Physical Modeling of Photonic Neural Networks

Thomas Ferreira de Lima, Bhavin J. Shastri, Mitchell A. Nahmias, Alexander N. Tait and Paul R. Prucnal
Lightwave Communications Laboratory, Department of Electrical Engineering
Princeton University, Princeton, NJ, 08544 USA, email: tlima@princeton.edu

***Abstract*—Brain-inspired distributed computing has attracted attention for its energy-efficient processing, and photonic neuromorphic hardware can overcome latency vs. fan-in tradeoffs from neuromorphic electronics. Here, we introduce system-level, physically detailed modeling tools for photonic neural networks, and use them to study the behavior of attractor networks.**

Neuromorphic computing systems are currently being developed with the goal of closing the gap between the processing efficiency of a computer chip to that of the human brain. The TrueNorth chip has achieved remarkable efficiency of $26\,\mathrm{pJ/MAC}$ [1][1]. These chips can be applied to large-scale neuroscience research and robot control, with inter-spike intervals comparable to biology ($\sim 10\,\mathrm{Hz}$). On the other hand, there is potential to brain-inspired computing for applications where a decision has to be made in nanoseconds. That would require inter-spike intervals of $\sim$GHz, but digital electronic approaches pose fundamental tradeoffs between latency and interconnect fan-in.

Photonics provides an alternative way to create a passive, non-blocking, and asynchronous network architecture dubbed broadcast-and-weight (B&W) via wavelength-division multiplexing (WDM) [2]. B&W could result in photonic neural networks with similar energy efficiency as the TrueNorth chip, but eight orders of magnitude faster.

Several laser designs exhibiting spiking behavior have been proposed and extensively modeled. Spike processing is the third and most recent generation of artificial neural models [3]. For example, a distributed-feedback (DFB) laser with saturable absorber using a III-V multi quantum well (MQW) stack [4] has been shown to emulate leaky integrate-and-fire (LIF) neural dynamics. As fabrication processes work toward integrating III-V layers with SOI [5], this platform offer monolithic compatibility with CMOS processors and controller circuits could enable fully-packaged devices with electrical I/O.

The DFB laser is pumped by a DC current bias at the gain section—with sufficient excitation, the laser enters a passive Q-switching regime, emitting a series of pulses similar to a leaky-integrate-and-fire (LIF) neuron model (Fig. 1). Close to its Q-switching threshold, this type of laser becomes *excitable*, capable of emitting a single, controlled ultrashort (ps) optical pulse. Then, optical pulses transferred to a broadcast waveguide can cause the excitation in a subsequent PNN, releasing another pulse and therefore propagating the original signal indefinitely.

A spiking laser neuron, together with a microring resonator (MRR) weight bank, a germanium PIN balanced photodetector

[1]MAC: multiply-and-accumulate operation

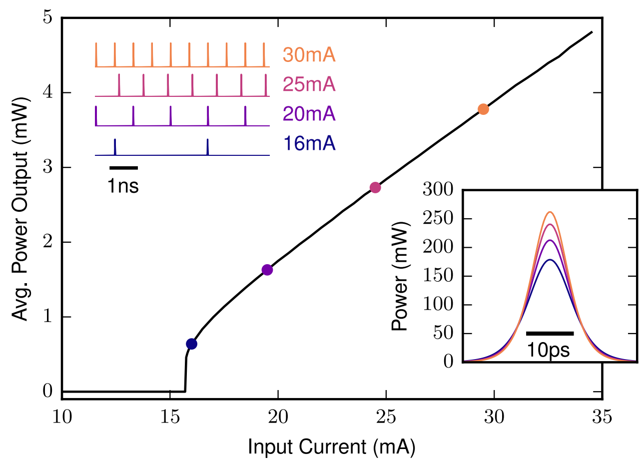

Fig. 1. Operation characteristics of a PNN with constant input current bias. The parameters are as in [6]. One can modulate the repetition rate of the spike train with the input current. The average optical power output follows a characteristic LIF-curve transfer function. The top-left inset shows the output power waveforms for four different operation conditions highlighted in the graph. The bottom-right inset shows the slight variation in the pulse profile for the same operation points.

(GePD) and a CMOS current-amplifying stage, constitute a *processing-network node* (PNN) that can be interconnected with a B&W network (Fig. 2). Despite the surge in interest in modeling spiking laser neurons, considerably little attention has been paid to network and system modeling.

In this abstract, we report a novel physical modeling framework for photonic spike processing networks and present preliminary studies of system cascadability and small-scale circuits.

Cascadability is assessed by examining the effect of input current bias on output spike rate and power in Fig. 1. Cascadability requires that one PNN be strong enough to drive at least N_{FO} other PNNs, where FO stands for *fan-out*. Above $20\,\mathrm{mA}$, the average output power increases linearly with the input current, with a slope efficiency of $0.21\,\mathrm{W/A}$. Considering the responsivity of a GePD to be $0.74\,\mathrm{A/W}$ [7] and further losses in the silicon waveguide, this corresponds to a propagation power penalty of at least $8.1\,\mathrm{dB}$ per cascaded stage. Therefore, the PNN needs a gain of at most $8.1\,\mathrm{dB} + 10\log_{10}(N_{\mathrm{FO}})$ between its input and output. To provide the extra gain stage, a CMOS current amplifier with variable gain is included between the balanced PD stage and the DFB laser, providing an extra control variable.

The simulations shown in this abstract involved the develop-

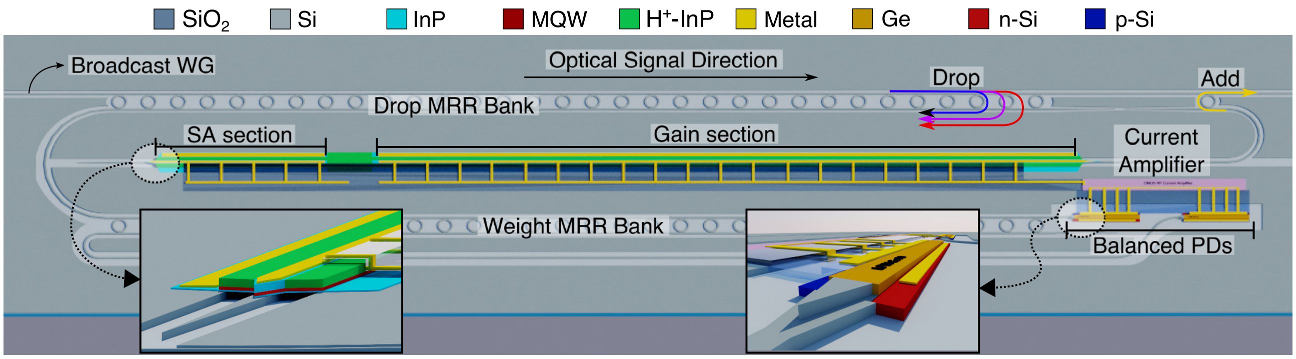

Fig. 2. A three-dimensional sketch of a Hybrid III-V/Si spiking PNN. Two MRR weight banks cascaded in series provide complementary weighted signals to a balanced Germanium Photodetector, which incoherently sums all channels while generating photocurrent (post-synaptic weighted addition). This photocurrent is fed into an RF current amplifier, which can be fabricated using CMOS processes. The amplified current is connected to the gain section of the laser. The laser output is added to the broadcast waveguide by a resonant MRR.

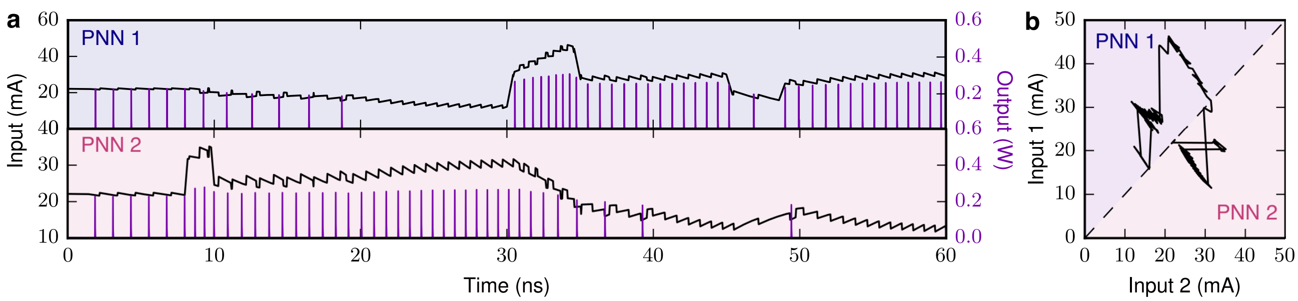

Fig. 3. Operation of two PNNs connected as a 2-neuron Winner-Takes-All Hopfield network. The current bias input to each neuron is $20\,\mathrm{mA}$. The connection weight matrix for this particular example is $W = \begin{bmatrix} 1 & -.7 \\ -.7 & 1 \end{bmatrix}$. The post-synaptic inputs to each neuron is amplified by a factor of 3 and smoothed by passive low-pass filter with RC time constant of $3\,\mathrm{ns}$, to prevent spike phase-locking.

ment of an optimized circuit-model dubbed SIMPEL [6]. We extended the simulator to support the current amplifying stage inside the PNN as well as the physics of fully-interconnected B&W networks. Since spiking behavior can be found in many other two-section laser configurations and are usually modeled by a dynamical system, SIMPEL will be made available shortly under GNU GPLv3 license [8]. SIMPEL can be used to realistically simulate small networks ($N < 100$) of PNNs representing any kind of dynamical system, and should be widely useful to researchers designing brain-inspired optoelectronic systems.

As a proof-of-concept, we show the simulated operation of a 2-PNN network connected in a Winner-Takes-All (WTA) configuration (Fig. 3). WTA is a classic example of lateral inhibition and competitive attractor dynamics in neural networks. The PNNs mutually inhibit each other while simultaneously activating themselves through self-connections. As a result, the only stable states of the network are the ones in which one neuron is firing and the other is not. More broadly, attractor networks could also be understood as a unit of a working memory. Hopfield networks with content-addressable memory could be instantiated in a similar way.

Fig. 3 shows the response of such a network to several external stimuli. Starting in a balanced equilibrium, a $2\,\mathrm{ns}$ stimulus is presented to PNN 2 (~$8\,\mathrm{ns}$ on x-axis). The network then converges to favor PNN 2 after $15\,\mathrm{ns}$. A $5\,\mathrm{ns}$ stimulus to PNN 1 (~$30\,\mathrm{ns}$ on x-axis) switches the system to the opposite attractor state. B&W networks of PNNs are well-suited to support ultra-low latency and high fan-in photonic neural processing networks on-chip, and system-level simulation tools will be an enabling component of their development.

References

[1] P. A. Merolla *et al.*, *Science*, vol. 345, no. 6197, pp. 668–673, 2014.
[2] A. Tait *et al.*, *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 4029–4041, 2014.
[3] W. Maass, *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.
[4] M. A. Nahmias *et al.*, *Optics Express*, vol. 23, no. 20, p. 26800, oct 2015.
[5] M. J. R. Heck *et al.*, *Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 4, p. 6100117, 2013.
[6] B. J. Shastri *et al.*, *Optics Express*, vol. 23, no. 6, p. 8029, 2015.
[7] H. Chen *et al.*, *Optics Express*, vol. 24, no. 5, p. 4622, mar 2016.
[8] "SIMPEL Project," https://github.com/lightwave-lab/SIMPEL, Accessed: 2015-04-16.