COMPUTATIONAL RESEARCH
in BOSTON and BEYOND SEMINAR

Ultra-low latency energy saving computing with silicon photonics

DR. JOHANNES FELDMANN
Co-Founder and CTO at Salience Labs, Oxford, UK.

ABSTRACT:

Silicon photonics (SiPh) achieves ultra-low latency processing and data movement and promises qualitative improvement in energy efficiency. We review existing SiPh technology, benefits, and remaining challenges. SiPh removes the gap in data communication between fiber-optical macro-level connectivity and the present electronic digital-based data movement at the system down to the chip level. SiPh also rejuvenates optical and electrical analog computing taking advantage of low-cost mass manufacturing of SiPh chips and their packaging with traditional electronic digital components, e.g., an amplitude-based modulation allows for a chip that clocks at 10's of GHz.

Matrix multiplication (MatMul) is the key compute operation in AI inference, signal processing, optimal control, numerical modelling, and other matrix-math heavy compute workloads. A hybrid electronic/SiPh chip can achieve orders of magnitude performance improvement in low precision MatMul, e.g., delivers a 60x reduction in latency for full AI inference workloads in our simulation and can be mass-manufactured at production-level foundries. Effective algorithmic/architecture co-design requires a better understanding of applications that can take advantage of reduced latency and benefit from low energy consumption. We are open for collaborations with potential customers in need for customized chips.

References:
1. Parallel convolutional processing using an integrated photonic tensor core. Nature, 2021.
2. All-optical spiking neurosynaptic networks with self-learning capabilities. Nature, 2019.

FRIDAY, MAY 5, 2023
12:00 PM – 1:00 PM

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

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

Massachusetts Institute of Technology