



Contribution ID: 205

Type: Poster

Efficient TrackML Data Access Using HDF5 for Scalable Particle Tracking

The TrackML dataset, a benchmark for particle tracking algorithms in High-Energy Physics (HEP), presents challenges in data handling due to its large size and complex structure. In this study, we explore using a heterogeneous graph structure combined with the Hierarchical Data Format version 5 (HDF5) not only to efficiently store and retrieve TrackML data but also to speed up the training and inference of the Graph Neural Network (GNN) models used for tracking.

We reorganize the TrackML dataset into a heterogeneous graph structure using PyTorch Geometric (PyG) to represent better the complex relationships in tracking detector data. In this representation, hit and track entities are modeled as distinct node types, with multiple edge types capturing interactions such as hit-hit spatial connections and hit-track associations. This heterogeneous structure enables more expressive GNN architectures that can leverage semantic information across node and edge types, leading to improved modeling of tracking behavior and enhanced flexibility for multi-relational learning tasks.

The conversion of TrackML CSV files to HDF5 enables rapid, scalable access to event-based particle tracking information while maintaining data integrity and structure. The HDF5 format significantly improves read speed, storage efficiency, and ease of data manipulation. The implementation supports fast indexing, event filtering, and compatibility with parallel processing workflows, which are critical for machine learning applications in particle physics. Benchmark results show compression gains and faster read performance than standard CSV and PyG parsing. This approach facilitates more efficient experimentation and prototyping in TrackML-based research and can be extended to other large-scale physics datasets.

Significance

References

Experiment context, if any

Authors: CIPRIJANOVIC, Aleksandra (Fermi National Accelerator Laboratory); LAZAR, Alina (Youngstown State University (US)); CERATI, Giuseppe (Fermi National Accelerator Lab. (US)); CHAN, Jay (Lawrence Berkeley National Lab. (US)); WHITELY, Meara (Youngstown State University); CALAFIURA, Paolo (Lawrence Berkeley National Lab. (US)); HEWES, V; JU, Xiangyang (Lawrence Berkeley National Lab. (US))

Presenter: LAZAR, Alina (Youngstown State University (US))

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research