



Contribution ID: 70

Type: **Poster and Flash talk**

Improving the Inference of Graph Neural Networks for Track Reconstruction

Monday 23 September 2024 16:33 (3 minutes)

Optimizing the inference of Graph Neural Networks (GNNs) for track finding is crucial for enhancing the computing performance of particle collision event reconstruction. Track finding involves identifying and reconstructing the paths of particles from complex, noisy detector data. By leveraging GNNs, we can model the relationships between detector hits as a graph, where nodes represent hits and edges represent potential connections between them. To speed up the inference of these GNN models, it is important to reduce computational overhead, improve model architecture, and exploit hardware accelerators such as GPUs. Techniques like quantization and pruning can be employed to minimize model size and inference time without sacrificing accuracy.

What of the following keywords match your abstract best?

GPUs

Please tick if you are a PhD student and wish to take part to the poster prize competition!

Other

Primary authors: LAZAR, Alina (Youngstown State University (US)); PASCHKE, Henry; JSGABORIAULTWHIT@STUDENT.YSU.EDU, James; CHAN, Jay (Lawrence Berkeley National Lab. (US)); DANG, Minh Tuan (CERN); CALAFIURA, Paolo (Lawrence Berkeley National Lab. (US)); JU, Xiangyang (Lawrence Berkeley National Lab. (US))

Presenters: LAZAR, Alina (Youngstown State University (US)); PASCHKE, Henry; JSGABORIAULTWHIT@STUDENT.YSU.EDU, James; CHAN, Jay (Lawrence Berkeley National Lab. (US)); DANG, Minh Tuan (CERN); CALAFIURA, Paolo (Lawrence Berkeley National Lab. (US)); JU, Xiangyang (Lawrence Berkeley National Lab. (US))

Session Classification: Flash talks / poster session