2014 International Symposium on Nonlinear Theory and its Applications
NOLTA2014, Luzern, Switzerland, September 14-18, 2014

[Invited Talk]

Recent Progress Towards a Fully-Integrated Hybrid Silicon Laser Neuron Platform

Mitchell A. Nahmias†, Alex N. Tait†, Bhavin Shastri†, and Paul R. Prucnal†

†Department of Electrical Engineering, Princeton University
Princeton, NJ 08542
Email: mnahmias@princeton.edu

Abstract—We combine advances in photonic integrated circuits (PICs) with principles from neuromorphic engineering to create a scalable, robust, and extremely high bandwidth bio-inspired computational system. We describe such a system in a wafer-bonded III-V/silicon platform, integrating the network (through passive silicon-on-insulator technology) and the computational elements (through active III-V laser devices) in a single substrate, and corroborate its underlying principles through preliminary bench-top demonstrations.

1. Introduction

The ability to map a processing paradigm to its physical implementation, rather than abstracting physical effects away entirely, represents an important step in streamlining efficiency and performance. The marriage of photonics with spike processing—a computational paradigm utilized in biological neurocircuits—is fundamentally enabled by the strong analogies of the underlying physics between the dynamics of biological neurons and lasers, both of which can be understood within the framework of dynamical systems theory. Integrated photonic platforms offer an alternative approach to microelectronics. The high switching speeds, high communication bandwidth, and low cross-talk achievable in photonics are very well suited for an ultrafast spike-based information scheme with high interconnection densities. In addition, the high wall-plug efficiencies of photonic devices may allow such implementations to match or eclipse equivalent electronic systems in low energy usage. Because of these advantages, photonic spike processors could access a picosecond, low-power computationally rich domain that is inaccessible by other technologies.

The photonic spike processor is a hardware building block that, like a logic gate, enables the scalability and noise robustness necessary to construct arbitrarily complex systems, but, unlike a logic gate, uses hybrid analog-digital codes to most naturally interact with the changing environment of radio spectra. The incorporation of a novel, biologically inspired signal processing model to lightwave devices imports the potential of high-complexity processing to high-performance photonic hardware. Spike processing circuits can be implemented in conventional device fabrication processes for integrated optical interconnects, like silicon and CMOS photonics: manufacturing process with billions of dollars of commercial investment. From these devices, our group has shown that unconventional circuits sufficient for information processing can be constructed (as shown in Figure 1). Using developing platforms such as hybrid silicon/III-V PICs [1], on-chip systems with tens of thousands of reconfigurable elements will be possible, allowing the large range of possible behaviors needed for the next generation of spectrally aware RF systems.

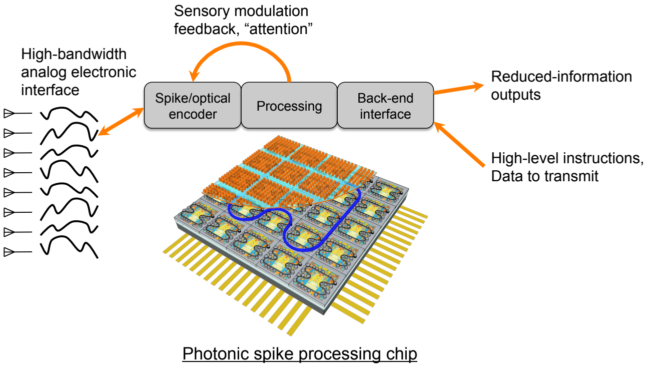
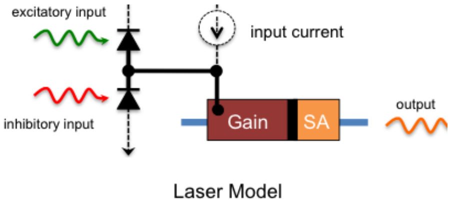


Figure 1: A diagram of a fully integrated photonic spike processor. The system could handle many fast-varying (GHz) signals simultaneously and perform complex operations such as classification, recognition, and adaptation. Such a unique processor could find use for spectrally aware RF systems or complex systems analysis.

2. Processor Node

Recent years have seen the emergence of a new class of optical devices that exploit a dynamical isomorphism between semiconductor photocarriers and neuron biophysics. The difference in physical timescales allows these photonic neurons to exhibit spiking behavior on picosecond (instead of millisecond) timescales [2]. Spiking is closely related to a dynamical system property that underlies all-or-none responses called excitability, which is shared by certain kinds of laser devices. Excitable laser systems have been studied in the context of spike processing with the tools of bifurcation theory by [3] and experimentally by [4].

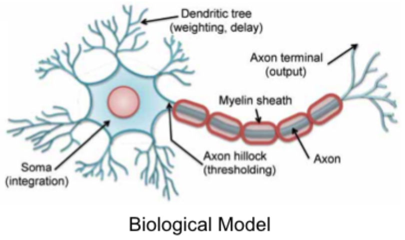


$$\frac{dG(t)}{dt} = -\gamma_G(G(t) - A) + \theta(t);$$

$$\text{if } G(t) > G_{\text{thresh}}, \text{ then}$$

$$\text{release a pulse, and set } G(t) \rightarrow G_{\text{reset}}$$

(a)



$$C_m\frac{dV_m(t)}{dt} = -\frac{1}{R_m}(V_m(t) - V_L) + I_{\text{app}}(t);$$

$$\text{if } V_m(t) > V_{\text{thresh}}, \text{ then}$$

$$\text{release a spike and set } V_m(t) \rightarrow V_{\text{reset}}.$$

(b)

Figure 2: (a) A circuit diagram of the proposed device. Excitatory and inhibitory inputs are incident on photodetectors, which represent synaptic variables. The induced photocurrent modulates carrier injection into a laser with embedded saturable absorber, which may cause the laser to transmit a pulse back into the network. (b) Analogy with a biological neuron. The soma performs integration and then applies a threshold to make a spike or no-spike decision. After a spike is released, the voltage is reset. The resulting spike is sent to other neurons in the network.

The LIF neuron model is a mathematical model of the spiking dynamics which pervade animal nervous systems, well-established as the most widely used model of biological neurons in theoretical neuroscience for studying complex computation in nervous systems. A simple model of a single-mode laser with saturable absorber (SA) section has been proven to be analogous to the equations governing an LIF neuron in certain parameter regimes (Figure 2) [2]. We have recently designed excitable lasers specifically designed for compatibility with common photonic integrated circuit (PIC) platforms [5], and tested the dynamical model in fiber-based bench-top experiments [6].

2.1. Bench-Top Model

We recently demonstrated [6] a fiber-based excitable laser as a proof of concept of excitability with an embedded SA. While the performance of this benchtop prototype is much less than that of an integrated version (bandwidth: 100 kHz, energy per spike: 10 nJ), it experimentally confirms the possibility of using laser systems to emulate the spike processing capabilities required for cortical processing: temporal integration of multiple inputs, threshold detection, and all-or-nothing pulse generation. Figure 3 is a demonstration of the system's ability to exhibit excitability when a series of spikes are incident upon it. Excitatory pulses increase the gain carrier concentration, which performs temporal integration. Enough excitation results in an excursion from equilibrium causing the laser to fire a pulse due to the saturation of the absorber to transparency. This is followed by a relative refractory period during which an excitatory pulse is unable to cause the laser to fire. The phase-space excursion resulting from an excitable response is stereotyped and repeatable while subthreshold activation results in no output: key all-or- nothing properties for pulse regeneration, reshaping, and signal integrity.

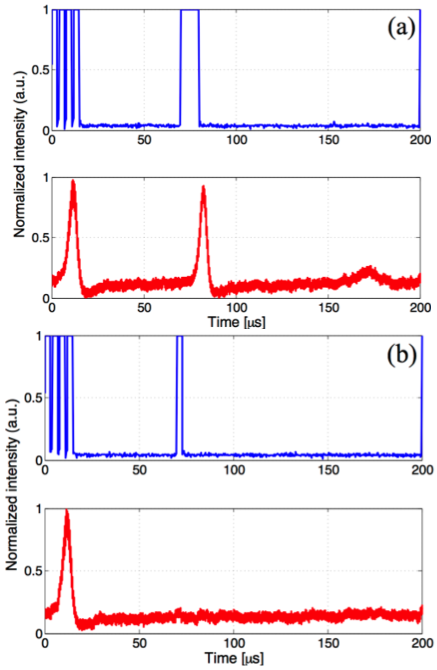


Figure 3: Experimental data of excitable fiber laser. (a) Traces showing a response to multiple integrated stimuli or one strong stimulus. (b) For too weak of an input, no change in output is detectable.

2.2. Integrated Excitable Laser

As illustrated in Figure 2, our device consists of three primary components: two photodetectors and an excitable laser. The photodetectors receive optical pulses from a network and produce a push-pull current signal which modulates the laser carrier injection. The excitable laser acts as a threshold decision maker and clean pulse generator anal-

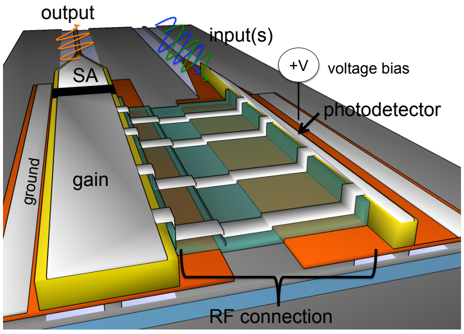


Figure 4: A schematic of an integrated excitable laser fully integrated into a hybrid silicon/III-V platform. This device can interface with a passive SOI network. Only one photodetector (PD) is shown.

ogous to the neural axon hillock. We chose to implement this model using a hybrid silicon evanescent DFB laser [5], shown in Figure 4.

Much of the energy cost of electronic conversion in optical systems comes from the need for high-speed clocked transistor circuitry and the need to demultiplex wavelength-division multiplexed (WDM) channels before conversion. In our case, electronic conversion does not have the goal of signal regeneration, but instead of exploiting electronic physics for intermediate analog processing. The use of passive integrated electrical wires does not sacrifice bandwidth or sensitivity in this device. Pulse reshaping is not required due to the clean pulse generation in the excitable laser. The conversion between optical and electronic domains also curtails the propagation of optical phase noise and the need for direct wavelength conversion, thus eliminating two major barriers facing scalable optical computing. Every device in the primary signal pathway performs both physical and computational roles, resulting in a robust, ultrafast, expressive, and extremely efficient signal processing primitive.

3. Network

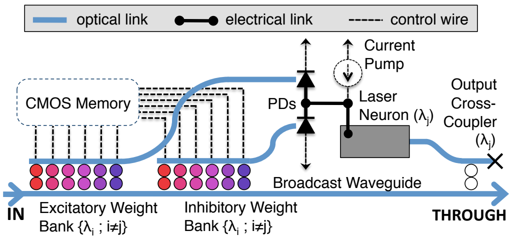


Figure 5: An example of a network node, complete with passive filters for neural weights.

The communication potential of optical interconnects (bandwidth, energy use, electrical isolation) have received attention for neural networking in the past; however, attempts to realize holographic or matrix-vector multiplication systems have encountered practical barriers, largely because they cannot be integrated. Techniques in silicon PIC fabrication is driven by a tremendous demand for optical communication links within conventional supercomputing systems [7]. The first platforms for systems integration of active photonics are becoming commercial reality [1], and promise to bring the economies of integrated circuit manufacturing to optical systems. Our work investigates the potential of modern PIC platforms to enable large-scale all-optical systems for unconventional and/or analog computing, using a standard device set designed for digital communication (waveguides, filters, detectors, etc.).

3.1. Broadcast-and-Weight

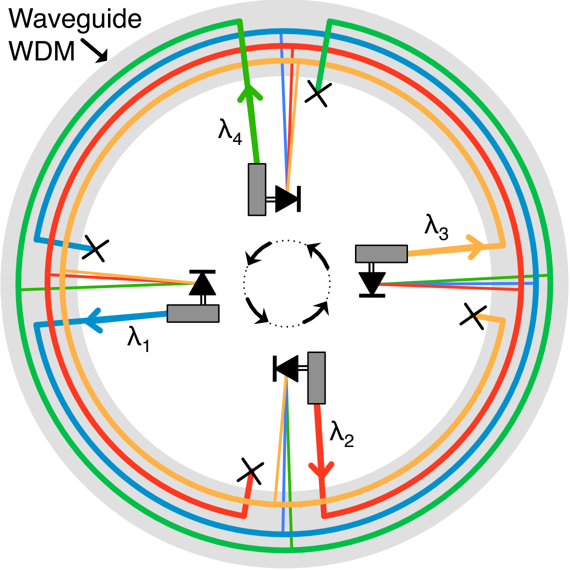


Figure 6: Diagram of a broadcast loop. Each unit has full spectrum access to the outputs of every other unit within the loop.

Our scheme leverages recent advances in PIC technology to address interconnect challenges faced by distributed processing. It has similarities with the fiber networking technique broadcast-and-select, which channelizes usable bandwidth using WDM; however, the protocol flattens the traditional layered hierarchy of optical networks, accomplishing physical, logical, and processing tasks in a compact network protocol. Although its processing circuits are unconventional, the required device set is compatible with mainstream PIC platforms. WDM effectively channelizes available bandwidth without spatial or holographic multiplexing and avoids coherent interference effects during fan-in. High-bandwidth optical channels are compatible with our proposed laser neuron devices, which could access a picosecond computational domain that impacts application areas where both complexity and speed are paramount.

In this scheme, a group of nodes (a node is illustrated in Figure 4) shares a common medium in which the output of every node is assigned a unique transmission wavelength and made available to every other node (Figure 6). Each node has a tunable spectral filter bank at its front-end. By tuning continuously between the ON/OFF resonant states, each filter drops a portion of its corresponding wavelength channel, thereby applying a coefficient of transmission analogous to a neural weight. The filters of a given receiver operate in parallel, allowing it to receive multiple inputs simultaneously. An interconnectivity pattern is determined by the local states of filters and not a state of the transmission medium between nodes. Routing in this network is transparent, massively parallel, and switchless, making it ideal to support asynchronous signals of a neural character.

3.2. Silicon Photonic Integration

Since routing is already performed by filters at the front end of each laser neuron described in the previous section, the broadcast medium must simply implement an all-to-all interconnection, supporting all N^2 potential connections between participating units. To satisfy these requirements, we use a loop-based architecture: a single integrated waveguide with topology of a loop (i.e. ring). A broadcast-and-weight cell thus consists of several laser neuron primitives coupled to a BL medium, as illustrated in Figure 6. This architecture allows each node in the network full access to signals from every other network node. Filtering at each node (Figure 5) allows neural weights to be applied.

The ability to control each connection, each weight, independently is critical for creating differentiation amongst the processing elements. A great variety of possible weight profiles allows a group of functionally similar units to compute a tremendous variety of functions despite sharing a common set of available input signals. Reconfiguration of the filters' drop states, corresponding to weight adaptation or learning, intentionally occurs on timescales much slower than spike signaling. Reconfigurable filters can be implemented by a micro-ring resonator whose resonance is tuned thermally or electronically. The broadcast medium could be a silicon-on-insulator (SOI) waveguide, which is fully compatible with the laser neuron structure described in the previous section.

4. Conclusion

We have described a potential platform—enabled by unique optoelectronic physics and the recent emergence of scalable photonic integrated circuits (PICs)—that emulates biological networks of neurons at ultrafast speeds. Silicon photonic platform development has revolved around point-to-point links for multi-core computing systems. We have examined an opportunity for this technology to extend to unconventional architectures that rely heavily on interconnect performance. Broadcast-and-weight is a new approach for joining neuron-inspired processing and optical interconnect physics. The LIF model with a synaptic variable, coupled with tunable routing in a passive SOI network on a scalable platform, could open computational domains that demand unprecedented temporal precision, power efficiency, and functional complexity. Its enormous bandwidth capabilities would allow for a system to emulate cortical functions at the time scale of radio frequency (RF) waves, creating a system fully cognizant of all aspects of the RF spectrum in real-time. In the long term, the platform could provide wide-band, reconfigurable, and robust communication with flexible spectrum access in the RF spectrum on a single chip.

Acknowledgments

The authors would like to thank Yue Tian, Mable Fok, and David Rosenbluth for their fruitful suggestions and support.

References

[1] Di Liang and John E Bowers. Recent progress in lasers on silicon. *Nature Photonics*, 4(8):511–517, 2010.

[2] Mitchell A Nahmias, Bhavin J Shastri, Alexander N Tait, and Paul R Prucnal. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. 19(5):1800212, Sep./Oct. 2013.

[3] B. Krauskopf, K. Schneider, J. Sieber, S. Wieczorek, and M. Wolfrum. Excitability and self-pulsations near homoclinic bifurcations in semiconductor laser systems. *Opt. Commun.*, 215(4-6):367–379, January 2003.

[4] S. Barbay, R. Kuszelewicz, and A. M. Yacomotti. Excitability in a semiconductor laser with saturable absorber. *Opt. Lett.*, 36(23):4476–4478, December 2011.

[5] M. A. Nahmias, A. N. Tait, B. J. Shastri, and P. R. Prucnal. An evanescent hybrid silicon laser neuron. In *Proc. IEEE Photonics Conf. (IPC)*, pages 93–44, Seattle, WA, USA, September 2013.

[6] B. J. Shastri, M. A. Nahmias, A. N. Tait, Y. Tian, B. Wu, and P. R. Prucnal. Graphene excitable laser for photonic spike processing. In *Proc. IEEE Photonics Conf. (IPC)*, Seattle, WA, USA, September 2013.

[7] Bahram Jalali and Sasan Fathpour. Silicon photonics. *Lightwave Technology, Journal of*, 24(12):4600–4615, 2006.