Ultra-Fast Photonic Neural Networks

James Garofolo
Rowan University, New Jersey, USA
Ben Zierdt, Taichu Shi, Yang Qi, Dr. Ben Wu

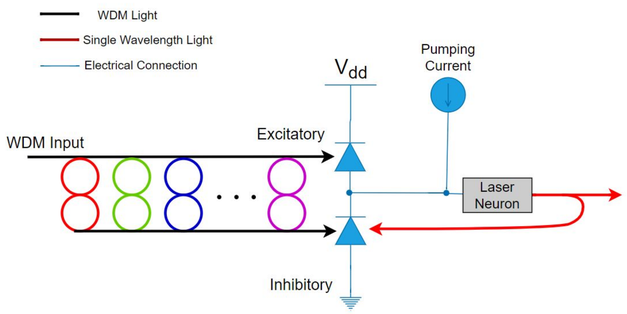
NFC Scan for GitHub!

Problem Statement:

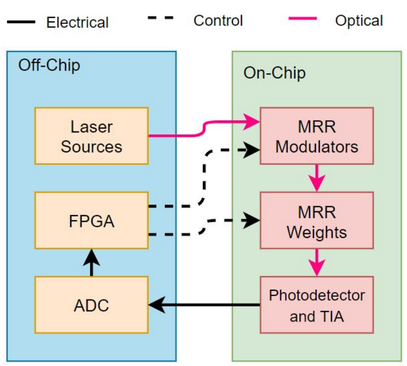
- High data rates of Analog and Optical channels create challenges for digital signal processing applications
- Artificial neural networks are of particular concern, as their effectiveness in signal processing applications is counteracted by their computational intensity for high speed applications
- Deep neural networks suffer more from this issue, requiring GPU's or specialized accelerators to achieve throughput rates in the KHz range
- Specialized hardware that performs calculations in the optical domain could be made to circumvent these challenges

Current State of the Art:

- Optical dot product engines allow for implementations of both constant value and spiking networks on photonic integrated circuits
- Minimum waveguide dimensions severely limit the neuron density of chips in static layer implementations



- Infinite scalability is achievable by way of time-division multiplexing of weight bank resources
- Adding layers in this configuration adds delay at a rate of O(n)
- Adding neurons to each layer adds delay at a rate of O(ceil(n/w)) where w is the number of available hardware neurons in a chip
- The most impactful bottleneck cited in this system is described to be the digitization of data between layers [1]

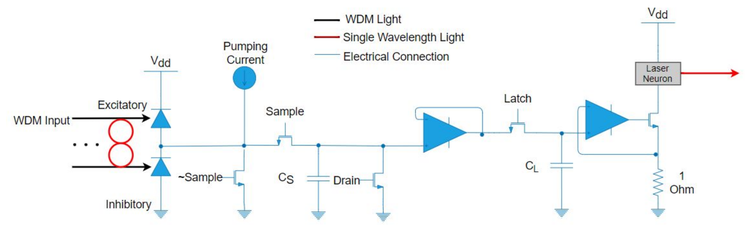


Research Goals:

- Design a circuit to replace the digital path of the infinitely scalable photonic neuromorphic model shown in [1]
- Verify functionality of the ideal model of the circuit through simulations
- Test compatibility with optical components using a low frequency discrete component model
- Examine potential limitations with respect to bandwidth and compare them to that of the existing design

Design and Methodology:

- The goal for the circuit can be simplified to saving a current and reproducing it indefinitely
- Sampling the current
  - Simple task with constant current values
  - Time dependent signals cannot be stored and reproduced
  - Time-averaging makes the most sense when planning for LIF neurons
- Saving the current
  - Capacitors can be used to store values as voltages
  - Circuit is designed for high frequency, so restorative circuitry shouldn't be necessary
  - Lower capacitances have higher bandwidths, higher capacitances are more robust to leakage current
- Reproducing current
  - Voltage to current conversion can be done with a MOSFET and a resistor to ground for a feedback voltage
  - Can represent a wide range of currents so long as the required drain voltage is significantly far from Vdd



Circuit Operation:

- After a short sample pulse τ, the voltage across C_S is defined by the equation

$$V_{CS} = \frac{1}{C_S}\int_{t_0}^{t_0+\tau} i_{in}(t)dt$$

- This equation exactly mirrors the time average integral if $\tau = C_S$
- Higher or lower capacitances can scale this value for larger or smaller expected input current ranges
- After the sample is taken, the latch signal is triggered and the value is stored in C_L
- Once $V_{CS} \approx V_{CL}$, the correct optical power is being output, the latch FET is turned off, and C_L is drained for reuse while the MRR weight bank is changed
- When $V_{CL}=0$ and the MRR weights have been successfully changed, the sampling process can begin again

Design Impact Factors:

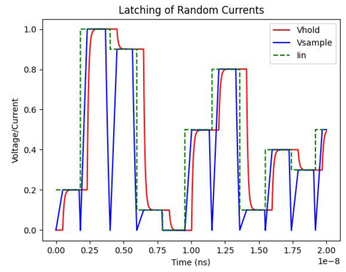
Constraints
- Should be able to handle both constant value and spiking input signals
- Must be fully parallelized for maximum data rate
- Should ideally have as small of a footprint as possible to allow for a full system-on-chip approach

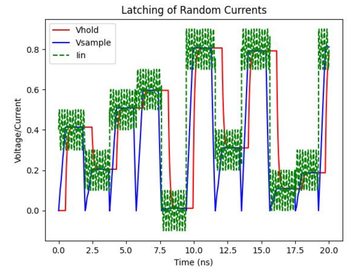
Engineering Standards
- Design should support common feed-forward network architectures like convolutional layers, pooling layers, etc.
- Optical outputs to communication fibers should have predictable wavelengths and bandwidths that adhere to FCC standards

Considerations
- Ease of use will affect consumer willingness to adopt product
- Smaller circuits with fewer transistors will reduce system cost
- Optical arithmetic units are shown to have better energy efficiency than CMOS technology
- Implementations must be dependable to be trusted for use on sensitive data

Verification and Physical Circuit:

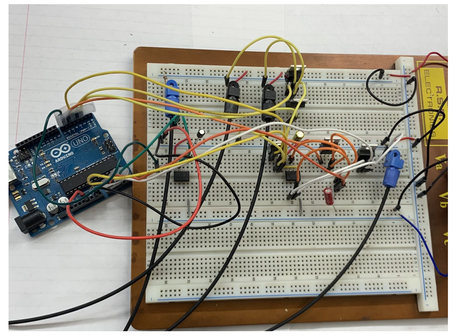
- Individual portions of the circuit were verified using Keysight's Advanced Design System circuit simulator
- These tests were used to construct the final circuit architecture
- Python mathematical model was created to more easily test the circuit's response to different signals
  - Capacitor differential equations were discretized to solve for values at successive time steps
  - Discrete input current signals were artificially generated prior to solving for circuit voltages
  - Thresholds were detected using "if" statements in the software
  - Both constant and time dependent inputs were simulated





- Maximum data rate was significantly improved when the sample capacitor was made to drain to a negative voltage rather than to 0 volts
- Current bottleneck is the latching time, which can be decreased by moving negative feedback node past the latching MOSFET to create op-amp overshoot
- Simulated results were promising enough to warrant construction of a proof-of concept circuit, intended to operate with a much slower data rate

- The model circuit was build on a breadboard using a set of photodetectors and LED's, discrete transistors and capacitors, and an Arduino Uno microprocessor for circuit monitoring and control
- The circuit is not yet in working order, as the MOSFETs being used for signal switching do not gate current bidirectionally
- Efforts have begun to change these MOSFETs to solid state switch ICs



Future Work:

- Replace signal gating MOSFETs with solid state switch ICs
- Operate and collect data from working circuit
- Analyze data to determine signal quality entering and exiting the circuit
- Use results to make predictions regarding the bandwidth of the proposed circuit in integration
- Write a conference paper on findings for submission to the Frontiers in Optics and Laser Science conference

Acknowledgements:

- This work is inspired by the work of the Princeton University Lightwave Communications Research Lab
- Primary reference:

[1] M. J. Filipovich, Z. Guo, B. A. Marquez, H. D. Morrison, and B. J. Shastri, "Training Deep Neural Networks in situ with neuromorphic photonics," IEEE Xplore, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9252268/. [Accessed: 27-Apr-2022].