FPGA Architectures for Distributed ML Systems for Real-Time Beam Loss De-Blending

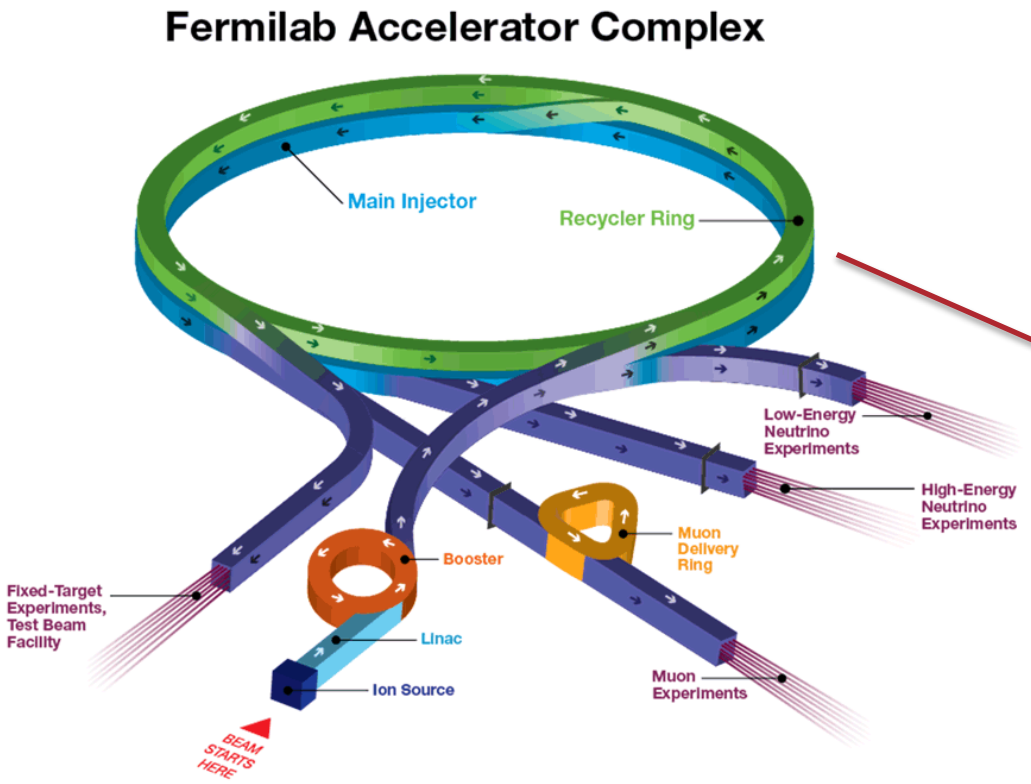
M.A. Ibrahim

This work was produced by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. Publisher acknowledges the U.S. Government license to provide public access under the DOE Public Access Plan.

Additional funding provided by Grant Award No. LAB 20-2261.

In partnership with:

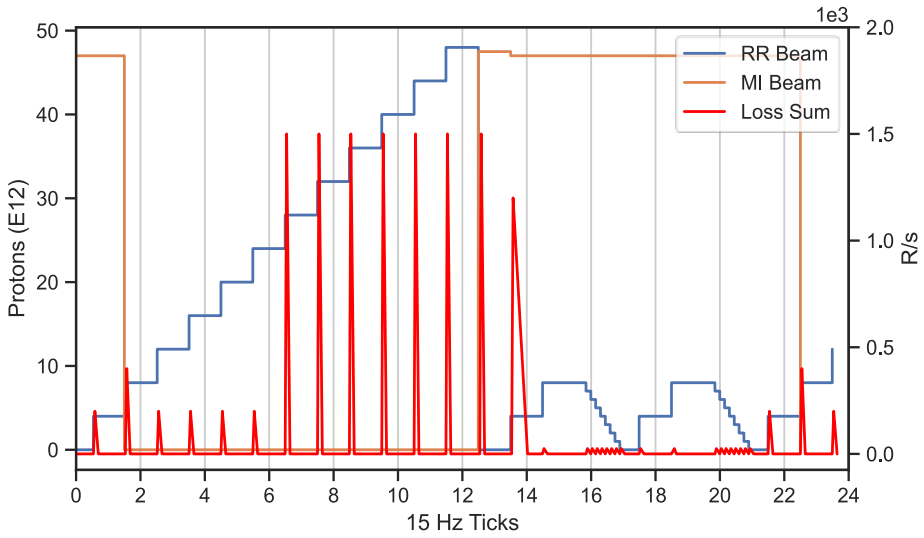
Fermilab Main Injector (MI) and Recycler (RR)



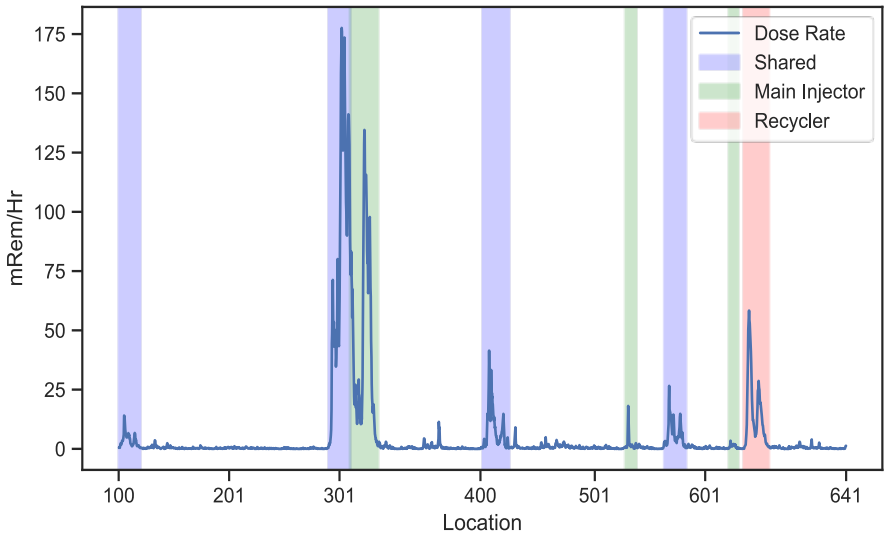
(Top) Recycler Beam Line
(Bottom) Main Injector Beam Line

- Both machines can and do often have high intensity beam in them simultaneous.
- Both machines can generate significant beam loss.
- The machine origin of a beam loss is often hard to distinguish.
- Often losses from one machine end up tripping the machine permit of the other resulting in unnecessary beam downtime.

Real-Time Edge AI for Distributed System (READS)



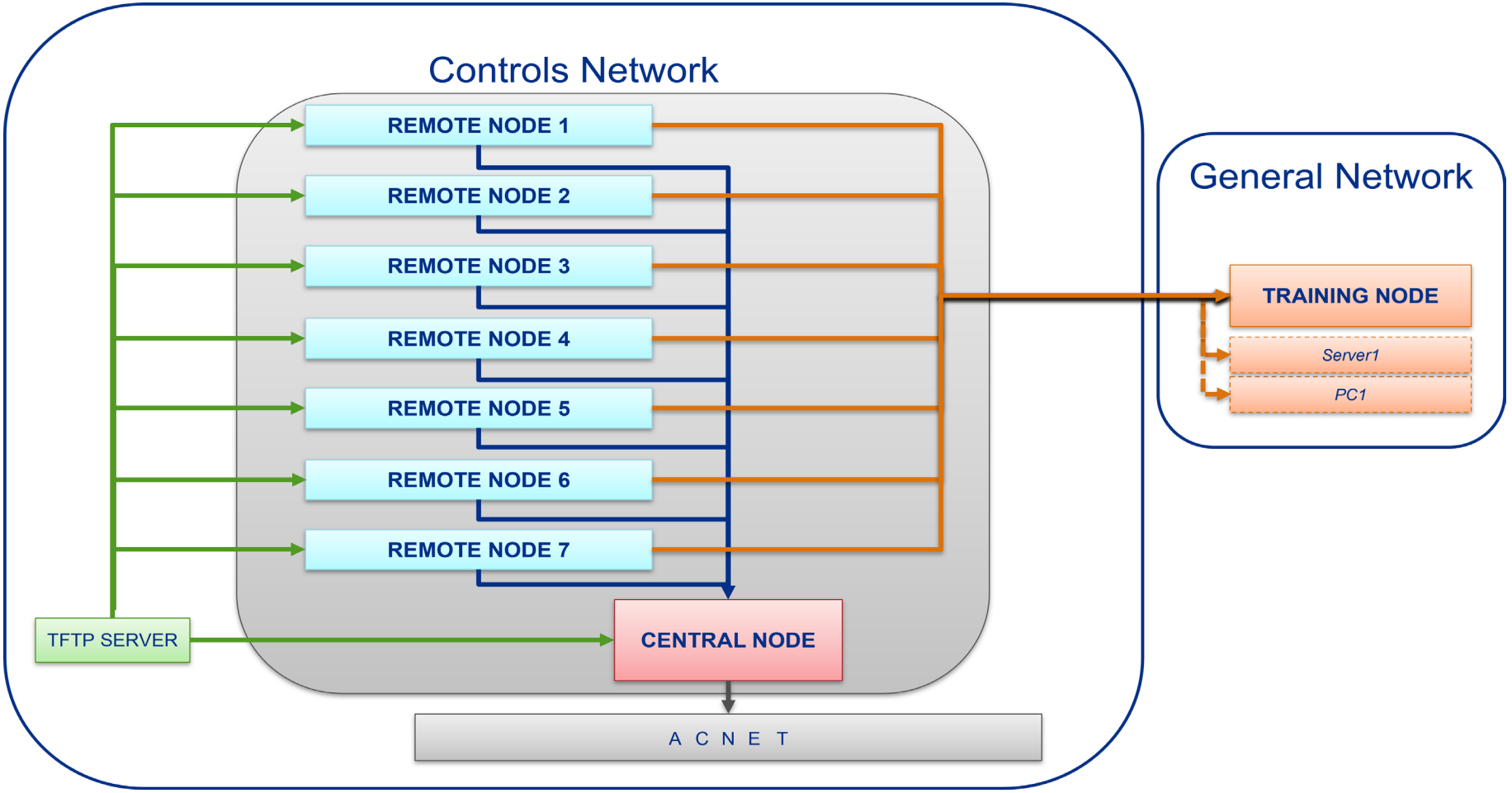
Example illustration of overlapping beam events and losses in the MI and RR accelerators



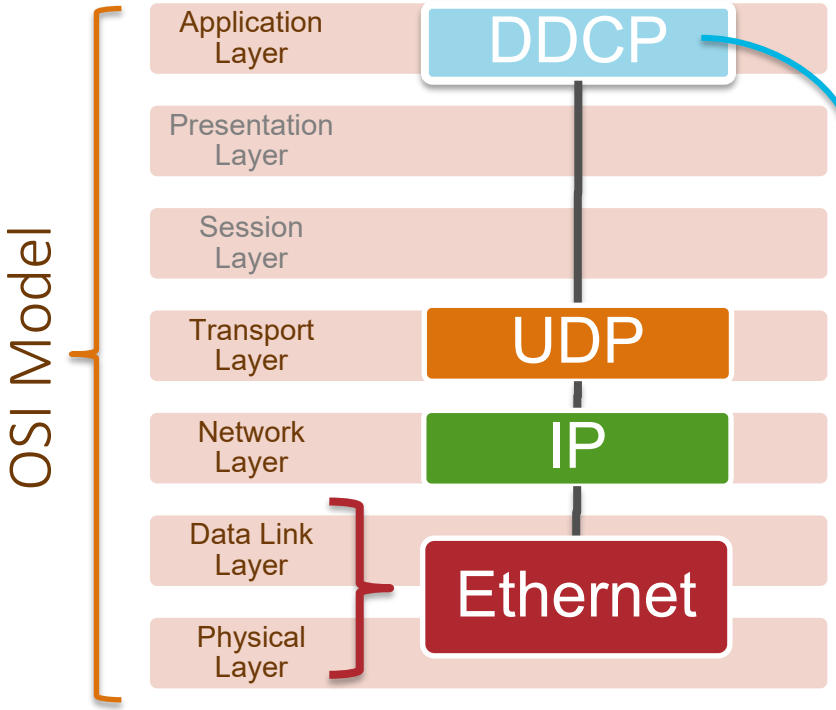
Location dependency of MI and RR beam loss as seen from tunnel activation residual doses

- The projects aims to deploy a machine learning model on a FPGA that when fed streamed beam loss readings from around the Main Injector complex, will infer in real-time the machine loss origin
- Using time, location and state of the machine, machine experts can sometimes attribute loss to a particular machine
- This suggests a Machine Learning (ML) model may be trainable to automatically attribute loss and replicate or improve upon the expert's ability

READS Distributed Network



READS DDCP Communications



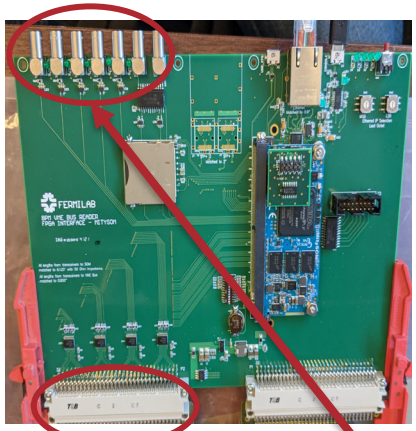
Mapping and Configuration Within Distributed Network

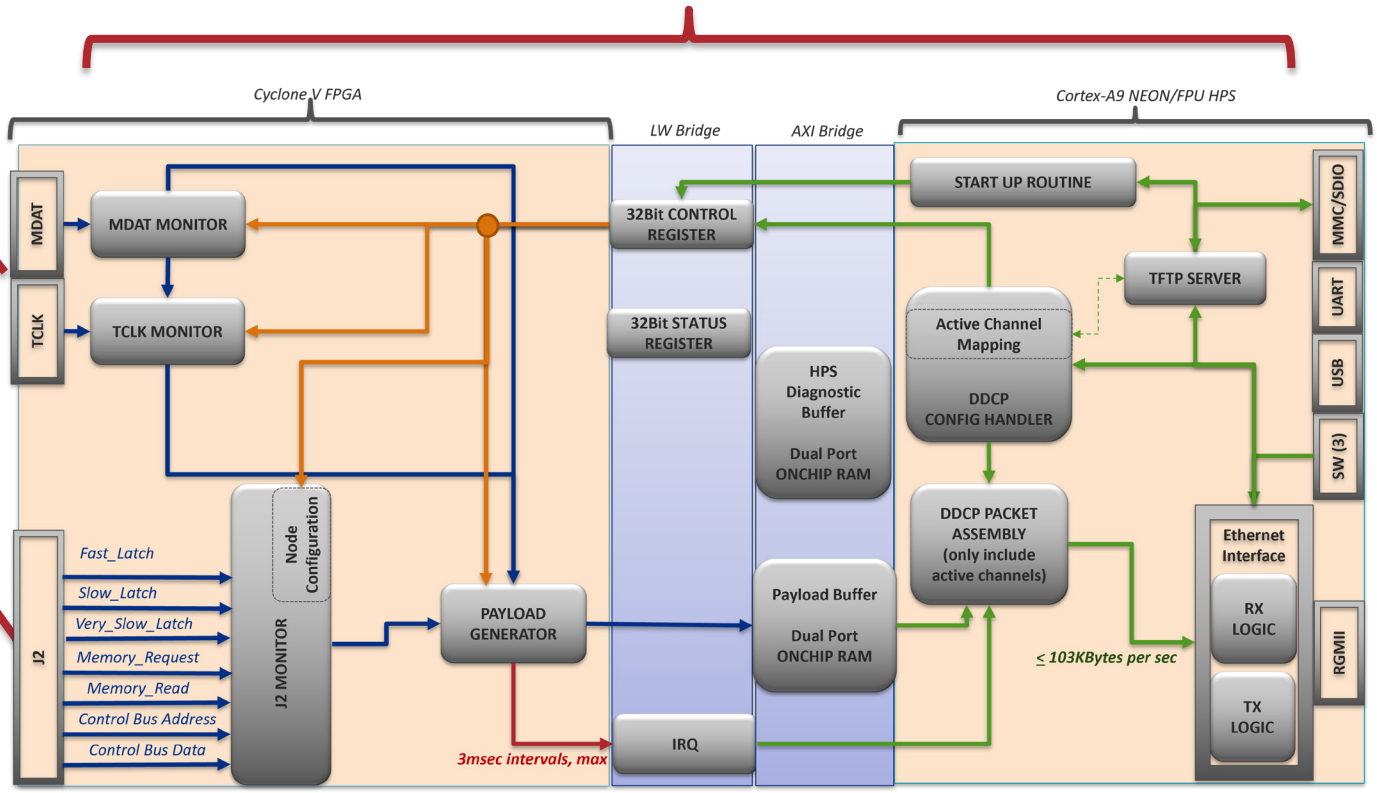
Slot	Maps to a device in the network.
Feature	Maps to a feature in the device (a control register, an array, etc.)
Operation	Operations that can be done on a feature.

Distributed Data Communications Protocol (DDCP) is a UDP application layer protocol, used to establish client-server relationships between the Remote DAQ Nodes and ML-related nodes:

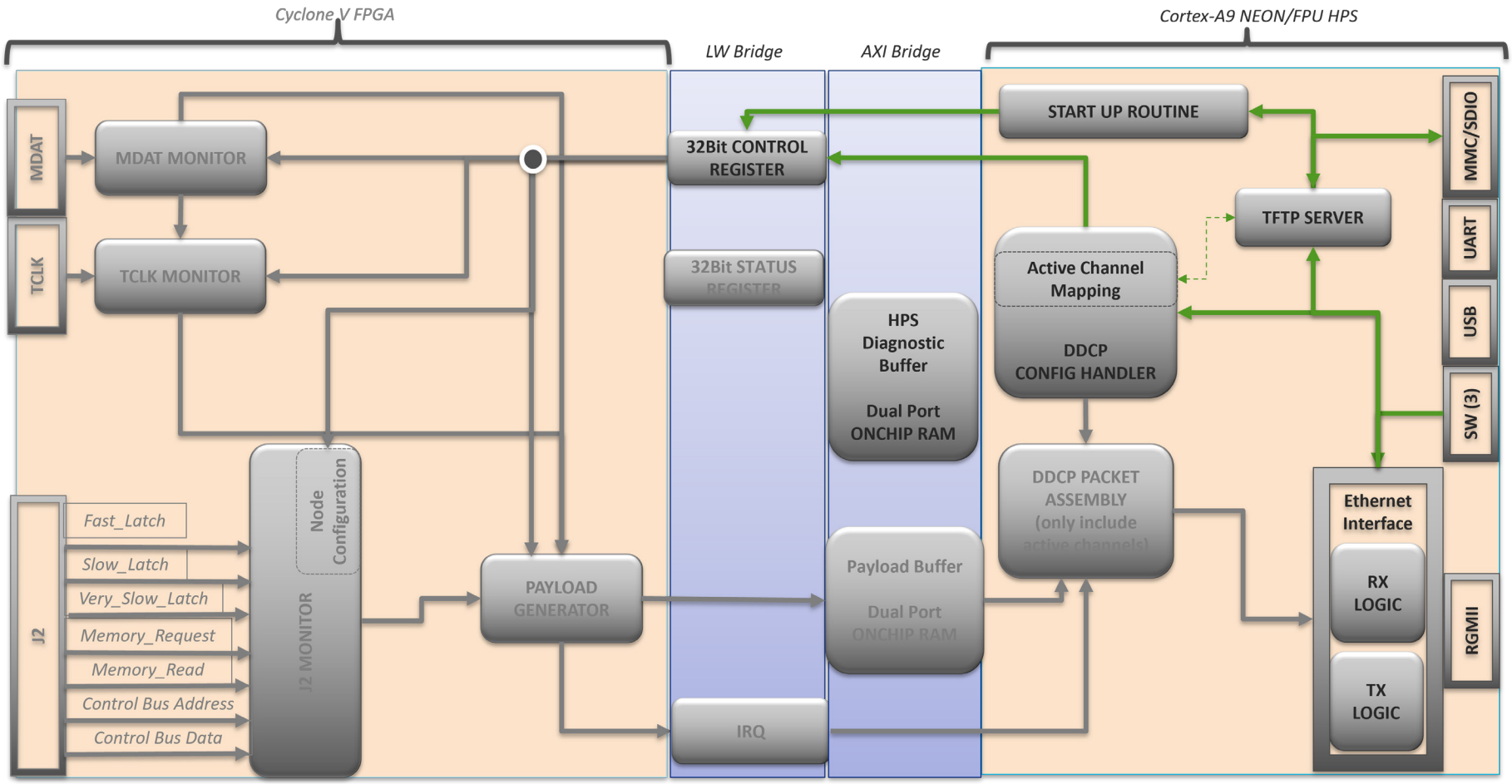
- Stream data between the captured data to the Central Node.
- Control hardware registers and registers in the software state machine in the VME Reader Cards.

READS Distributed Remote Nodes

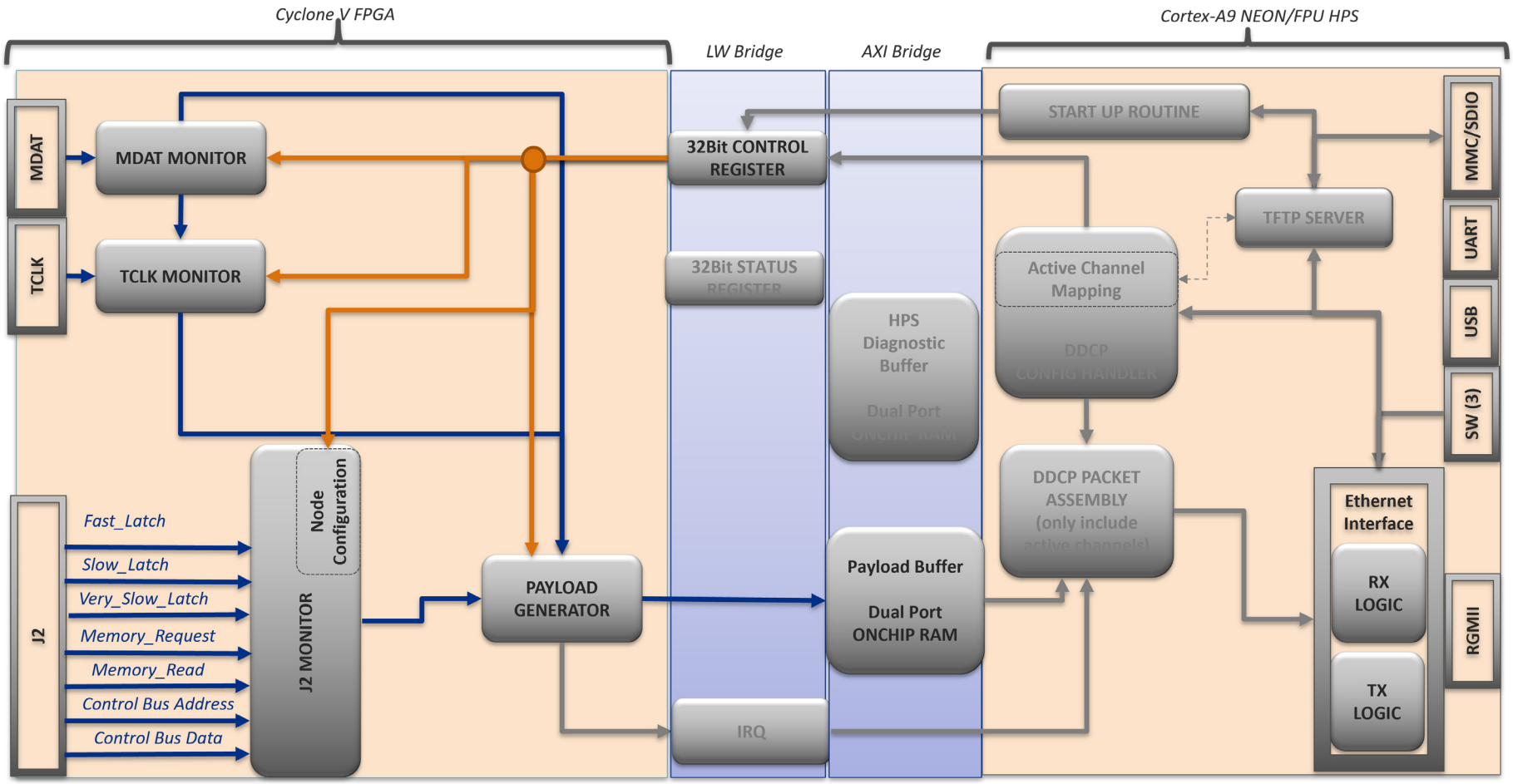




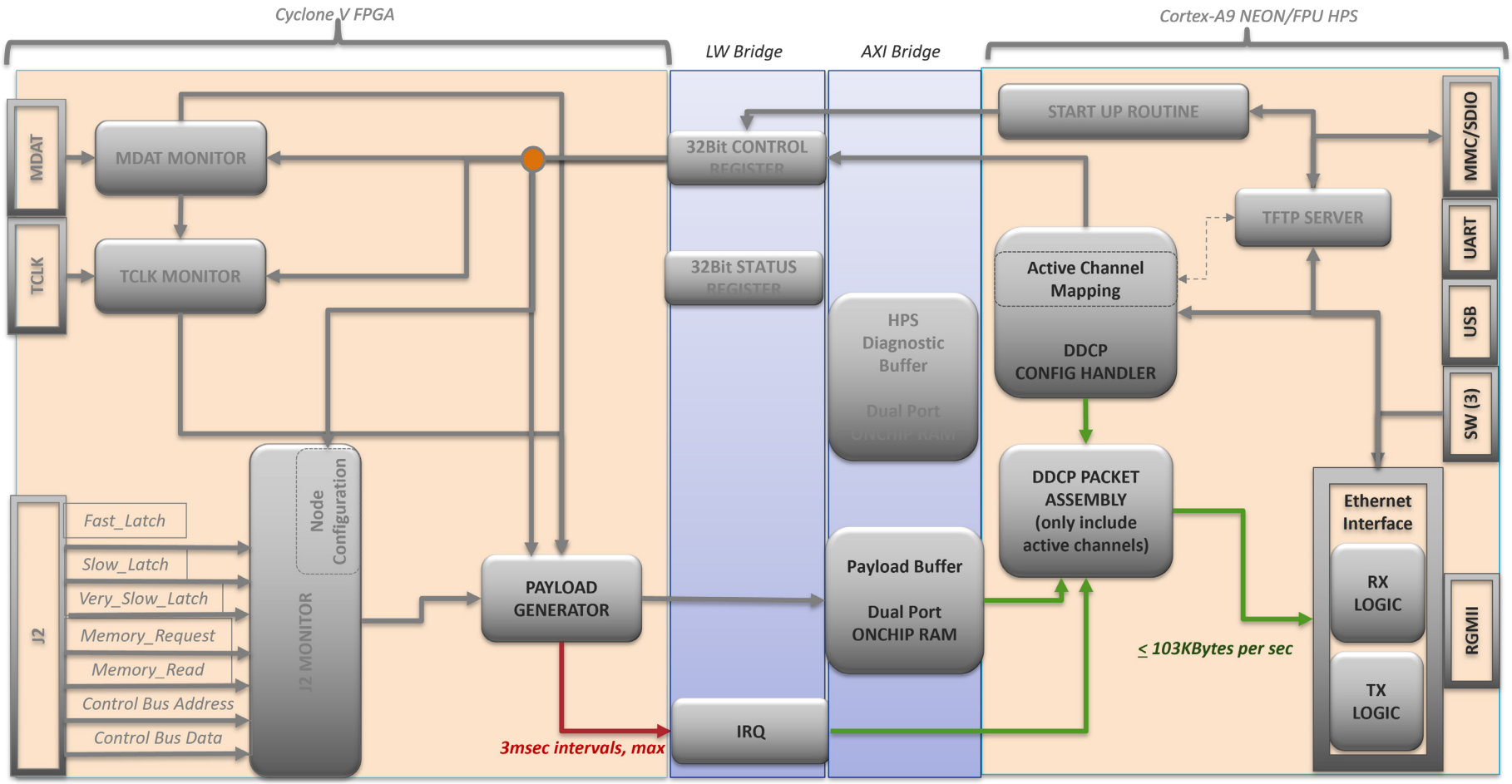
Data Path Thru Remote Node (1/3)



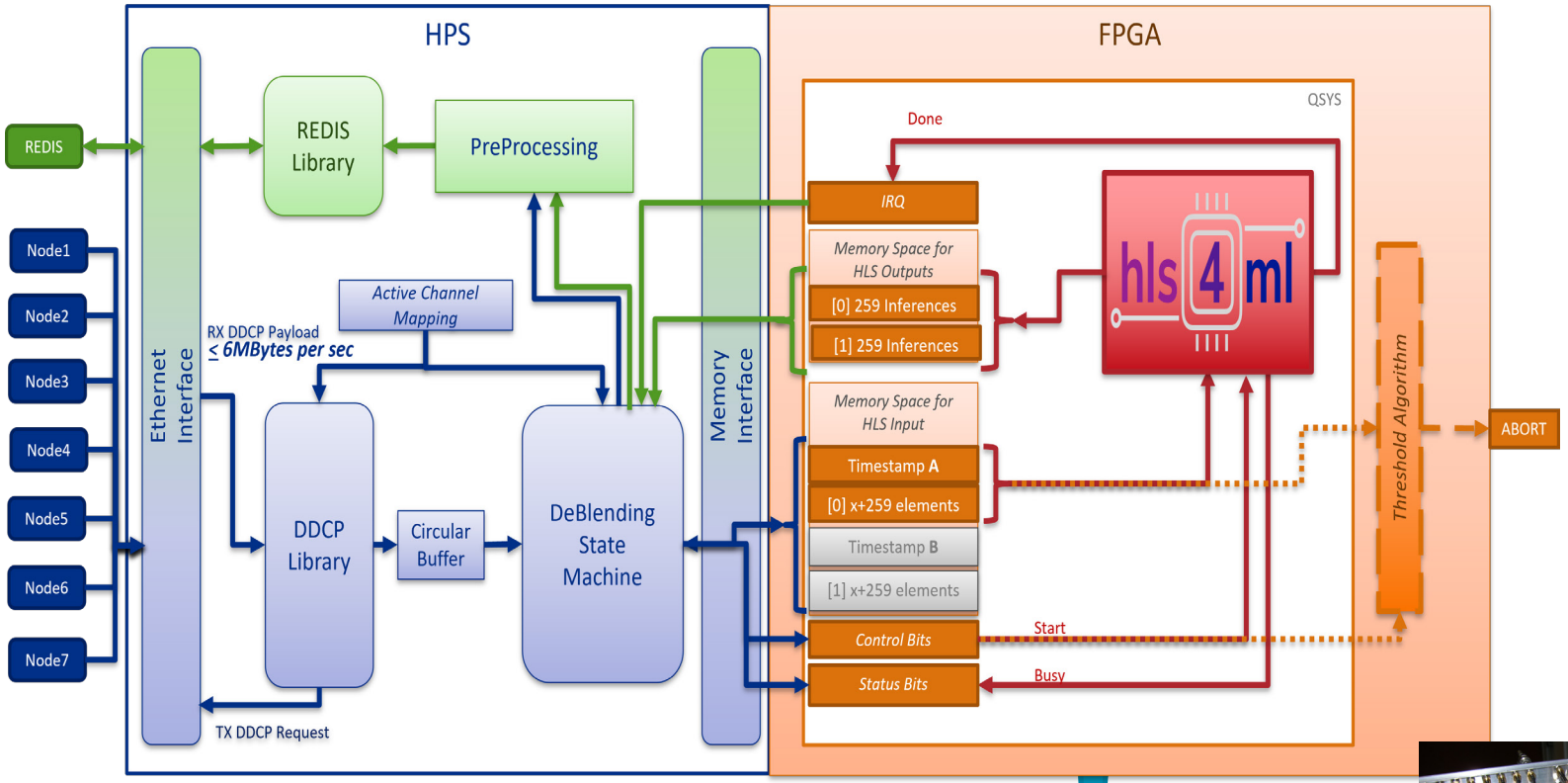
Data Path Thru Remote Node (2/3)

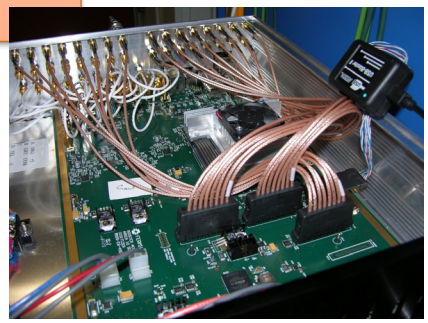


Data Path Thru Remote Node (3/3)

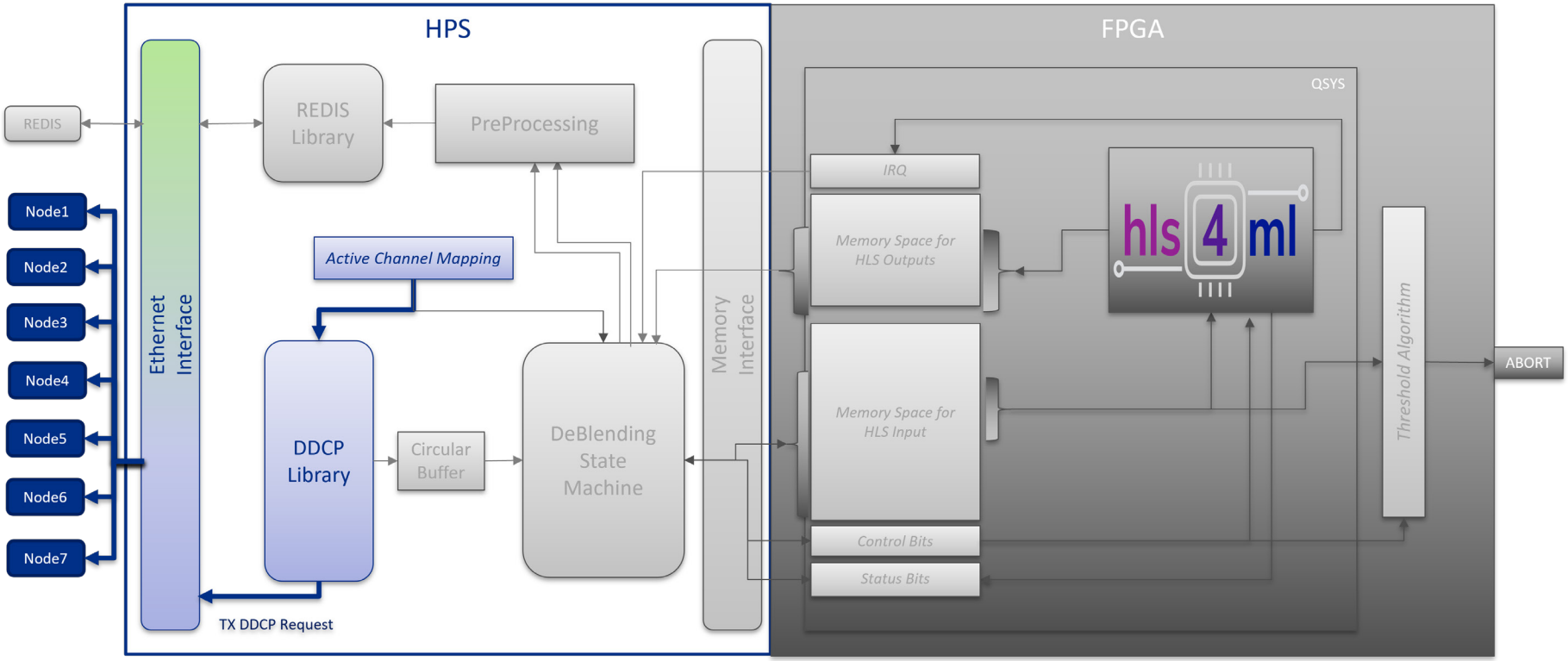


READS Central Deblending Node

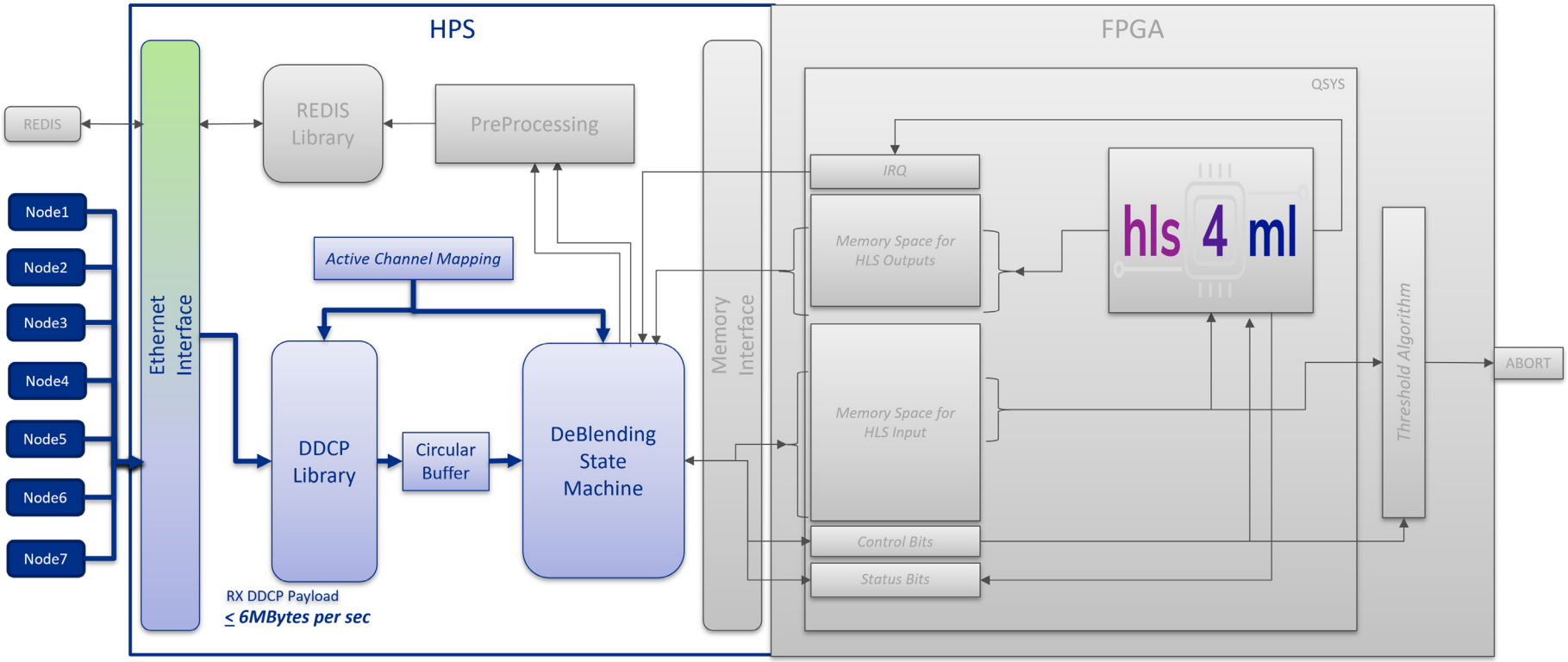




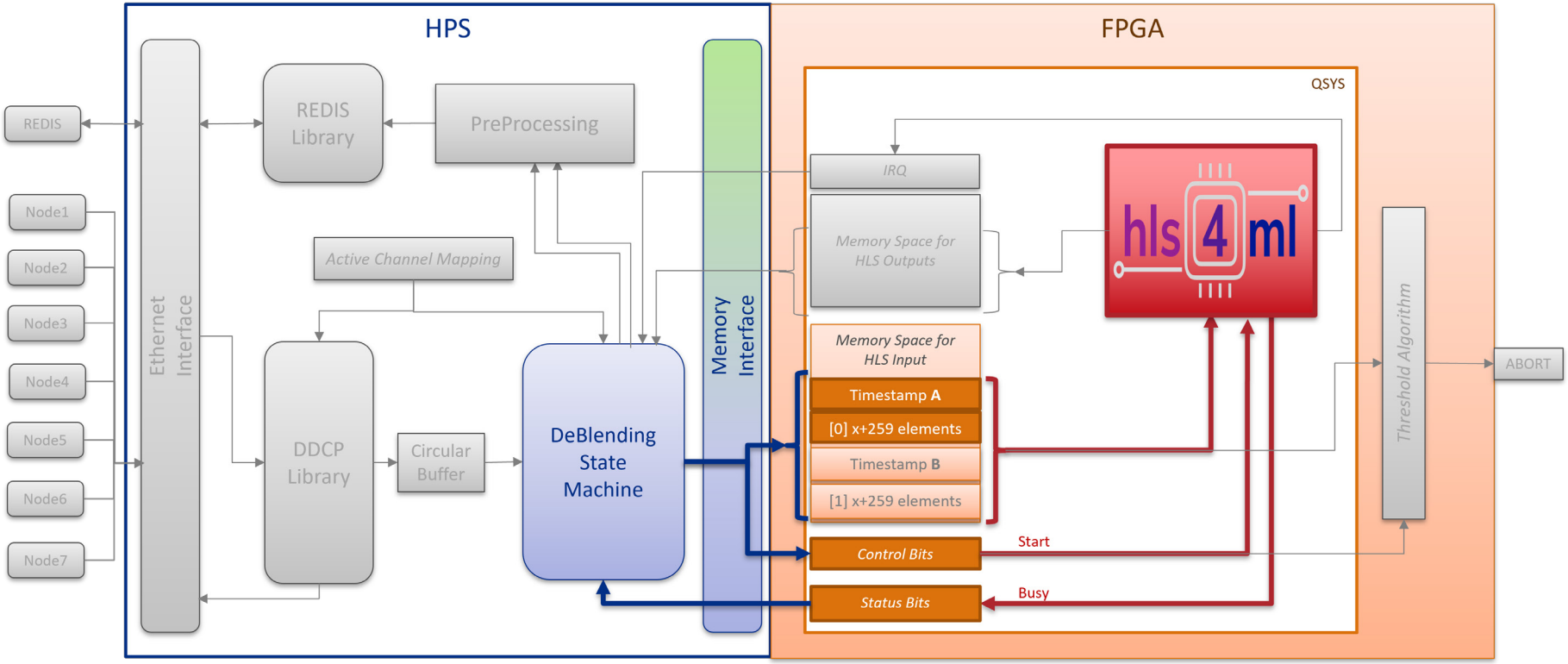
Data Path Thru Central Node (1/5)



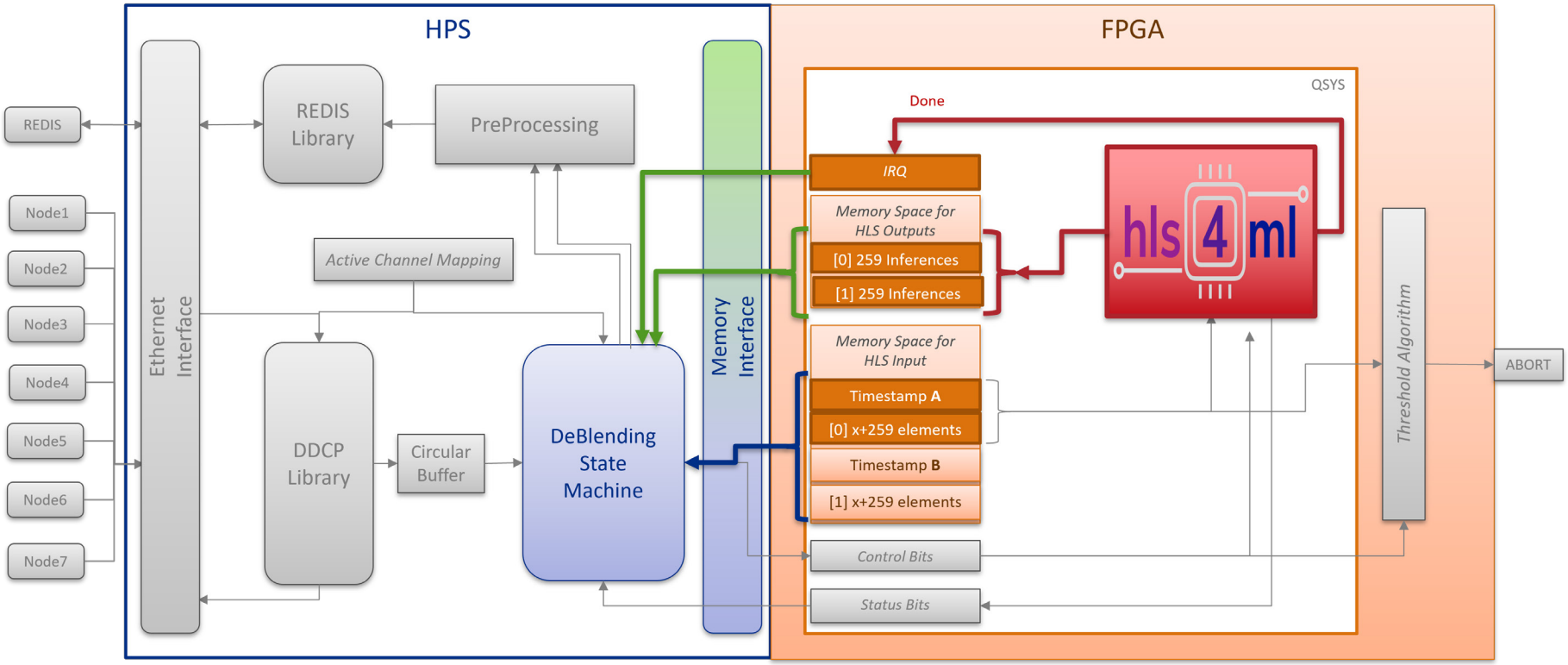
Data Path Thru Central Node (2/5)



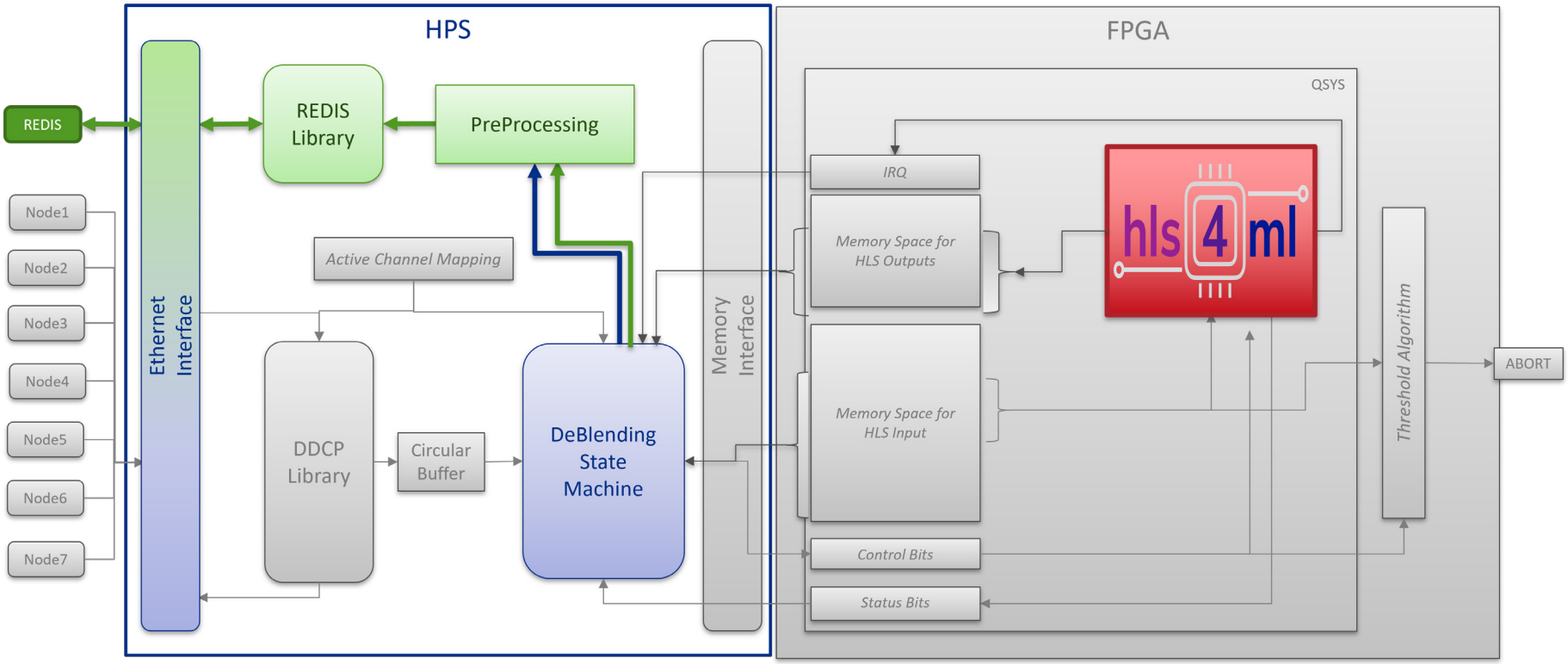
Data Path Thru Central Node (3/5)



Data Path Thru Central Node (4/5)

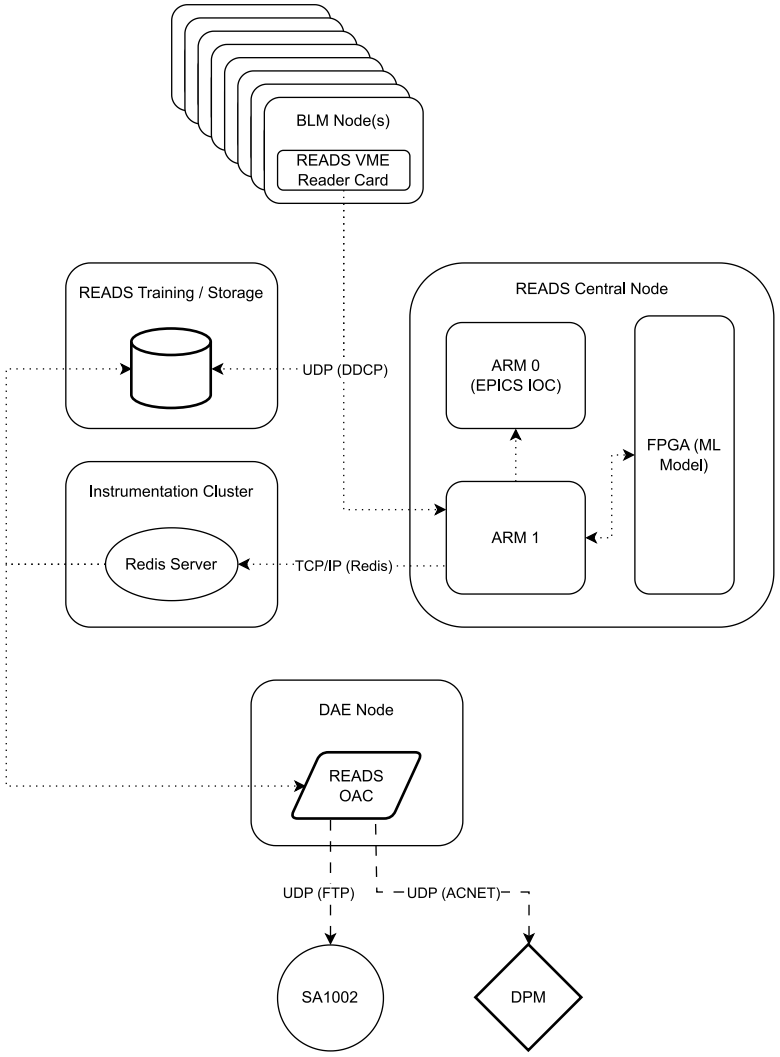


Data Path Thru Central Node (5/5)

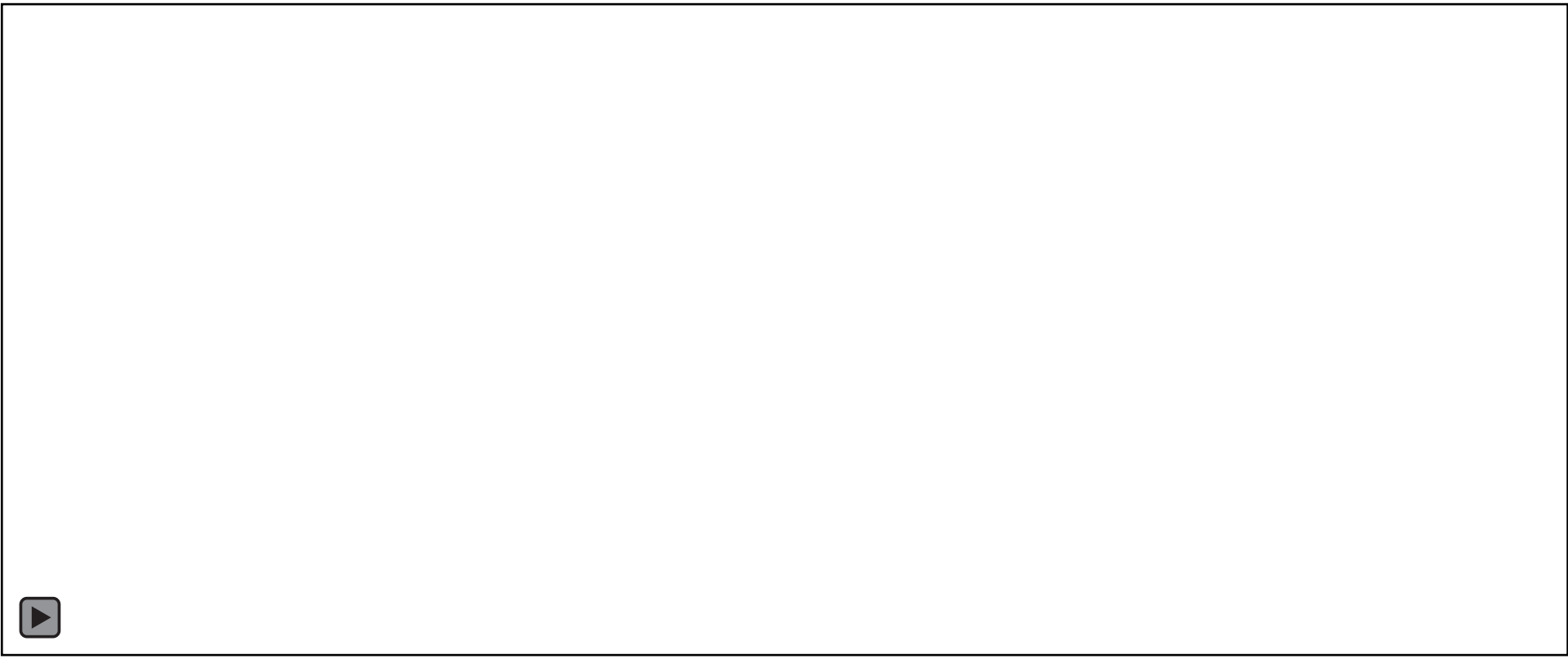


Complete Data Path

- Custom Remote Node board designed and deployed to process readings and stream ML inferences
- Proof of concept using Redis for inference streaming
- An EPICS IOC is providing CA readings
- An Open Access Client translates the inferences into 520+ ACNET devices at 320Hz
- Operators and experts are able to use these readings for everyday tuning using existing ACNET console applications (parameter page, FTP)
- First deployment of Realtime Edge AI on the accelerator controls system and complex



Recycler study under Switchyard slow spill, loss purposefully lost in Recycler



Summary

- Large amounts of data have been collected and processed
- A ML model has been successfully trained to disentangle Main Injector and Recycler beam losses
- The trained ML model has been successfully synthesized and implemented on an FPGA
- Realtime beam loss attribution inferences are live on the accelerators controls network (ACNET)
- READS project ends fall 2023
  - More investigation into improving high frequency data UNet model performance
  - More investigation into model robustness and accuracy before incorporating into machine protection
  - A lot of documentation to be done, papers to write
  - Datasets and code will be packaged and made public after export review

SPECIAL THANKS TO ALL READS COLLABORATORS

Fermilab
M.R Austin, J.M. Arnold, K.J. Hazelwood, P. Hanlet, M.A. Ibrahim, A. Narayanan, D. J. Nicklaus, G. Pradhan, A.L. Saewert, B.A. Schupbach, K. Seiya, R.M. Thurman-Keup, N.V. Tran

Northwestern University
J. Jang, H. Liu, S. Memik, R. Shi, M. Thieme, M. Valentin

READS Publications And Presentations

- Accelerator Real-time Edge AI for Distributed Systems (READS) Proposal (March 2020)
  https://arxiv.org/abs/2103.03928
- Real-Time Edge AI for Distributed Systems (READS): Progress on Beam Loss De-Blending for the Fermilab Main Injector and Recycler (August 2021)
  https://jacow.org/ipac2021/papers/mopab288.pdf
- Optimizing Mu2e Spill Regulation System Algorithms (August 2021)
  https://jacow.org/ipac2021/papers/THPAB243.pdf
- Synchronous High-Frequency Distributed Readout for Edge Processing at the Fermilab Main Injector and Recycler (August 2022)
  https://napac2022.vrws.de/papers/mopa15.pdf
- Semantic Regression for Disentangling Beam Losses in the Fermilab Main Injector and Recycler (August 2022)
  https://napac2022.vrws.de/papers/mopa28.pdf
- Machine Learning for Slow Spill Regulation in the Fermilab Delivery Ring for Mu2e (August 2022)
  https://napac2022.vrws.de/papers/mopa75.pdf
- Disentangling Beam Losses in the Fermilab Main Injector Enclosure Using Real-time Edge AI (November 2022)
  https://indico.bnl.gov/event/16158/contributions/69569/attachments/44217/74598/2022-11-02_icfaml_2022_reads.pdf
- Machine Learning for Slow Spill Regulation in the Fermilab Delivery Ring (November 2022)
  https://indico.bnl.gov/event/16158/contributions/69563/attachments/44212/74590/ICFA%20SRS%20Presentation.pdf
- FPGA-Based Architectures and DAQ Pathways for Distributed ML Systems (February 2023)
  https://indico.fnal.gov/event/58280/contributions/264467/attachments/166403/221599/FPGA_BasedArchitectures_DAQPathways.pdf
- 4+ more planned this fall, final report early next year

THANK YOU!


IBIC2023
12th INTERNATIONAL BEAM
INSTRUMENTATION CONFERENCE
SASKATOON, CANADA
September 10-14, 2023
Canadian Light Source
Centre canadien de rayonnement synchrotron