This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under GA No 101004730.

Task 10.6: Machine Learning Techniques for Accelerator and Target Diagnostics

I.FAST WP10 meeting, 2021-11-15

Thomas Shea (Task leader) and
Irena Dolenc Kittelmann (Deputy Task Leader)
ESS
for the Task 10.6 team

iFAST

10.6 Task Summary

- Long term mission: Develop low-latency Machine Learning (**ML**) techniques to improve performance and availability of high-power facilities at the intensity frontier.
- Goal: Identify signatures of potential errant beam conditions
- Scope:
  - Assess the predictive capabilities of selected ML models
  - Prototype: proof of principle demonstration
    - The most promising ML model will be implemented on a low-latency network of FPGAs processing signals from array of detector channels.

Collaborators and responsibilities

Beneficiaries

- ESS
  - Expertise: Diagnostic scientists, Data manager, FPGA engineers
  - Contribution:
    - Demo platform: Linac
    - FPGA-based diagnostics
    - Low latency communications
- RTU – Faculty of Computer Science and Information
  - Expertise: AI, ML and FPGA
  - Contribution
    - Full-time master student
    - Master student supervision (senior researcher/professor)

Additional participants

- CosyLab – via subcontract from ESS
  - FPGA implementation
- SNS/ORNL – collaboration
  - > decade of operational data
  - Active machine learning project

Timeline – Now in preparation and development phase

- Model **preparation** (Q3 2021 – Q1 2022)
  - ML model exploration
  - Data format preparation
- Assess the predictive capabilities of selected ML models, trained with (Q1 2022 – Q2 2023, **milestone Q4 2022**: selection for demo):
  - simulation results
  - existing SNS operation data (13 years of operation data)
  - ESS commissioning data within the timeline of the project
- ML FPGA platform **development** (Q2 2021 - Q4 2022)
  - Low-latency network development and optimisation
  - Acceleration of ML model on FPGA platform
- Demonstration (Q1 2023 - Q2 2024)
  - In the lab
  - In Linac
- Report (March 2024)

1. Intelligent Trigger - Data Acquisition

- Goal:
  - Identify “off-normal/interesting” and “normal” events
  - Trigger DoD***** acquisition for each “off normal” event

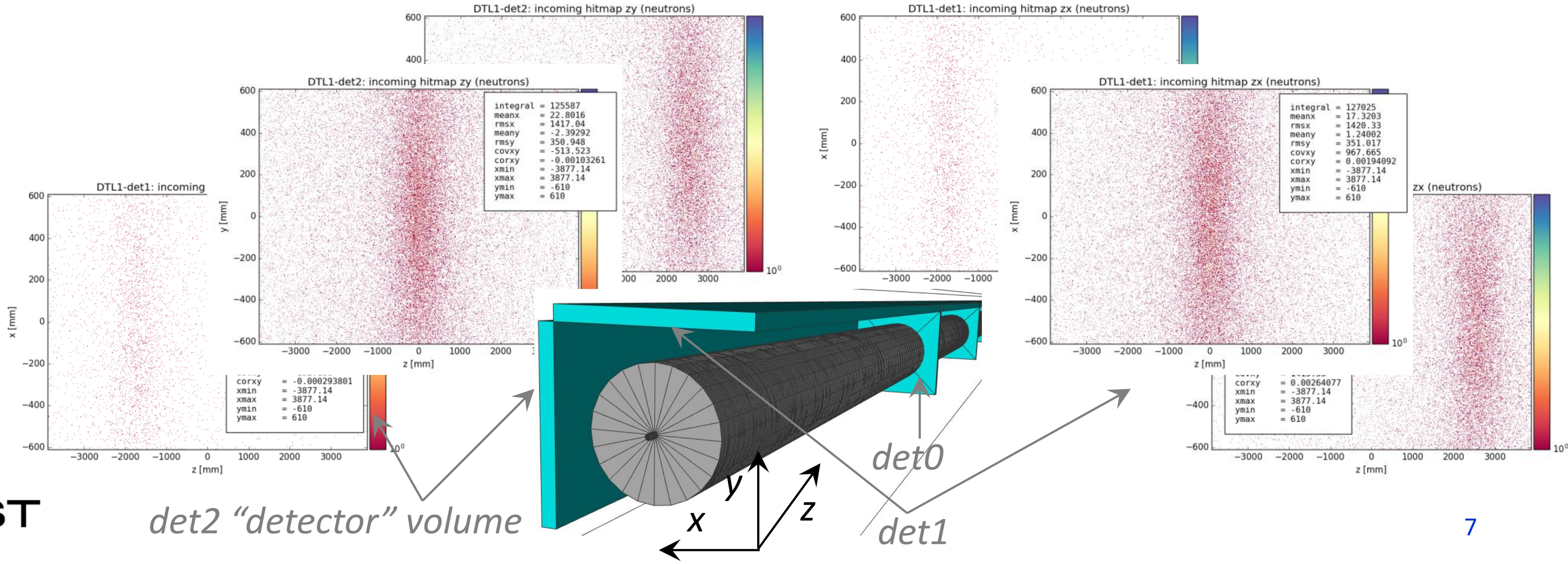
    Data-On-Demand (DoD) is event acquisition: raw or detailed data, buffered and then extracted from the FPGA level on trigger occurrence (on demand)
- Relevant systems: Beam Loss, Beam Position, Beam Current
- Challenges
  - Select appropriate model for anomaly detection
  - Conditions will change with time – may require “online learning” (**unsupervised** ML model; **adaptive** ML)
  - Identify relevant data
    - dataset type and structure (what data to take from each system)
    - dataset triage/”cleaning” (for example remove beam study data)

2. Machine Protection

- Goal: inhibit beam production when damaging conditions are recognised/predicted
- Relevant systems: Loss and Current; also Position, Faraday Cup, etc for protecting specific beam destinations.
- Can be considered an advanced version of ”intelligent trigger” application
  - Correlation of known failures with the datasets extracted with ”intelligent trigger” as precursors (offline analysis)
  - Use these identified precursor datasets (loss patterns or other errant conditions) as training datasets for ML based machine protection

Example: Beam Loss in the ESS Linac

- Figure: Incoming neutron hit maps for different localized loss locations along a DTL tank
- This map is sampled by small number of detectors
- Resulting signal has complex relationship to the beam properties at the potential damage point



Data

- Selected NeXus variant of HDF5 standard (https://www.nexusformat.org)
- Used in the neutron and X-ray communities
- With ORNL, developed emittance data structure.
  - We will build on this to cover ML-related data that we intend to exchange
- Coverage:
  - Operations data from SNS
  - Commissioning data from ESS
  - Simulation data from ESS

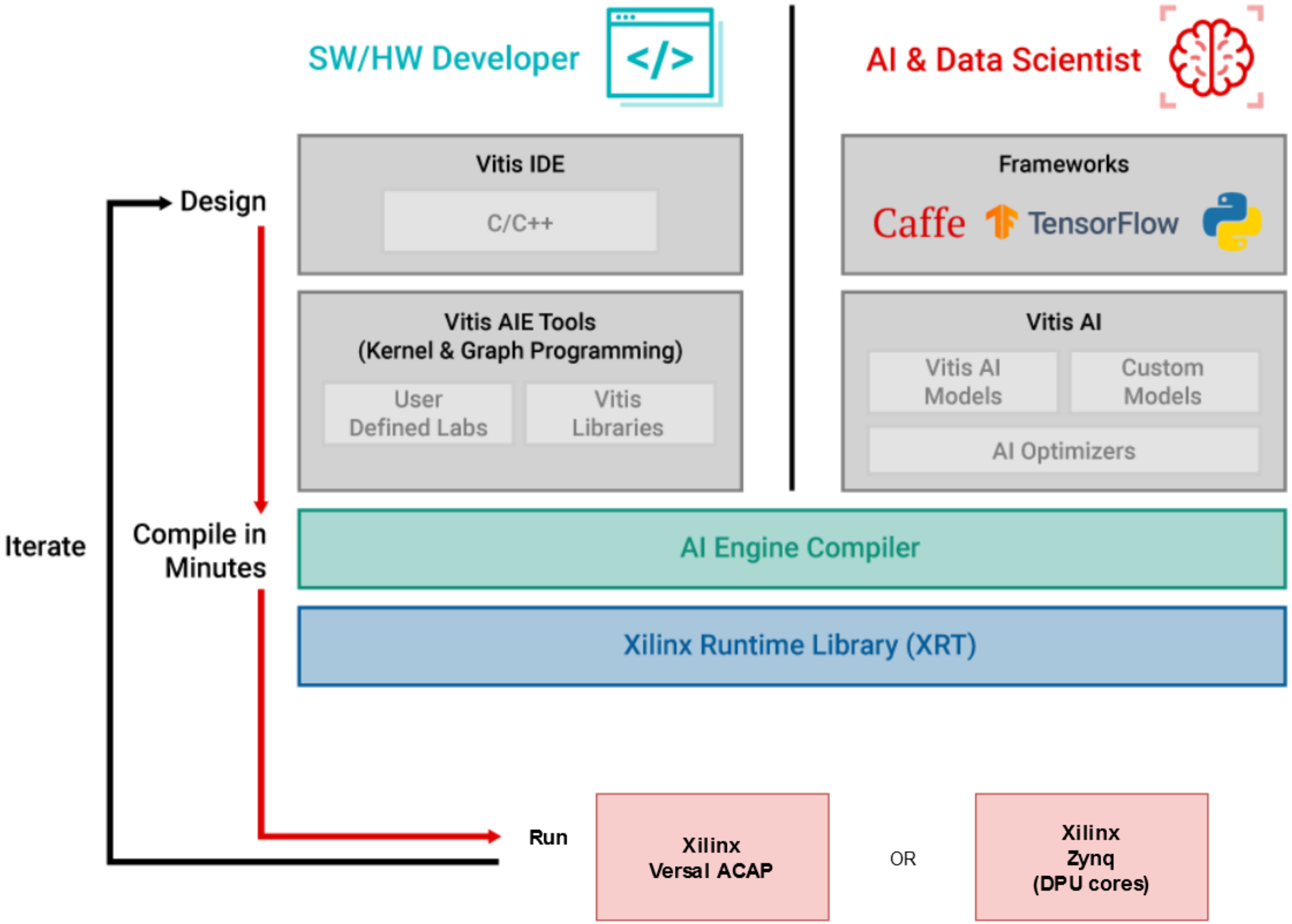
NeXus data format (example: ESS diagnostics data)

NXEntry: Shift ID and/or Asset ID and Timestamp (start and end)

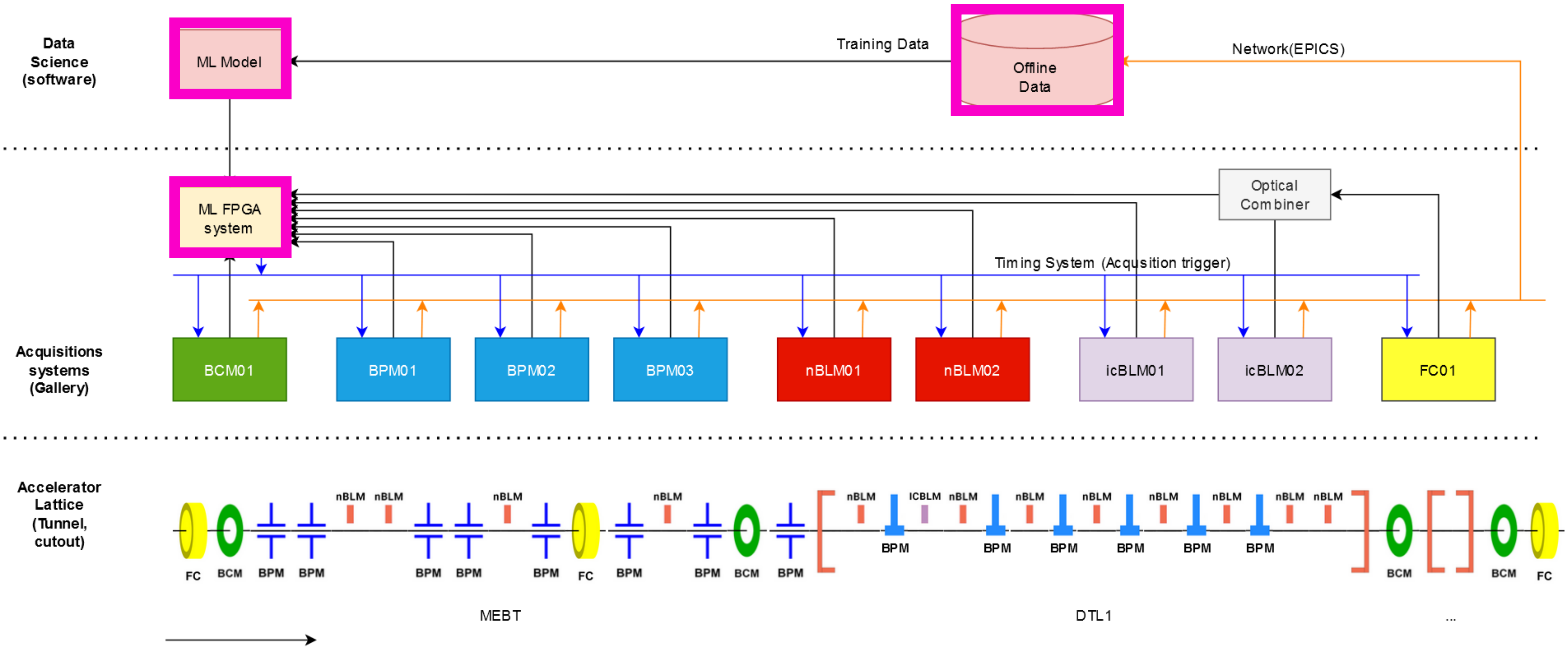
- NXUser
  - Operator Name
  - Affiliation
  - In-Kind ID number
  - role (optional)
- NXInstrument
  - WQe shoudl created new classes for the equipments we use
  - NXSource, NXRf, NXPbi, NXMagnets and NXTiming (new nexus classes)
  - Machine Description can be here (add on NXequipment class and add attributes to describe it)
- NX Parameters
  - Process parameters
- NXNote (optional)
  - Further information needed
- NXData
  - Any data pertinent to the experiment/calibration or verification
  - It is possible to create link between the instrument and data (should be looked into in more details)
- NXLog
  - Special class for time series data

Tools

- 2 platforms
  1. Xilinx Zynq: Current generation
  2. Xilinx Versal ACAP: Bleeding edge with dedicated AI cores
- Advantage: short turn-around and flexibility when exploring ML applications
- The downside: compiled software -> latency
  1. Zynq: Software feeds the accelerator cores.
  2. Versal: can feed AI cores directl from FPGA logic



Demo in Linac



Next steps

- Summarize applications/techniques in I.FAST collaboration space
- Stand up parallel ML toolchains at ESS and RTU, consistent with FPGA platform capabilities
- With SNS, continue collaboration on data exchange and protection applications
- Launch periodic collaboration meetings between ESS and RTU. Add SNS and others as needed to target specific topics.
- Explore synergy with other I.FAST tasks

Thomas Shea, Irena Dolenc Kittelmann, Kaj Rosengren, Roxana Tarkeshian, Elena Donegani, Clement Derrez – ESS

Toms Torims, Agris Nikitenko – DTU

Willem Blokland – ORNL

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under GA No 101004730.

Motivation and Background

High power hadron facilities

- Performance on the intensity frontier is enabled by controlling loss/activation and mitigating beam-induced damage
- Tails of the beam distribution are important contributors to activation
- Subtle changes might be predictors of future damage events
- Challenging to directly measure tails and subtle changes without invasive techniques
- Result: predictions and decisions are based on incomplete and imprecise measurements

Strategy

- Graded approach

  Deployed as trust increases.

  - Intelligent trigger: gather the relevant data for model selection and initial training
  - Alarms to operators
  - Automatic functions: protection and feedback

- Enablers

  Many provided outside of the IFAST program. Task 10.6 extends and leverages these technologies to achieve goals.

  - Triggered, buffered Data Acquisition at the network edge
  - Low latency communication for sensor aggregation
  - Platform to host low latency algorithms
  - Adequate data transport and storage – collaboration with ESS Neutron Science
  - Tools for curation, analysis and offline training
  - At the heart of this: Standard data structures for exchange within the team and with other facilities: NeXuS

Low-latency applications under consideration

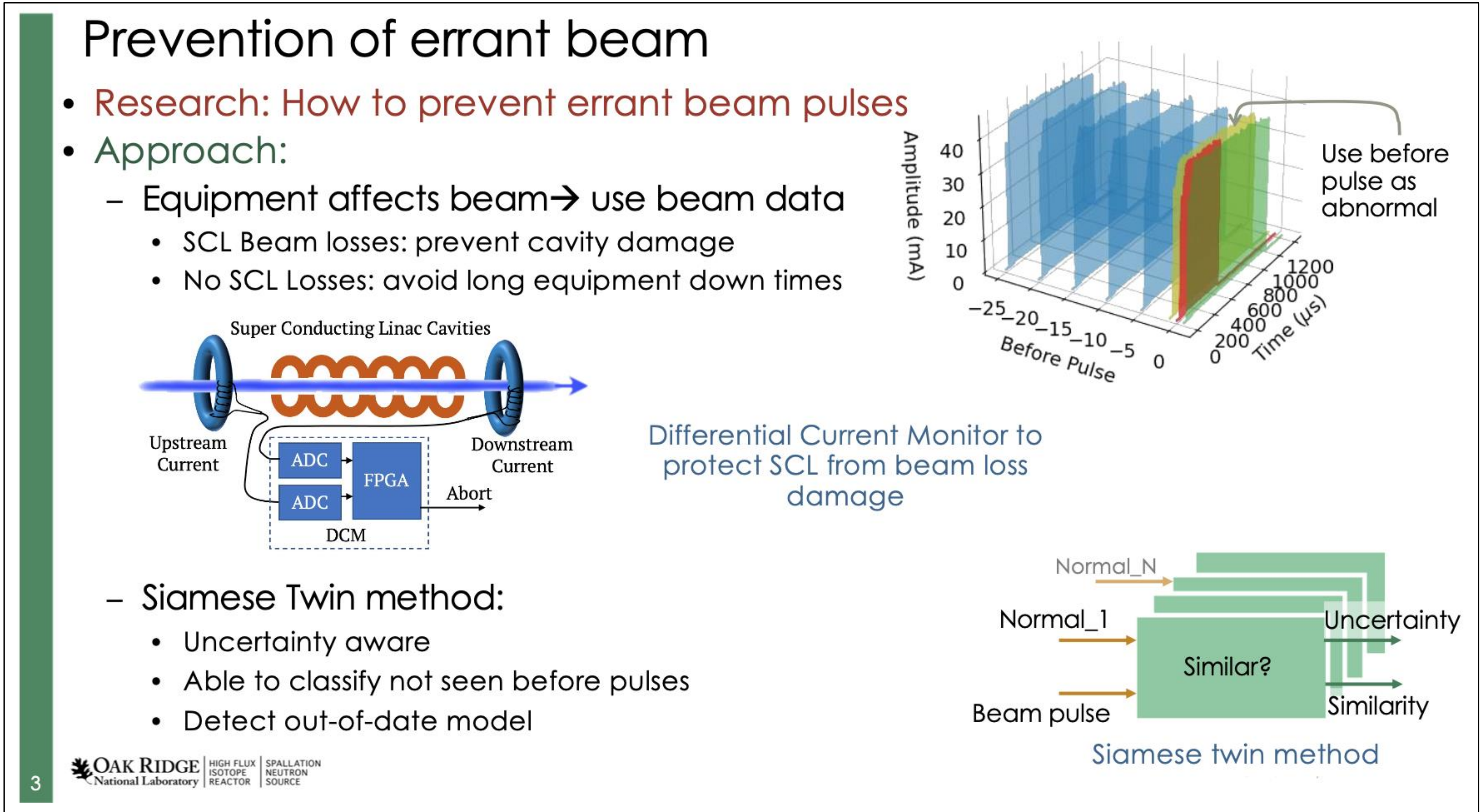
Currently exploring several Low-Latency applications

- ML based Intelligent Trigger for Data-On-Demand (DoD)
- ML based machine protection
  - Beam Loss
  - Beam destination protection
  - Additional sensor aggregation applications
- Interesting low-latency, single-device applications considered but de-emphasized (particle discrimination, image recognition, etc)

Example: Energy deposition on beam destinations

- Damage in Faraday cups and beam stops depends on beam properties that are not directly measured
  - Transverse density distribution
  - Low energy tail of longitudinal distribution
- Relevant measurement devices
  - BPM (centroid phase -> energy)
  - FC – BCM current difference
  - Option: BLM (combine with others to account for upstream loss)
- Use these imperfect but low-latency signals to predict thermo-mechanical response and the resulting damage potential

ORNL/SNS Example

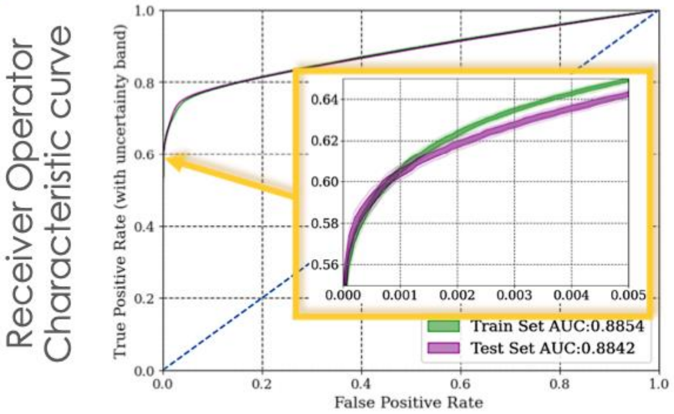


from
W. Blokland

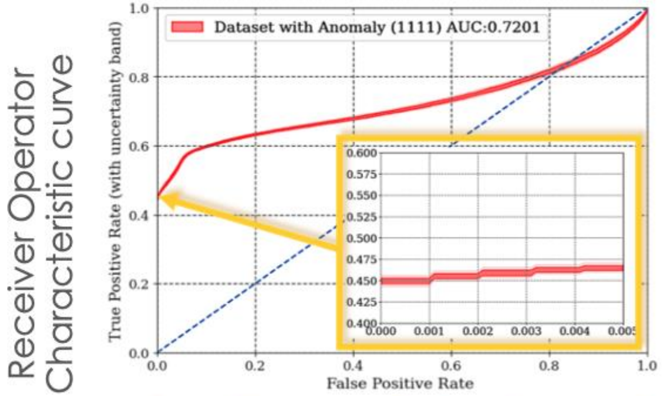
ORNL/SNS Example

Prevention of errant beam

- Status:
  - Determined ML method goal:
    - Keep False Positive rate (FP) < 0.1%
  - Applied method to March 2021 data:
    - ~6100 trips, of which 126 with SCL beam loss
    - We can keep well under 0.1% FP with around 50% True Positive (TP) rate.
- Plan:
  - Implementing Siamese Twin (RT) and Random Forest (FPGA) on DCM:
    - Collect statistics first passively, then with abort active
  - Add equipment classification!
    - We want to predict what equipment is going to fail: keep metrics, determine beam hold-off, and adjust or service equipment
  - Use BPM data (phase)



Results on trained events (no SCL beam losses)



Results on untrained events (with SCL beam losses)

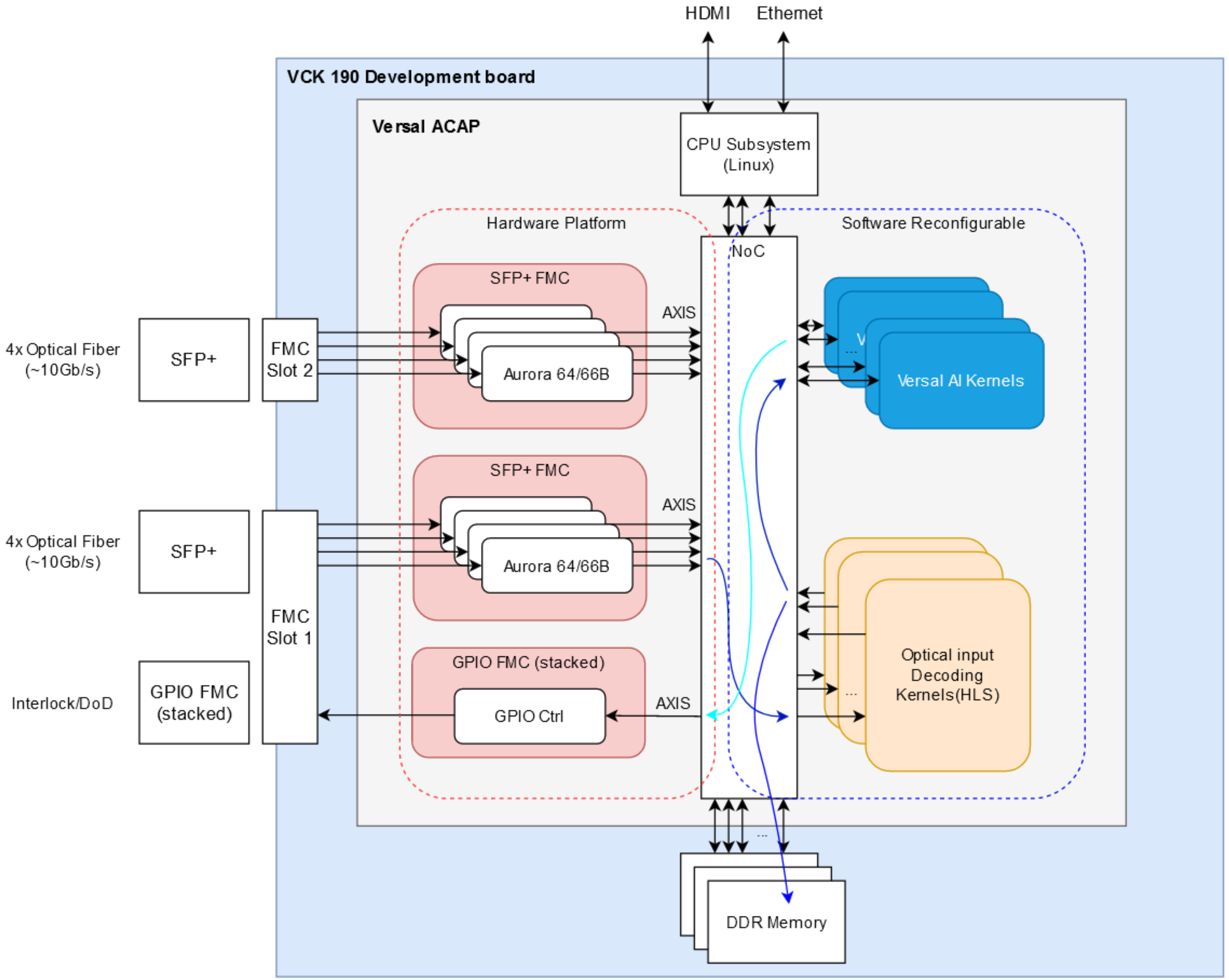
4 OAK RIDGE National Laboratory | HIGH FLUX ISOTOPE REACTOR | SPALLATION NEUTRON SOURCE

ATAC Review
November 1-4, 2021

from
W. Blokland

Platform

- Versal “Adaptable Compute Acceleration Platform, ACAP”
  - The NoC and Versal AI kernels in the diagram is made of dedicated hardware (not FPGA logic)
- The design can be divided into two parts (compare with tool diagram)
  - The “Hardware Platform” part is made by FPGA designer
  - The “Software Reconfigurable” part can be reconfigured in the SW compile flow with quick turnaround.
- Data can be streamed directly from input to the AI kernels for processing.



iFAST