

Development of the neural network algorithm on Versal AI-engine

Wednesday, 20 August 2025 11:40 (20 minutes)

Next generation high-energy experiments aim to probe unexplored regimes in particle physics. Real-time data processing of terabyte-per-second is a critical technological bottleneck, especially, track reconstruction suffer from degraded efficiency under high-background environment. To address this challenge, we are designing a heterogeneous data processing system that integrates a real-time graph neural network(GNN) algorithm into the Xilinx Versal ACAP, as a potential upgrade for future detector. To demonstrate the feasibility of the system, we implement baseline neural network deployments, including an online deep neural network (DNN) algorithm and an offline GNN algorithm, on the AI Engine of the Xilinx Versal ACAP. As a result, the best latency achieved for the DNN deployment is approximately 2 microseconds per event, while the GNN deployment achieves 1 milli second latency with a 27 ϵ -nearest neighbor (ϵ -NN) graph construction input. In this report, we will show the design and performance of the development.

Primary author: 刘, 兆志

Presenter: 刘, 兆志

Session Classification: Session