Contribution ID: **82**   Type: **Tutorial**

Super Neural Architecture Codesign Package (SNAC-Pack)

Monday 1 September 2025 11:00 (1h 30m)

Machine learning has become a critical tool for analysis and decision-making across a wide range of scientific domains, from particle physics to materials science. However, the deployment of neural networks in resource-constrained environments, such as hardware accelerators and edge devices, remains a significant challenge. This often requires specialized expertise in both neural architecture design and hardware optimization.

To address this challenge, we introduce the Super Neural Architecture Codesign Package (SNAC-Pack), an integrated framework that automates the discovery and optimization of neural network architectures specifically tailored for hardware deployment. SNAC-Pack combines two powerful tools: Neural Architecture Codesign, which performs a two stage neural architecture search for optimal models, and the Resource Utilization and Latency Estimator, which predicts how an architecture will perform when implemented on FPGA software.

SNAC-Pack streamlines the neural architecture design process by enabling researchers to automatically explore diverse architectures optimized for both task performance and hardware efficiency. By providing quick estimates of resource utilization and latency without requiring time-consuming synthesis, SNAC-Pack accelerates the development cycle. State-of-the-art compression techniques, such as quantization-aware training and pruning, further optimize the models, resulting in architectures that can be deployed to FPGA hardware.

This tutorial provides a hands-on introduction to SNAC-Pack, guiding participants through the complete workflow from dataset preparation to hardware deployment. By the end of the tutorial, attendees will be able to run SNAC-Pack for their own applications, achieving improvements in accuracy, latency, and resource utilization compared to naive hand-crafted approaches.

Authors: DEMLER, Dmitri; WEITZ, Jason (Univ. of California San Diego (US))

Co-authors: HAWKS, Ben (Fermi National Accelerator Lab); DUARTE, Javier Mauricio (Univ. of California San Diego (US)); TRAN, Nhan (Fermi National Accelerator Lab. (US))

Presenter: DEMLER, Dmitri

Session Classification: Tutorials