

## CryoAI –Prototyping cryogenic chips for machine learning at 22nm

*Tuesday, 4 October 2022 16:45 (15 minutes)*

We present our design experience of a prototype System-on-Chip (SoC) for machine learning applications that run in a cryogenic environment to evaluate the performance of the digital backend flow. We combined two established open-source projects (ESP and HLS4ML) into a new system-level design flow to build and program the SoC. In the modular tile-based architecture, we integrated a low-power 32-bit RISC-V microcontroller (Ibex), 200KB SRAM-based scratchpad, and an 18K-parameter neural-network accelerator. The network is an autoencoder working on audio recordings and trained on industrial use cases for the early detection of failures in machines like slide rails, fans, or pumps. For the hls4ml translation, we optimized the reference architecture using quantization and model compression techniques with minimal AUC performance reduction. This project is also an early evaluation of Siemens Catapult as an HLS backend for hls4ml. Finally, we fabricated the SoC in a 22nm technology and are currently testing it.

**Primary authors:** SYAL, Chinar (Fermi National Accelerator Lab. (US)); GIRI, Davide (Columbia University); FAHIM, Farah (Fermilab); DI GUGLIELMO, Giuseppe (Fermilab); ZUCKERMAN, Joseph (Columbia University); CARLONI, Luca (Columbia University); CASSEL DOS SANTOS, Maico (Columbia University); Mr VALENTIN, Manuel (Northwestern University); TRAN, Nhan (Fermi National Accelerator Lab. (US)); MEMIK, Seda (Northwestern University)

**Presenter:** Mr VALENTIN, Manuel (Northwestern University)

**Session Classification:** Contributed Talks