Contribution ID: **44** Type: **Parallel Presentation**

Automated and Holistic Co-design of Smart Readout ASICs with Embedded Machine Learning

Thursday, November 21, 2024 11:15 AM (15 minutes)

Extreme edge-AI systems, such as those in readout ASICs for radiation detection, must operate under stringent hardware constraints such a micron-level dimensions, sub-milliwatt power, and nanosecond-scale speed while providing clear accuracy advantages over traditional architectures. Finding ideal solutions means identifying optimal AI and ASIC design choices from a design space that has explosively expanded during the merger of these domains, creating non-trivial couplings which together act upon a small set of solutions as constraints tighten. It is impractical, if not impossible, to manually determine ideal choices among possibilities that easily exceed billions even in small-size problems. Existing methods to bridge this gap have leveraged theoretical understanding of hardware to create proxies for key metrics such as ASIC area and power and used them in neural architecture search. However, the assumptions made in computing such theoretical metrics are too idealized to provide sufficient guidance during the difficult search for a practical implementation. Meanwhile, theoretical estimates for many other crucial metrics (like delay) do not even exist and are similarly variable, dependent on parameters of the process design kit (PDK). To address these challenges, we employ intelligent search using multi-objective Bayesian optimization, integrating both neural network search and ASIC synthesis in the loop. This approach provides reliable feedback on the collective impact of all cross-domain design choices. We showcase the effectiveness of our approach by finding several Pareto-optimal design choices for effective and efficient neural networks that perform real-time feature extraction from input pulses within the individual pixels of a readout ASIC. The proposed optimization approach was used to realize a smart readout ASIC for segmented radiation detectors. The chip, which was designed in 65 nm CMOS technology, contains 23 independent sensing channels. Each channel features a low-noise analog front-end, single-ended to differential converter, ADC driver, high-speed 12-bit ADC, digital signal processor (DSP), and artificial neural network with on-chip weights for performing regression or classification tasks. The DSP was realized using a high-level synthesis design flow. Each channel contains 1.8 kb of on-chip memory and consumes approximately 14.3 mW at the nominal sampling rate of 25 MS/s.

Primary author: MANDAL, Soumyajit (Brookhaven National Laboratory)

Co-authors: Mr GORNI, Dominik (Brookhaven National Laboratory); Dr DEPTUCH, Grzegorz (Brookhaven National Laboratory); Dr MAJ, Piotr (Brookhaven National Laboratory); Dr MUKIM, Prashansa (Brookhaven National Laboratory); Dr WOO, Shinjae (Brookhaven National Laboratory); Dr KHAREL, Shubha (Brookhaven National Laboratory); Dr REN, Yihui (Brookhaven National Laboratory)

Presenter: Dr KHAREL, Shubha (Brookhaven National Laboratory)

Session Classification: RDC 04 - Readout and ASICs Parallel Session