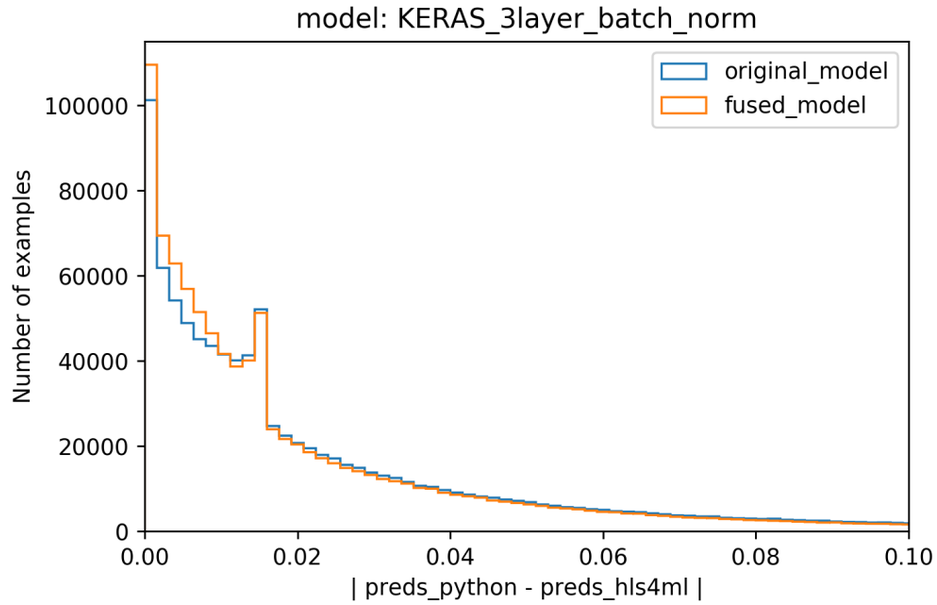
# HLS4ML

Vladimir, Sioni, Jennifer
Petr, Adrian, Hasib, Dejan

# Completed

- Fixed Convolutional layers issues
  - Implemented support for *channels first* and *channels last* options
- Fusion of dense and batch normalization layers
  - Include BN parameters in the preceding dense layer
  - Remove BN layer
    - fused_W = gamma * W / sqrt(variance + eps)
    - fused_b = gamma * (b - mean) / sqrt(variance + eps) + beta
    - Fewer layers on the board = reduced latency
    - (Slightly) increased precision
- Started developing support for multiple backends
  - Vivado, Mentor Catapult, Intel




| | Latency [clock cycles] |
|---|---|
| Original | 27 |
| Fused | 19 |

# Next Steps

- Fuse convolutional layers with batch normalization
- Develop additional features for HLS4ML
- Compare HLS4ML performance for 2D CNN with Micron