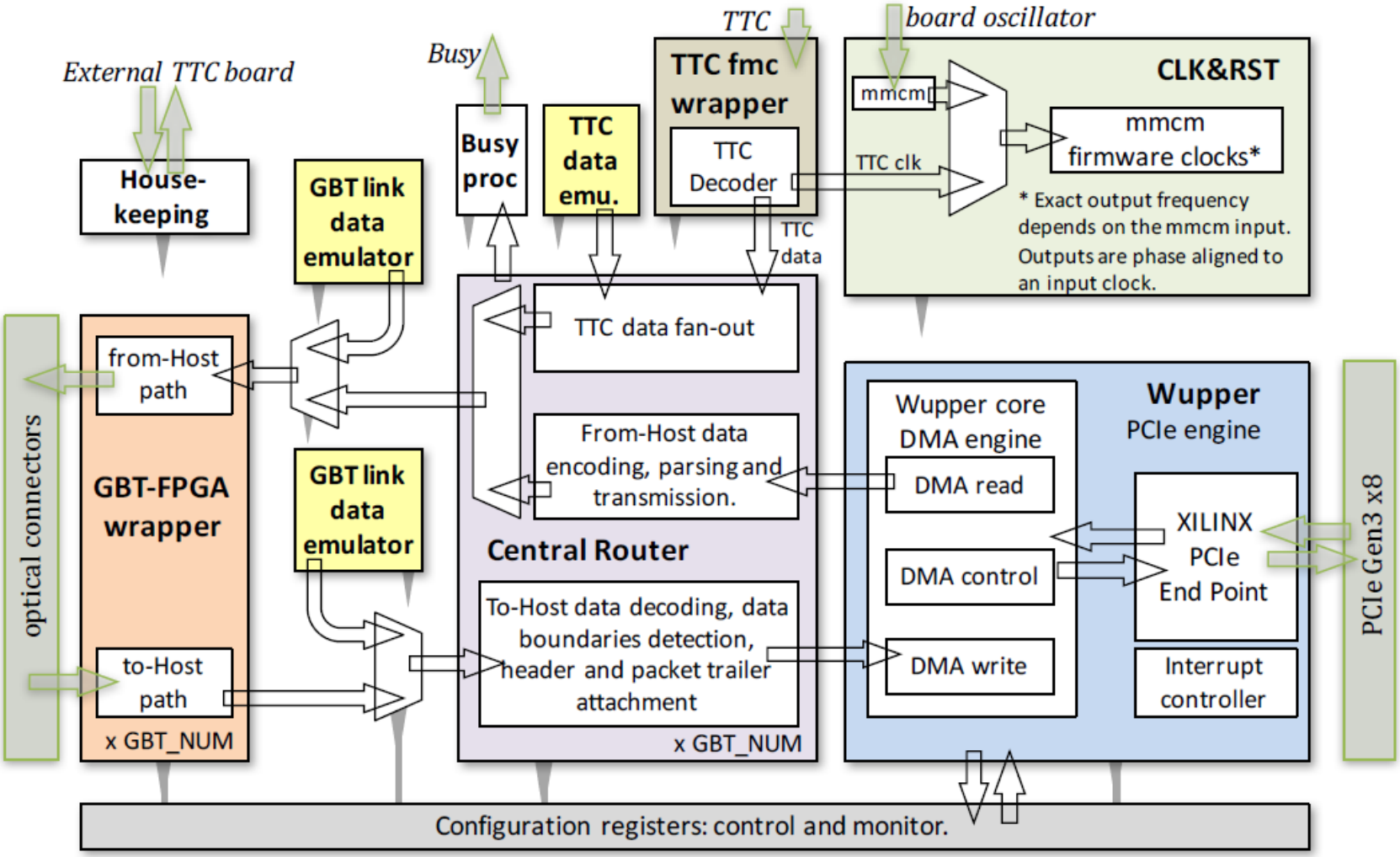
Hardware setup



A second FELIX is connected to the first FELIX through the optical transceivers, as a **dedicated FPGA hardware for smart control and real-time decision making** for TPC readout in the selective data streaming architecture.

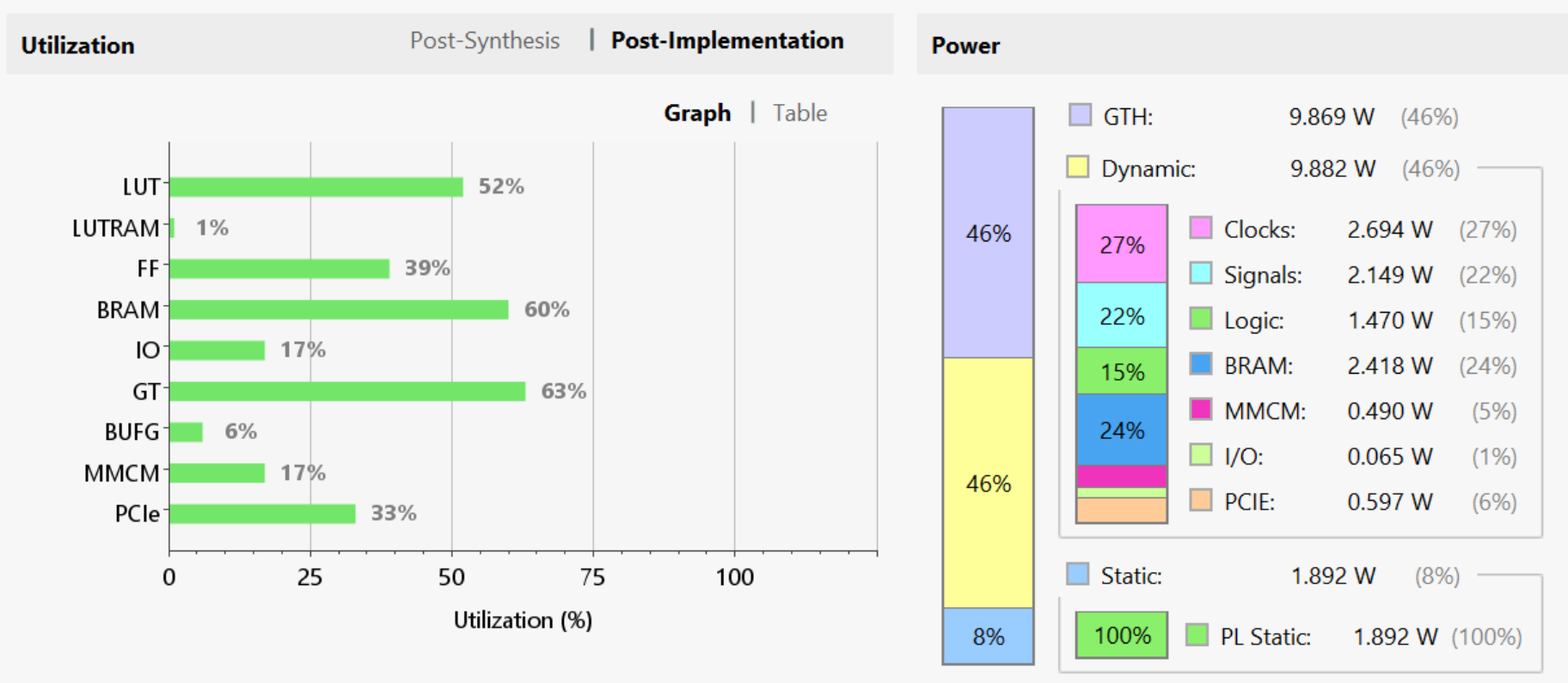
Fermilab

FELIX Firmware



Fermilab

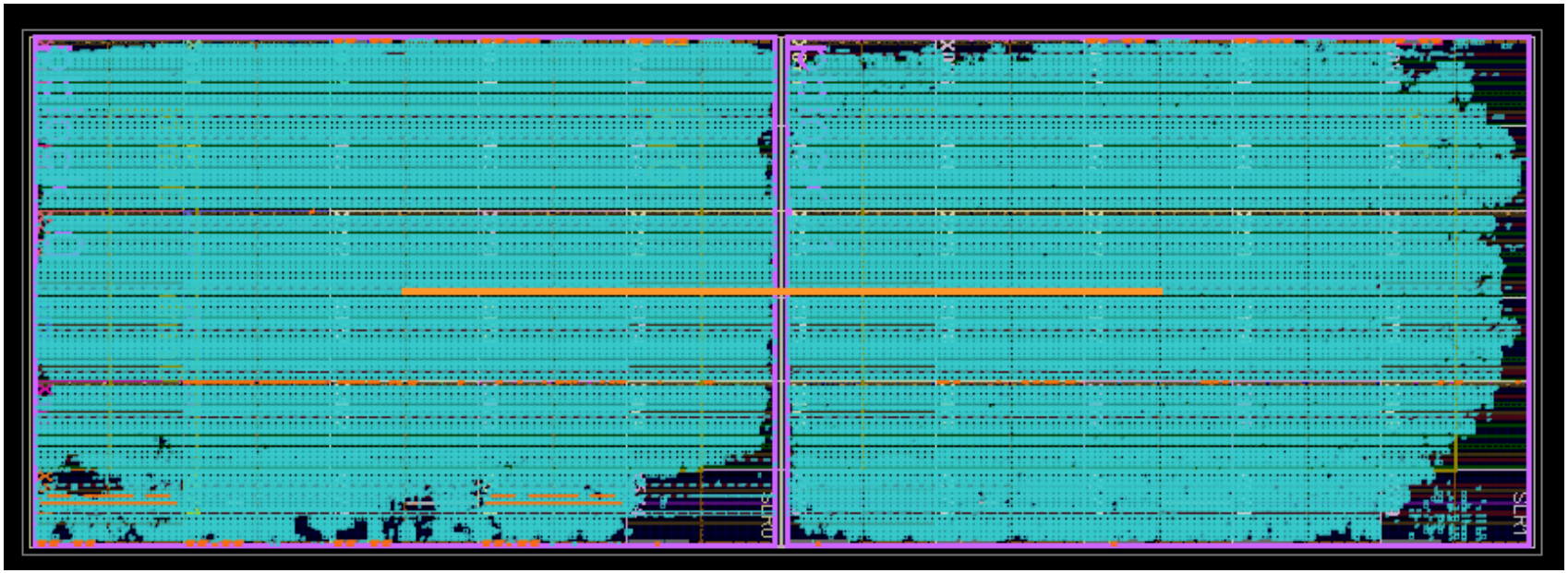
FELIX Firmware Utilization



BNL711 FELIX Firmware Report Utilization

FELIX Firmware Challenges

1 Fit a NN in the FELIX Firmware complying with timing constraints



BNL711 FELIX Firmware Floorplanning

- Fit the NN in the already crowded FELIX Firmware
- The workflow of **hls4ml** generates an IP considering available the resources of the target FPGA. We need to route the NN considering the already routed FELIX Firmware

➡ **Meeting the timing in this condition is the real challenge**

Fermilab

FELIX Firmware Challenges

2 Test the functionality of the NN and the data routing mechanism

The FELIX card is a **data router**. It needs an application to be instructed on the data movement.
The **application** relies on an **OS** and a **driver** interfacing the FELIX card.

- Inserting the NN in the FELIX Firmware will change how the driver interfaces with the card
- The testing part right now relies on the interplay of Hardware, Firmware and Software

→ **Interfacing the board to test the NN is the other real challenge**

Task 3 schedule

1. The Fermilab **test stand** is been set up: the FELIX board **BNL711** is set up in a Host Computer.

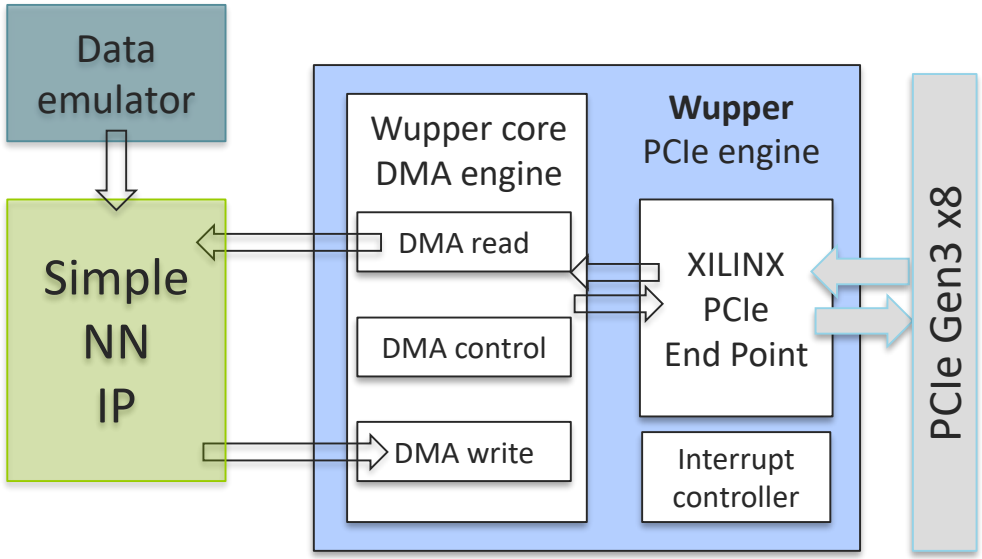
2. On the Host Computer, the **FELIX driver** and the **software application** is been installed.

3. The **data movement** between the BNL711 and the Host Computer is been tested. The data are emulated inside the BNL711.

Task 3 schedule

5 The **hsl4ml** workflow is been tested with the Kintex Ultrascale XCKU115 as a target generating the IP of a **simple NN**

6 The IP of the simple NN is in the integration phase with the **Wupper** module (**PCIe engine**)



NEXT:

7 The analysis of the FELIX Firmware in terms of resources to drive the **hls4ml** implementation of the NN

Fermilab

Thank you!