Contribution ID: **60** Type: **not specified**

Graph neural networks for FPGAs

Wednesday, April 22, 2020 7:50 PM (25 minutes)

Graph neural networks have been shown to achieve excellent performance for several crucial tasks in particle physics, such as charged particle tracking, jet tagging, and clustering. An important domain for the application of these networks is the Level-1, FGPA-based trigger, which has strict latency and resource constraints. We discuss how to design distance-weighted graph networks that can be executed with less than 1 μs latency on an FPGA. To do so, we consider representative tasks associated with particle reconstruction and identification in a next-generation calorimeter operating at a particle collider. We use graph architectures developed for these purposes and simplified in order to match the computing constraints of real-time event processing at the CERN Large Hadron Collider. The trained models are compressed using pruning and quantization. Using the hls4ml library, we convert the compressed models into firmware to be implemented on an FPGA. We show results both in terms of model accuracy and computing performance.

Consider for young scientist forum (Student or postdoc speaker)

Second most appropriate track (if necessary)

Primary authors: DUARTE, Javier Mauricio (Univ. of California San Diego (US)); HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US)); IIYAMA, Yutaro (CERN)

Presenter: IIYAMA, Yutaro (CERN)

Session Classification: Recording sessions