Contribution ID: **55**   Type: **Lightning Talk**

Graph Neural Networks on FPGAs with HLS4ML

Monday 25 September 2023 18:10 (5 minutes)

Graph structures are a natural representation of data in many fields of research, including particle and nuclear physics experiments, and graph neural networks (GNNs) are a popular approach to extract information from that. Simultaneously, there is often a need for very low-latency evaluation of GNNs on FPGAs. The HLS4ML framework for translating machine learning models from industry-standard Python implementations into optimized HLS code suitable for FPGA applications has been extended to support GNNs constructed using PyTorch Geometric (PyG). To that end, the parsing of general PyTorch models using symbolic tracing using the torch.FX package has been added to HLS4ML. This approach has been extended to enable parsing of PyG models and support for GNN-specific operations has been implemented. To demonstrate the performance of the GNN implementation in HLS4ML, a network for track reconstruction in the sPHENIX experiment is used. Future extensions, such as an interface to quantization-aware training with Brevitas, are discussed.

Primary authors: SCHULTE, Jan-Frederik (Purdue University (US)); LIU, Miaoyuan (Purdue University (US)); HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US)); LONCAR, Vladimir (Massachusetts Inst. of Technology (US))

Presenter: SCHULTE, Jan-Frederik (Purdue University (US))

Session Classification: Contributed Talks

Track Classification: Contributed Talks