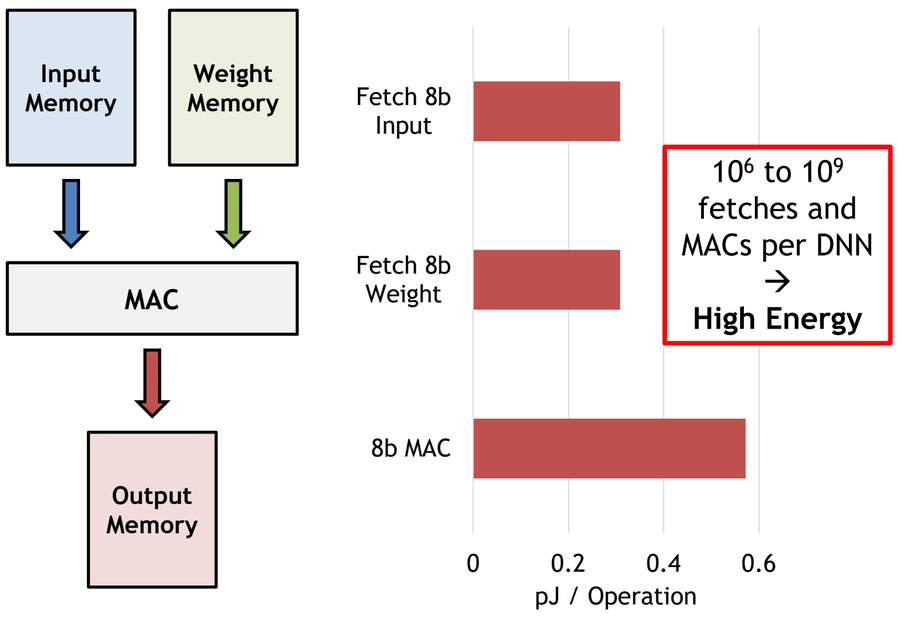
Architecture-Level Modeling of Photonic Deep Neural Network Accelerators

Tanner Andrulis, Gohar Irfan Chaudhry, Vinith M. Suriyakumar, Joel S. Emer, Vivienne Sze
Massachusetts Institute of Technology

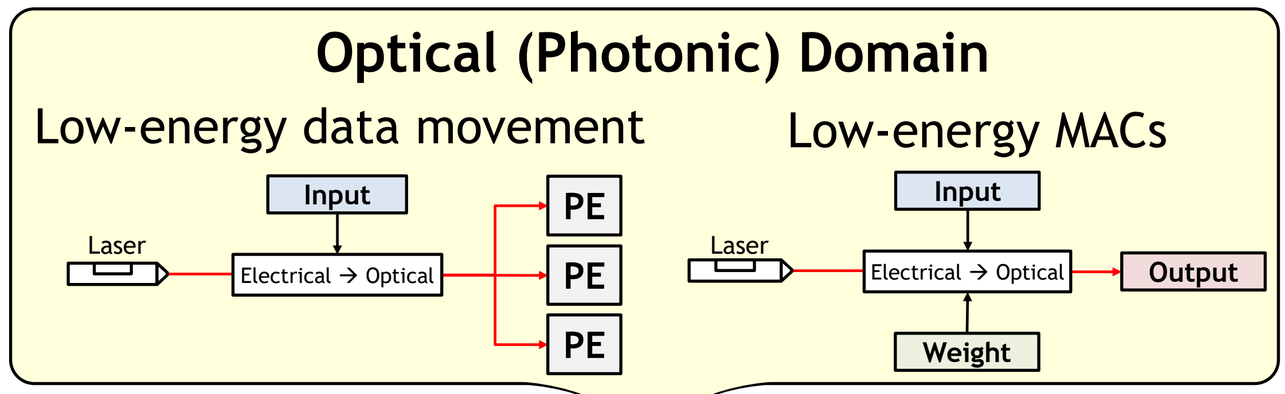
Motivation: Photonic Deep Neural Network (DNN) Accelerators

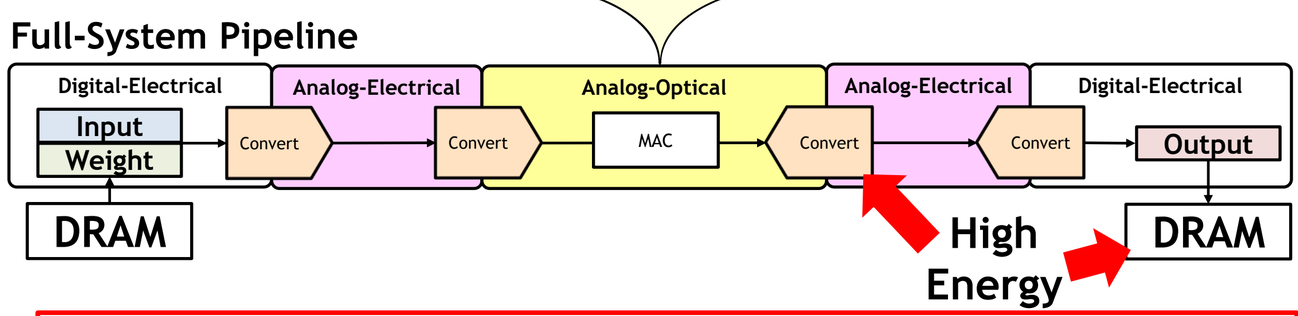
Conventional DNN Accelerator

High energy for data movement & multiply-accumulate (MAC)



Optical (Photonic) Domain

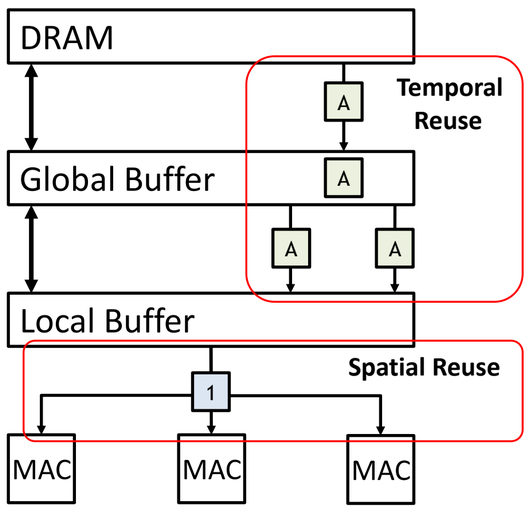


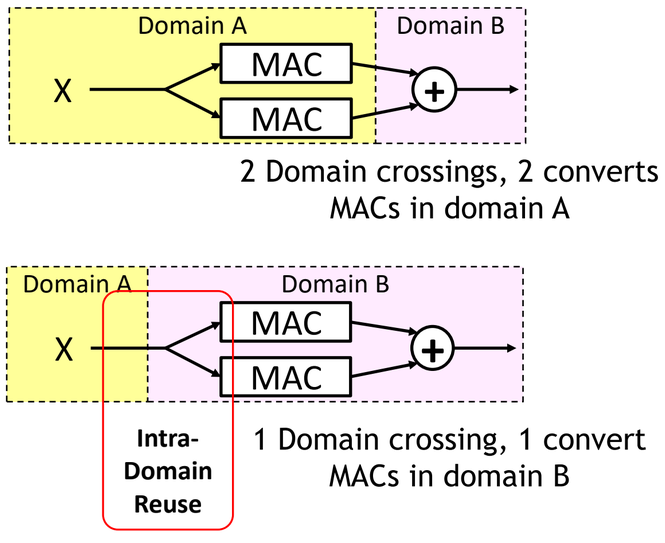


Optical isn't the whole story!
Must balance optical domain benefits and conversion costs and integrate into full systems with DRAM

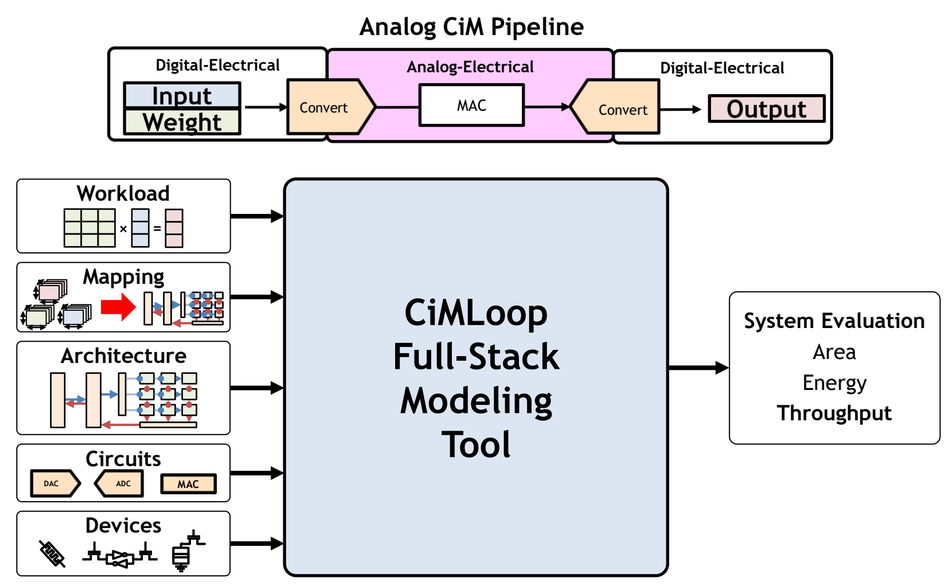
Approach: Key Tradeoffs in & Modeling of Photonic DNN Accelerators

Spatial & temporal reuse → Reduce data movement energy
Intra-domain reuse → Reduce data converter energy



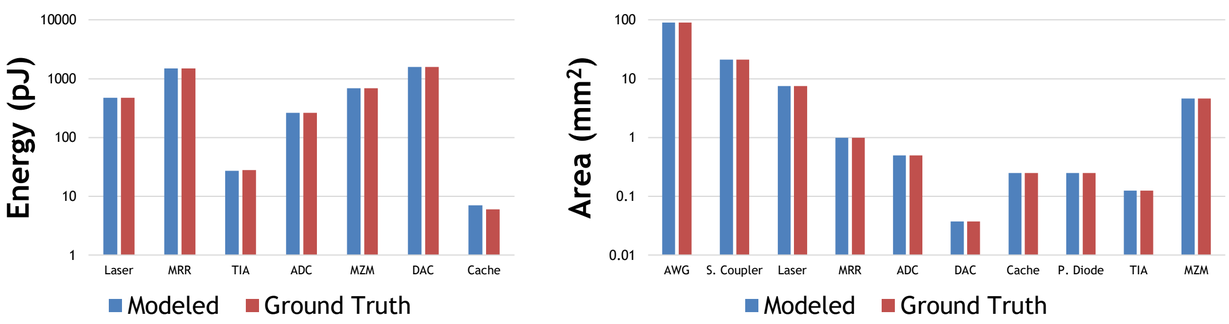


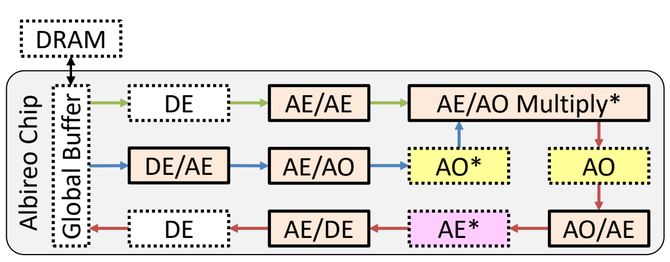
Analog compute-in-memory (CiM) uses multiple domains too
→ **Model photonics with CiM tools**

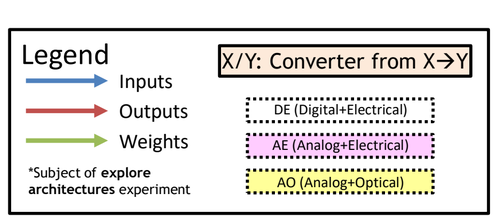


Results: Validation and Case Studies

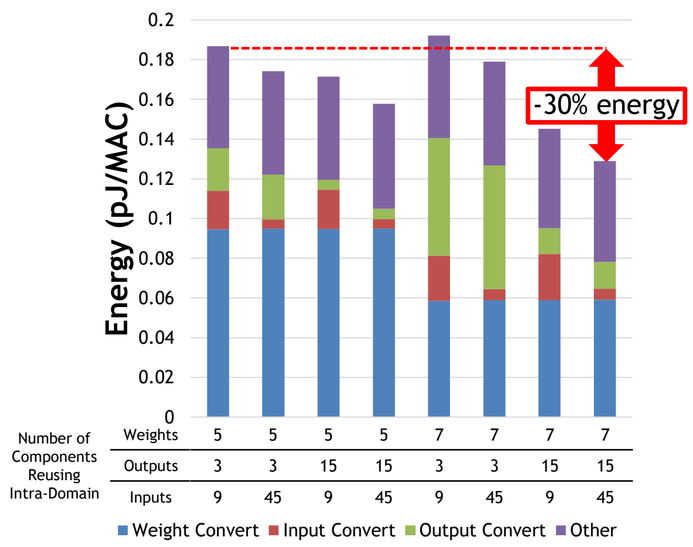
Validated model of the Albireo photonic accelerator [**Shiflett**, ISCA 2021]



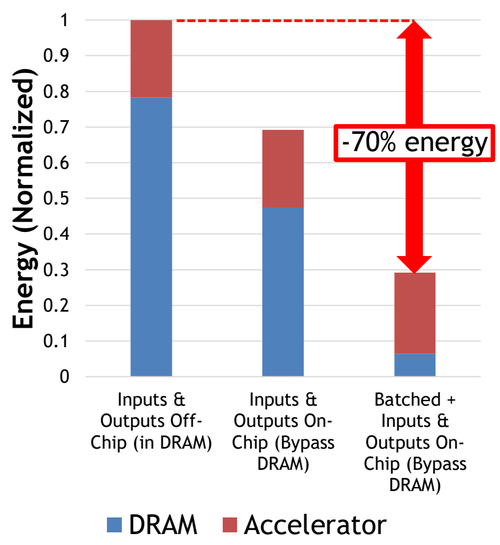




Explore architectures
Conversion consumes high energy. Leverage intra-domain reuse to reduce energy by 30%.



Evaluate full systems
DRAM consumes high energy. Batch & keep operands on-chip to reduce energy by 70%.



Energy-Efficient Multimedia Systems Group
(www.rle.mit.edu/eems)

This work was funded in part by Ericsson, TSMC, the MIT AI Hardware Program, and MIT Quest.

Open-source model

https://github.com/mit-emze/cimloop