# BigPanDA and DUNE's GPU payload

Sergey Panitkin

(BNL)

# DUNE's GPU payloads

- The DUNE Collaboration is exploring the use of ML methods for online data readout and offline data processing
- Particular emphasis on use of DNNs for event identification and reconstruction
- In February Alexei and Torre met with BNL DUNE folks and learned that they need resources to run DNN
- That conversation has started this project with DUNE GPU payloads in BigPanDA
- I worked with Arbin Timilsina from BNL who provided DUNE codes and documentation for:
  - *Pixel-wise classification network to perform cosmic ray and beam particle separation in prototype DUNE detector*
- This work coincided with my ATLAS ML activities at OLCF
  - Containers for ML on Titan
  - ML stack installation on Summit

# DUNE's GPU payload on Titan

- Titan still is one of the largest supercomputers with GPUs (18,688 K20X Keplers – old, but many of them)
- Old OS and CRAY software stack makes installation of ML software difficult
- Containers provide a good platform for ML stack installation.
  - Allow use of modern OS with all dependencies required by latest ML software versions
  - Allow reuse of existing Docker and Singularity containers  with desired versions of ML software stack
- Singularity containers were build for Titan with CUDA, TensorFlow and Keras
- DUNE's  DNN model was successfully ran on Titan in container
  - Pixel-wise classification network to perform cosmic ray and beam particle separation in prototype DUNE detector

# DUNE's GPU payload on Summit

- Since February ATLAS/BigPanDA has short term project (CSC343) approved on Summit.
  - We asked OLCF for the project scope expansion to run DUNE's payloads
- ML software stacks was installed by us on Summit in the project area
  - Still no official container support on Summit
  - TensorFlow (v1.12), Keras, and all necessary dependencies. Some components needed compilation. *Why this is not available as a standard system software module on Summit?*
  - Current NVidia driver and CUDA versions on Summit are constraints for the ML stack installation by users
- ML stack was tested with:
  - DL image recognition example with MNIST dataset
  - ATLAS b-tagging example
- DUNE's DNN model was trained on Summit (manual job submission)
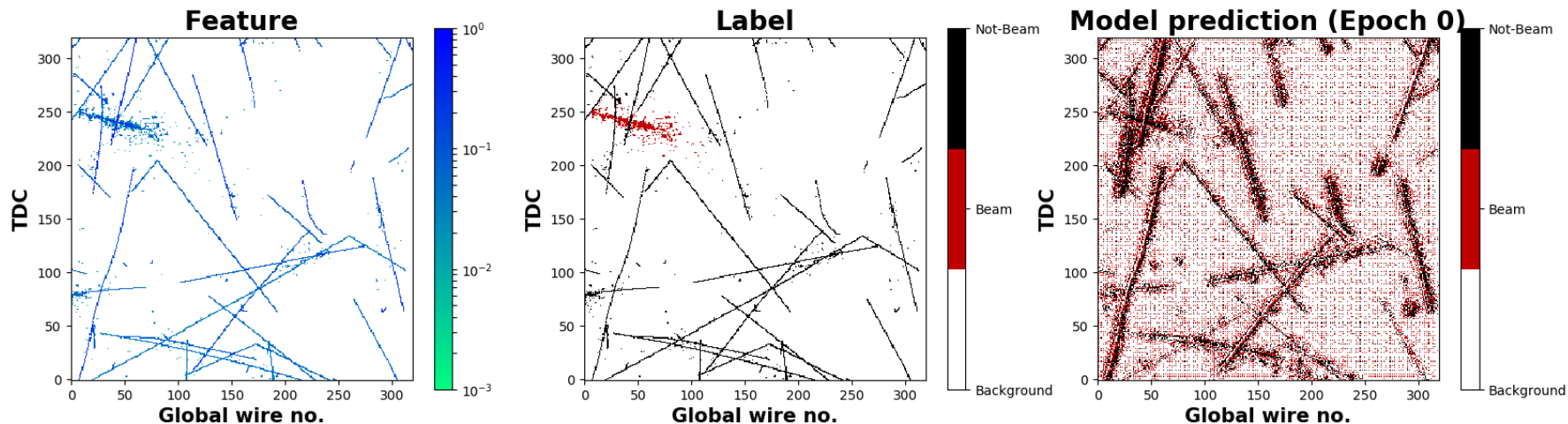
# Summary and Plans

- We started in February 2019 work with DUNE ML codes at machines at OLCF
- We started with creating a container with ML stack on Titan and used it to run DUNE's DNN model
- We installed ML software on Summit.
- The same model was ran on Summit (no containers)
- Dune was accepted by OLCF as a co-project on BigPANDA CSC343 project
- Armin Timilsina applied for an account at OLCF as CSC343 project member


- Harvester is expected to be integrated with Summit soon (Danila)
  - Waiting for installation of the remote job submission client for LSF on OLCF's DTNs
- Dedicated PanDA queue for DUNE can be created on PanDA server on EC2
  - We probably need to discuss general issue of PanDA and Harvester for non-ATLAS projects (Pavlo)
- Perhaps another dedicated PanDA queue is needed to use with the NGE system on Summit
- Expecting in May an update of GPU drivers on Summit and corresponding CUDA upgrade to v10
  - This will allow to use latest ML stack and simplify installation/upgrade procedure (IBM PowerAI)
- We will then update to latest TF and Keras
- Container support and container builder is expected to become available on Summit in May/June timeframe. Perhaps ML stack should be build in a container
- We need to apply for CSC343 project extension and provide a closeout report