

COMPUTATIONAL RESEARCH in BOSTON and BEYOND SEMINAR

AI Hardware Accelerators Survey

ALBERT REUTHER

MIT - Lincoln Laboratory

ABSTRACT:

Certain aspects of Moore's law arguably have ended, as have a number of related laws and trends including Denard's scaling (power density), clock frequency, core counts, instructions per clock cycle, and instructions per Joule (Koomey's law). Taking a page from the system-on-chip (SoC) trends first seen in automotive systems and smartphones, advancements and innovations are still progressing by developing and integrating accelerators for often-used operational kernels, methods, or functions. Over the past several years, startups and established technology companies have been announcing, releasing, and deploying a wide variety of artificial intelligence accelerators. The focus of these accelerators has been on accelerating deep neural network (DNN) models, and the application space spans from very low power embedded voice recognition to data center scale training. Understanding the relative benefits of these technologies is of particular importance to applying AI/ML to many industry challenges and National Security domains under significant constraints such as size, weight, and power, both in embedded applications and in data centers.

This seminar will share the results of an on-going, five-year study that has been surveying AI accelerators (and accelerators, in general), including their architectures, their capabilities, and their applicability to various embedded and data center applications. The survey has grown to include well over 100 ML accelerators, and they provide a basis with which we will discuss the trends of the accelerators and what to expect in the coming years.

FRIDAY, APRIL 7, 2023

12:00 PM – 1:00 PM

<https://math.mit.edu/sites/crib/>

<https://mit.zoom.us/j/96155042770>