

# Neural Architecture Codesign for Physics Applications

*Friday 30 August 2024 14:40 (20 minutes)*

We develop an automated pipeline to streamline neural architecture codesign for physics applications. Our method employs neural architecture search to enhance these models, including hardware costs, leading to the discovery of more hardware-efficient neural architectures. We exceed performance and show further speedup through model compression techniques such as quantization-aware-training and neural network pruning. We synthesize the ideal models to high level synthesis code for FPGA deployment with the hls4ml library. Additionally, our hierarchical search space provides greater flexibility in optimization, which can easily extend to other tasks and domains. We demonstrate this with two case studies: Bragg Peak finding in materials science and jet classification in high energy physics.

**Presenters:** DEMLER, Dmitri; WEITZ, Jason