Excitable laser processing network node in hybrid silicon: analysis and simulation

Mitchell A. Nahmias,* Alexander N. Tait, Bhavin J. Shastri, Thomas Ferreira de Lima, and Paul R. Prucnal

Electrical Engineering Dept., Princeton University, Princeton, NJ 08544, USA.

**mnahmias@princeton.edu*

Abstract: The combination of ultrafast laser dynamics and dense on-chip multiwavelength networking could potentially address new domains of real-time signal processing that require both speed and complexity. We present a physically realistic optoelectronic simulation model of a circuit for dynamical laser neural networks and verify its behavior. We describe the physics, dynamics, and parasitics of one network node, which includes a bank of filters, a photodetector, and excitable laser. This unconventional circuit exhibits both cascadability and fan-in, critical properties for the large-scale networking of information processors based on laser excitability. In addition, it can be instantiated on a photonic integrated circuit platform and requires no off-chip optical I/O. Our proposed processing system could find use in emerging applications, including cognitive radio and low-latency control.

OCIS codes: (140.3538) Lasers, pulsed; (200.4700) Optical neural systems; (250.3140) Integrated optoelectronic circuits; (320.7085) Ultrafast information processing.

References and links

1. L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," Nat. Commun. **2**, 468 (2011).
2. D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," Nat. Commun. **4**, 1364 (2013).
3. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutiérrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing," Opt. Express **20**, 3241–3249 (2012).
4. F. Rogister, A. Locquet, D. Pieroux, M. Sciamanna, O. Deparis, P. Mégret, and M. Blondel, "Secure communication scheme using chaotic laser diodes subject to incoherent optical feedback and incoherent optical injection," Opt. Lett. **26**, 1486–1488 (2001).
5. C. Juang, T. Hwang, J. Juang, and W.-W. Lin, "A synchronization scheme using self-pulsating laser diodes in optical chaotic communication," IEEE J. Quantum Electron. **36**, 300–304 (2000).
6. K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," Nat. Commun. **5**, 3541 (2014).
7. A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, "Network of time-multiplexed optical parametric oscillators as a coherent ising machine," Nat. Photonics **8**, 937–942 (2014).
8. 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**, 367–379 (2003).
9. J. L. A. Dubbeldam, B. Krauskopf, and D. Lenstra, "Excitability and coherence resonance in lasers with saturable absorber," Phys. Rev. E **60**, 6580–6588 (1999).
10. M. Turconi, B. Garbin, M. Feyereisen, M. Giudici, and S. Barland, "Control of excitable pulses in an injection-locked semiconductor laser," Phys. Rev. E **88**, 022923 (2013).

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

11. S. Beri, L. Mashall, L. Gelens, G. Van der Sande, G. Mezosi, M. Sorel, J. Danckaert, and G. Verschaffelt, "Excitability in optical systems close to z2-symmetry," Phys. Lett. A **374**, 739–743 (2010).
12. M. A. Nahmias, B. J. Shastri, A. N. Tait, and P. R. Prucnal, "A leaky integrate-and-fire laser neuron for ultrafast cognitive computing," IEEE J. Sel. Top. Quantum Electron. **19**, 1–12 (2013).
13. M. A. Nahmias, A. N. Tait, B. J. Shastri, and P. R. Prucnal, "An evanescent hybrid silicon laser neuron," in "Photonics Conference (IPC)" (IEEE, 2013), pp. 93–94.
14. A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," J. Lightwave Technol. **32**, 3427–3439 (2014).
15. M. A. Nahmias, A. N. Tait, B. J. Shastri, , and P. R. Prucnal, "A receiver-less link for excitable laser neurons: Design and simulation," in *Summer Topicals* (IEEE/OSA, 2015), pp. 99–100 (2015).
16. A. W. Fang, H. Park, Y.-h. Kuo, R. Jones, O. Cohen, D. Liang, O. Raday, M. N. Sysak, M. J. Paniccia, and J. E. Bowers, "Hybrid silicon evanescent devices," Mater. Today **10**, 28–35 (2007).
17. P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," Science **345**, 668–673 (2014).
18. K. Boahen, "Point-to-point connectivity between neuromorphic chips using address events," IEEE Trans. Circ. Syst. **47**, 416–434 (2000).
19. J. Schemmel, J. Fieres, and K. Meier, "Wafer-scale integration of analog neural networks," in *IEEE International Joint Conference on Neural Networks* (IJCNN, 2008), pp. 431–438.
20. D. A. Miller, "Rationale and challenges for optical interconnects to electronic chips," Proc. IEEE **88**, 728–749 (2000).
21. T. Sorrentino, C. Quintero-Quiroz, A. Aragoneses, M. C. Torrent, and C. Masoller, "Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback," Opt. Express **23**, 5571–5581 (2015).
22. F. Selmi, R. Braive, G. Beaudoin, I. Sagnes, R. Kuszelewicz, and S. Barbay, "Relative refractory period in an excitable semiconductor laser," Phys. Rev. Lett. **112**, 183902 (2014).
23. T. V. Vaerenbergh, K. Alexander, J. Dambre, and P. Bienstman, "Excitation transfer between optically injected microdisk lasers," Opt. Express **21**, 28922–28932 (2013).
24. W. Coomans, L. Gelens, S. Beri, J. Danckaert, and G. Van der Sande, "Solitary and coupled semiconductor ring lasers as optical spiking neurons," Phys. Rev. E **84**, 036209 (2011).
25. B. Romeira, J. Javaloyes, C. N. Ironside, J. M. L. Figueiredo, S. Balle, and O. Piro, "Excitability and optical pulse generation in semiconductor lasers driven by resonant tunneling diode photo-detectors," Opt. Express **21**, 20931–20940 (2013).
26. B. J. Shastri, M. A. Nahmias, A. N. Tait, A. W. Rodriguez, B. Wu, and P. R. Prucnal, "Dynamical laser spike processing," arXiv:1507.06713 (2015).
27. J. W. Goodman, "Fan-in and fan-out with optical interconnections," Opt. Acta **32**, 1489–1496 (1985).
28. A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, "Photonic neuromorphic signal processing and computing," in *Nanophotonic Information Physics,* , M. Naruse, ed. (Springer Berlin Heidelberg, 2014), pp. 183–222.
29. A. Fang, M. Sysak, B. Koch, R. Jones, E. Lively, Y. hao Kuo, D. Liang, O. Raday, and J. Bowers, "Single-wavelength silicon evanescent lasers," IEEE J. Sel. Top. Quantum Electron. **15**, 535–544 (2009).
30. A. Tait, M. Nahmias, T. Ferreira de Lima, B. Shastri, A. Wu, E. Zhou, E. Blow, and P. Prucnal, "Continuous control of microring weight banks," in "Proc. IEEE Photonics Conf. (IPC)," (IEEE, 2015).
31. H. Park, A. W. Fang, R. Jones, O. Cohen, O. Raday, M. N. Sysak, M. J. Paniccia, and J. E. Bowers, "A hybrid algainas-silicon evanescent waveguide photodetector," Opt. Express **15**, 6044–6052 (2007).
32. K. Preston, N. Sherwood-Droz, J. Levy, and M. Lipson, "Performance guidelines for wdm interconnects based on silicon microring resonators," in "2011 Conference on Lasers and Electro-Optics (CLEO)" (OSA, 2011), pp. 1–2.
33. D. Liang, G. Roelkens, R. Baets, and J. E. Bowers, "Hybrid integrated platforms for silicon photonics," Materials **3**, 1782–1802 (2010).
34. C. Zhang, S. Srinivasan, Y. Tang, M. J. R. Heck, M. L. Davenport, and J. E. Bowers, "Low threshold and high speed short cavity distributed feedback hybrid silicon lasers," Opt. Express **22**, 10202–10209 (2014).
35. A. Destexhe, Z. F. Mainen, and T. J. Sejnowski, "Kinetic models of synaptic transmission," Methods Neuro. Modeling **2**, 1–25 (1998).
36. W. Maass, "Networks of spiking neurons: The third generation of neural network models," Neural Networks **10**, 1659 – 1671 (1997).
37. S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," Phys. Rev. Lett. **80**, 197–200 (1998).
38. B. J. Shastri, M. A. Nahmias, A. N. Tait, B. Wu, and P. R. Prucnal, "Simpel: Circuit model for photonic spike processing laser neurons," Opt. Express. **23**, 8029–0844 (2015).
39. D. A. B. Miller, "Are optical transistors the logical next step?" Nat. Photonics **4**, 3–5 (2010).
40. S. Le Beux, J. Trajkovic, I. O'Connor, G. Nicolescu, G. Bois, and P. Paulin, "Optical ring network-on-chip

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

(ORNoC): Architecture and design methodology," in "Design, Automation Test in Europe Conference Exhibition (DATE), 2011," (2011), pp. 1–6.
41. D. G. Feitelson, *Optical Computing: A Survey for Computer Scientists* (MIT Press, USA, 1988).

1. Introduction

Recently, there has been a surge of interest in the hybridization of photonic and electronic physics to achieve unique processing capabilities. In this context, there has been significant research in using laser dynamics for both information processing [1–3] and communication [4,5]. Multiplexing in optical networks have also been exploited for similar gains in performance [6, 7]. A more specific approach involves studying the dynamical property of excitability in lasers, in which discrete pulses are generated in response to perturbations that exceed a threshold [8–11]. Excitable lasers can process information in a way that resembles spiking in biological neuron models. In comparison, however, lasers can exhibit dynamics roughly eight orders of magnitude faster while being mathematically isomorphic to their biological counterparts [12].

In this manuscript, we simulate a unified, realistic model of a reconfigurable excitable laser processor, recently proposed in [13]. This model includes a detailed analysis and simulation of a processing-network node (PNN) as defined in a recent laser networking scheme [14]. Although parts of the signal pathway have been proposed [13–15], no single model has fully characterized the signal pathway of a PNN. The device uniquely allows for a large fan-in ($\sim$10s to 100s per unit) without routing or packet switching, and utilizes the ultrafast dynamics of lasers for high bandwidth ($\sim$GHz) processing. We describe the photonic circuit techniques used in our approach and simulate the device based on experimentally measured parameters in a standard hybrid III-V/silicon platform [16]. The model allows us to explore critical properties such as energy consumption, cascadability, and signal bandwidth, and verifies that the PNN can be fabricated in a realistic device structure.

Spiking neural networks (SNNs)—systems whose communication channels code information in events rather than bits—have received significant attention as an alternative to the von Neumann paradigm of computation. Hardware SNNs take advantage of distributed, sparse, and robust coding schemes to perform computations to minimize size, weight and power, and have been utilized both as biological network simulators and low-power data processors.

Interconnects in neural network architectures require many-to-one fan-in and multicast communication, as opposed to von Neumann processors, where a number of point-to-point memory-processor communication bottlenecks arise [17]. In distributed systems, processing performance is more closely related to inter-node communication. Distributed electrical wires can introduce performance challenges in a multicast network due to RC and radiative physics, in addition to the typical bandwidth-distance-energy limits of point-to-point electronic links. Since biological-time SNNs typically operate far slower than underlying electronics, many large-scale systems employ a form of time-division multiplexing (TDM) or packet switching, notably, address event representation (AER). TDM and AER abstractions allow networks to have virtual interconnectivity densities that exceed the wire density by a factor related to the sacrificed bandwidth, which can be orders of magnitude [18].

While AER is effective for SNNs running at biological time scales, networks running at faster-than-biological time scales experience new challenges and limitations. The fastest large-scale SNN emulator is capable of an approximately 10 MHz synaptic simulation (10^4 speed-up over biological-time) [19]. In the GHz regime, the gap between this operation bandwidth and feasible real signal bandwidths shrinks, which creates a harsh tradeoff between operation bandwidth and virtual interconnectivity. Neuron emulators relying on electrical wires cannot, in the worst case, simultaneously cope with virtual interconnectivity and GHz signal bandwidths. This is a significant limitation for neuromorphic networks, which rely on high interconnectivity

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

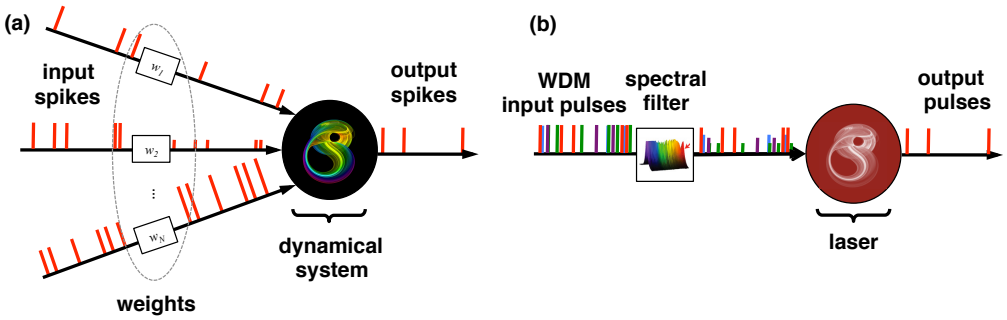

Fig. 1. (a) A general depiction of a processor in a SNN. Input channels are weighted by w_i, where they are summed together into a dynamical system (the neuron), which outputs its own set of spikes. (b) Concept of a photonic spike processor. Inputs are wavelength division multiplexed (WDM) channels on a single input channel. The spectral filter provides the weighted sum of inputs, which are summed together and drive a single excitable laser. The laser emits its a set of its own pulses into the network.

densities for complex, parallelized operations. In contrast, it is well known that photonic technology possesses favorable properties for communication and can come to bear on problems in on-chip interconnect performance [20].

The research community has proposed a number of dynamical laser models [12, 21–26]. Unfortunately, many of these models suffer from network or interconnection limitations. Interconnected systems require each node to robustly drive a number of other nodes—referred to as cascadability and fan-in (Fig. 1). If inputs are optically injected into the gain section, the pump wavelength must be shorter than that of the signal for population inversion to be achieved. As a result, network feedback requires wavelength conversion, a fairly expensive operation in terms of size, weight, and power. Alternatively, pumping the intensity directly requires a match with the cavity resonance, limiting inputs and outputs to the same wavelength. Optical fan-in with a single wavelength requires coherent optical summing to avoid beat noise, a practical impossibility in unsynchronized laser systems [27]. These complications would offset many of the advantages of using optical physics over more traditional RF electronic approaches.

To address the issues of cascadability and fan-in for networks of ultrafast spiking laser neurons, a scalable networking architecture was recently proposed [28]. High interconnect density is achieved through the use of wavelength division multiplexing (WDM) signals fanning-in to a single photodetector, which directly drives an electronically-modulated laser neuron (see Fig. 1). In this setting, electronic physics act as a powerful complement to optics, which, as a result of the lack of charge and bosonic properties of photons, are unwieldy for cascadable computation on their own. O/E/O conversion is useful in many contexts, although it is conventionally associated with a high performance cost to demultiplex, digitize, and remodulate signals. In the neural networking context, the electronic link between a photodetector-laser pair can make a cascadable node without necessarily performing these costly functions (see Fig. 2).

This work simulates a unified and realistic model, verifying that fabricating a PNN is within today's technological capabilities in a hybrid silicon/III-V platform. This is an important stepping-stone towards larger scale simulations and the verification of these concepts in an experimental device. Application of the technology, once developed, would enable a new processing domains, including the manipulation of ultrafast physical phenomena, low-latency control, and—perhaps most pertinently—novel communication and classification schemes in the radio frequency (RF) domain. Examples of the latter include blind source separation, RF fingerprinting, spectral hole exploitation (i.e. non-interfering transmission and reception), and

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

novel information encoding/retrieval through the direct modulation of carrier waveforms.

2. Methods

As illustrated in Fig. 2, the PNN consists of three primary components: reconfigurable spectral filters, photodetectors, and an excitable laser. Although an inhibitory photodetector is shown in the schematic diagram, the behavior of the devices are similar, aside from a reversal of the ground and signal pads. In this analysis, we only examine the excitatory photodetector pathway for brevity and conciseness. In this scheme, WDM spike signals arrive along a dedicated waveguide. Weights are applied to each channel via a set of tunable spectral filters.

Inputs from other laser neurons are weighted in the optical domain before reaching the photodetector. The photodetectors produce a photocurrent summing the total optical power. Demultiplexing many input channels is not necessary because the incoherent sum of all WDM channels is intentionally computed by the photodetector. The photodetectors receive optical pulses from a network and produce a current signal which modulates the laser carrier injection. The excitable laser performs nonlinear discrimination and regenerates the pulsed signal, analogous to the neural axon hillock. The photodetector front-end proposed here allows for significant signal fan-in, while tunable filters allow adjustments of the weights between neurons, allowing for network reconfigurability.

2.1. Technology Platform

We describe an instantiation of the PNN in the hybrid/III-V platform. This platform includes III-V materials that are bonded to underlying passive silicon photonic interconnection networks. In a typical device, optical modes are hybridized between the silicon and III-V layers simultaneously [29]. Nanostructures necessary for waveguides, resonators, and gratings are fabricated strictly in silicon, while the III-V layers provide optical gain. This platform was chosen for its convenient ability to spatially organize many active elements together onto a single PIC. The analysis below uses realistic parameters derived from experimental papers, and is divided into three primary sections: (a) filtering (i.e. neural weighting) of input WDM-pulsed signals, (b) summation and electrical conversion from photodetector to the adjacent laser, and (c) dynamics present within the laser cavity itself. A schematic of the full modeling structure is shown in Fig. 3.

2.2. Filter Bank

In the proposed WDM network configuration, a large number of processing nodes share a single waveguide and selectively couple light in and out through a bank of adjacent add-drop filters. The pulses coupled from the broadcast waveguide are emitted from various lasers at different wavelengths, and have approximately equal amplitudes. The filter are tuned on and off resonance to control the weight or strength of connection between each element. The result is a series of pulses at different frequencies with an array of different amplitudes [30].

In a given broadcast ring with N nodes, N filters are associated with each node to adjust the strength of a total of N^2 connections. In this analysis, we include only four in a given node for simplicity. We model the transmission function of a resonator drop filter with a Lorentzian function of the form:

$$T(\delta) = \frac{1}{1+\delta^2}, \text{ where } \delta = \frac{Q}{\omega_0}(\omega - \omega_0) \tag{1}$$

where δ is the linewidth-normalized frequency, Q is quality factor ($Q \approx 10,300$), and ω_0 is the peak center frequency. Let δ_{tun} represent the tunability of each filter, and $\Delta\delta$ the spacing between adjacent filters. We assume each tunable filter is composed of a single add-drop ring,

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

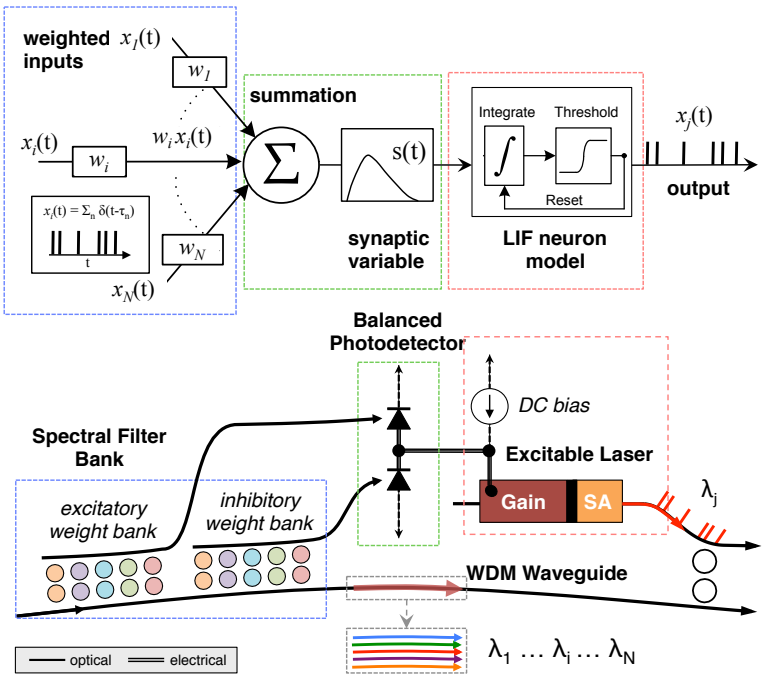

Fig. 2. (Top) A depiction of an LIF neuron with a synaptic variable, embedded within a network. (Bottom) A schematic of the proposed laser neuron, complete with filters, balanced photodetectors, and an excitable laser. In this model, only the excitatory photodetector pathway is investigated.

and choose $\delta_{tun} = 4.4$ (0.66 nm) and $\Delta\delta = 8.8$ (1.3 nm) to achieve tolerable levels of the extinction ratio, crosstalk and insertion loss [14]. A filter bank will receive a series of pulses from other lasers, each governed by some output power function $P_{out[j]}(t)$ for neuron j at an associated wavelength λ_j.

There is also a linewidth associated with the output of each laser. Single mode distributed feedback (DFB) lasers, the structures assumed here, tend to have narrow linewidths. Nonetheless, since the output is really a time dependent signal $P_{out[j]}(t)$, the linewidth is fundamentally broadened by modulation, and in addition, can be further broadened by non-idealities (such as carrier-induced chirp). Nonetheless, we will first consider an ideal case in which bandwidth-limited pulses are assumed. This allows us to approximate the output power spectrum as $S_{out[i]}(\lambda,t) = P_{out[j]}(t)\delta(\lambda-\lambda_i)$ for each laser i in a given loop where $\delta(\lambda)$ is a Dirac delta function centered at λ_i. Then the broadcast loop signal $B(\lambda,t)$ contains a sum of all output power spectra, i.e.

$$B(\lambda,t) = \sum_i S_{out[i]}(\lambda,t) \tag{2}$$

Each filter i in a given filter bank is associated with a wavelength λ_i corresponding to a laser i within the broadcast space. Although there will be some crosstalk due to the Lorentzian shape of the filter, we have minimized crosstalk by design (< 13 dB), so we can neglect its effect to first order. The power spectrum of the signal dropped from the broadcast waveguide for neuron

j will take the form:

$$S_{in[j]}(\lambda,t) = \sum_i T_{ij}(\lambda) S_{out[i]}(\lambda,t) \tag{3}$$

where $T_{ij}(\lambda)$ is the transmission function of ring i for neuron j. Therefore, for a given photodetector spectral responsivity linear system R_λ, the current produced by the photodetector is as follows:

$$i_{p[j]}(t) = \sum_i \int_\lambda R_\lambda \left\{ T_{ij}(\lambda) S_{out[i]}(\lambda,t) \right\} d\lambda \tag{4}$$

Based on our assumptions, we can approximate the filter linewidth as significantly larger than the signal bandwidth.The 40 ps $sech^2$ pulses used in this simulation have transform-limited bandwidth of $\Delta\delta = .42$ (.063 nm), which is much smaller than the bandwidth of the filters (as shown in Fig. 4(b)), allowing this approximation to be valid. We assume the photodetector response is approximately spectrally flat over the C-band [31], and neglect carrier diffusion limitations to frequency response, which are small compared to the RC limitations included in this model. This allows us to approximate the responsivity as a constant $R_{\lambda_i}\{P(t)\} = R_{PD}P(t)$. Therefore, we can simplify the photodiode current to the following expression:

$$i_{p[j]}(t) \approx \sum_i R_{PD} T_{ij}(\lambda_i) P_{out[i]}(t) \tag{5}$$

A simulation of a filter bank with four filters is shown in Fig. 4. Pulses of different frequencies experience different weights—or amplitude shifts—based on the spectral position of each tunable ring. Based on the parameters used here, there can be a total of $N = 34$ filters per node within 45 nm gain band in hybrid III-V/Si. This number is synonymous with the upper bound on the fan-in per neuron. The networks themselves can be designed to have much larger numbers of neurons by using multi-loop strategies that can sidestep fan-in limitations [28]. In addition, the upper limit can be expanded: 62 channels have been demonstrated simultaneously in an experimental silicon photonics platform, a number that can be extended further using more advanced techniques [32]. Nonetheless, fan-in puts an upper limit on the complexity of the computation that a single unit can perform. The scalability of neural fan-in must be balanced with respect to spectral stability of the lasers, the linewidth of the pulses and the Q factor of the filters (this point is expanded in the Discussion section).

2.3. Electronic Junction

The analysis of the electrical junction parasitics describes the signal as it is received from the photodetector and drives the adjacent excitable laser. It is converted from a series of pulses at different wavelengths to a single current signal (Fig. 4). Although current diffusion, saturation, and other nonidealities other can also play a role in the limiting the frequency response of the device, RC charging times tend to dominate [31]. We model the electronic junction as the lumped circuit (Fig. 3) using parasitic values that were determined in several experimental papers [29, 33, 34].

In this analysis, we only consider the bandwidth of a single photodetector driving a laser for simplicity, although multiple photodetectors may also be used for push-pull excitatory and inhibitory inputs, if desired. Multiple pulses of different wavelengths are incident on the photodetector, which provides the role of summation and converts signals to current pulses. A schematic of the junction is shown in Fig. 2, including a superimposed image of the parasitic circuit model. The purpose of the model is to find the relationship between the photodetector current $i_p(t)$ and laser driving current $i_e(t)$:

$$i_e(t) = F\{i_p(t)\} \tag{6}$$

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

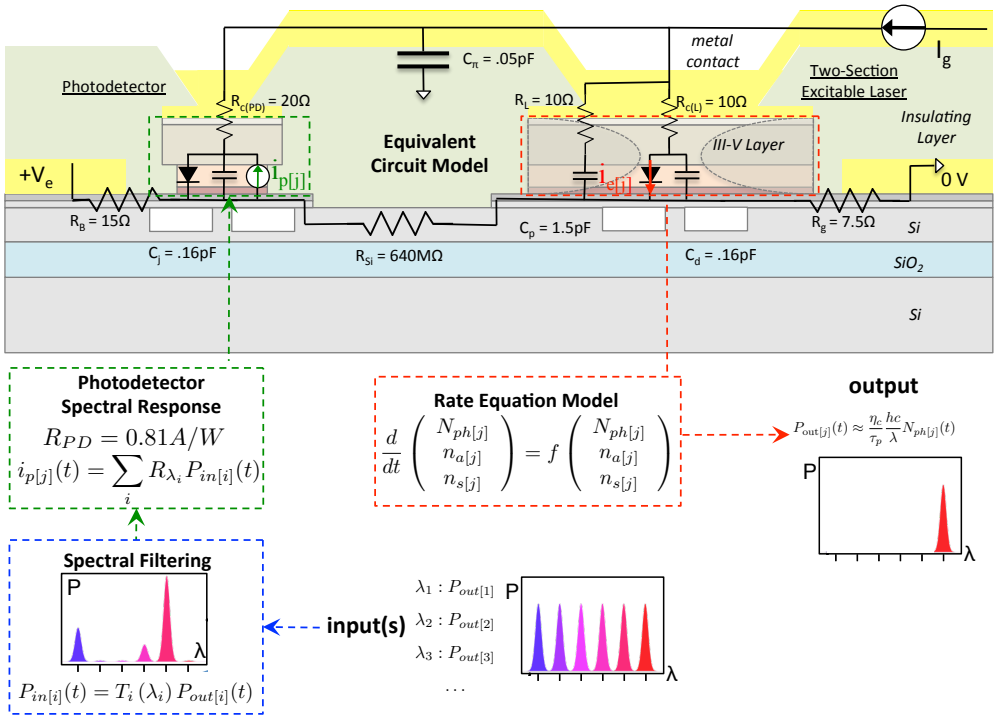

Fig. 3. Cross section of the device and full modeling structure for the hybrid silicon/III-V laser neuron (inhibitory photodetector not included). A series of pulses along n different wavelength channels $\lambda_1 ... \lambda_N$ are spectrally filtered (i.e. weighted). This results in an excitatory photodetector current response, which propagates into the equivalent circuit depicted above. The interaction between the photons in the cavity, gain and SA sections are modeled using rate equations. The resulting output power along wavelength λ_j becomes the input to other neurons in a given network.

where F is the linear transfer function of the modeled circuit.

A simulation of multiple input pulses is shown in Fig 5. The laser is pumped with a stable current source I_p, while the PD is reverse biased with a large voltage (>5 V) to offset the influence of I_p. Responsivity is assumed to be $0.81A/W$ [16]. The junction is kept fairly short ($\sim$100s of μm) to avoid transmission line effects.

While the general expression for the complex impedance of the link is more involved, in practice, the dominant parasitics are the capacitance of the metal wire C_π and the contact resistances R_{PD}, R_L. The metal wire capacitance C_π is the easiest to adjust lithographically, either by changing the height of the oxide layer or the area occupied by the metal bridge to change the characteristics of the junction.

A common first order model for synaptic dynamics is an integrator with decay [35]:

$$\frac{ds}{dt} = -\frac{s}{\tau} + I(t)$$

where $I(t)$ is the input, τ is the time constant and $s(t)$ is the synaptic variable. A neuron will typically sum signals from multiple synapses $s_i(t)$ and receive them as inputs. In biological systems, synaptic variables typically represent neurotransmitter concentrations. In our case, it represents the RC charged signal. This behavior takes place independently of temporal integra-

Table 1. Hybrid/III-V Laser Parameters

series Param.	series Description	series Value
V_g	gain section volume	1.68×10^{-11} cm^3
V_α	SA section volume	3.36×10^{-12} cm^3
Γ	QW conf. factor	0.056
n_0	transparency carrier density	1.75×10^{18} cm^{-3}
v_g	group velocity	c/3.49
g_0	QW gain coefficient	966 cm^{-1}
τ_g	gain carrier lifetime	1.1 ns
τ_α	SA carrier lifetime	100 ps
τ_{ph}	photon lifetime	2 ps
n_{sp}	spontaneous noise factor	2
I_g	gain pumping current	21 mA
η_i	injection efficiency	0.6
η_c	differential quantum efficiency	0.26
λ	lasing wavelength	1550 nm

tion in the soma. The synaptic time constant thus has a correspondence with this circuit's RC time constant.

As shown in Fig. 4, the electronic junction does not significantly degrade the bandwidth of input pulses with FWHMs on the order of ∼ps. In fact, the RC time constant can be as low as 30 ps using standard lithographic techniques [15]. This shows that optical injection is not necessary to maintain the information content in high bandwidth pulses, if the electrical junction is short. Nonetheless, a slower synaptic time constant—which is useful for various processing applications—is still possible by engineering a larger capacitance across the wire junction.

2.4. Laser Neuron

2.4.1. Dynamics

The dynamical system underlying the behavior of our processing model is a gain-absorber cavity, describing single mode lasers with gain and saturable absorber (SA) sections. Despite its simplicity, it can exhibit a large range of possible behaviors [9], and has been investigated in various contexts as the basis for an optical processor [22]. The system in its simplest form can be described using the following undimensionalized equations:

$$\dot{G}(t) = \gamma_G[A - G(t) - G(t)I(t)] + \theta(t) \tag{7a}$$

$$\dot{Q}(t) = \gamma_Q[B - Q(t) - aQ(t)I(t)] \tag{7b}$$

$$\dot{I}(t) = \gamma_I[G(t) - Q(t) - 1]I(t) + \varepsilon f(G) \tag{7c}$$

where $G(t)$ models the gain, $Q(t)$ the absorption, and $I(t)$ the laser intensity. A is the gain bias current, B is the absorption level, γ_G is the gain relaxation rate, γ_Q is the absorber relaxation rate, γ_I is the inverse photon lifetime , and a is a differential absorption relative to the gain factor. We represent spontaneous noise via $\varepsilon f(G)$, for small ε, and time-dependent input perturbations as $\theta(t)$.

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

Under a desired parameter regime, the internal dynamics can be compressed so that pulse generation is instantaneous. The behavior simplifies to [12]:

$$\frac{dG(t)}{dt} = -\gamma_G(G(t)-A) + \theta(t); \tag{8a}$$

$$\text{if } G(t) > G_{\text{thresh}} \text{ then} \tag{8b}$$

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

where the input $\theta(t)$ can include spike inputs of the form $\theta(t) = \sum_i \delta_i(t-\tau_i)$ for spike firing times τ_i, G_{thresh} is the gain threshold, and $G_{\text{reset}} \sim 0$ represents the gain at transparency. This system is analogous to a leaky integrate-and-fire (LIF) neuron model, commonly employed in computational neuroscience for modeling biological neural networks. Although it is one of the simpler spike-based models, the LIF model is capable of universal computations [36], and is capable of coding information in the timing between spikes [37].

2.4.2. Device

We consider a two-section distributed feedback (DFB) excitable laser in the hybrid platform. We use a DFB cavity to guarantee a single longitudinal lasing mode, defined through the lithographic definition of lasing wavelength via silicon-on-insulator grating pitch. The quarter wavelength-shifted defect is placed asymmetrically in the cavity to direct the light in one direction. An ion implantation region electrically isolates the semiconductor absorber and gain sections, and a separate implant provides a smaller lifetime in the absorber section. Small modulation currents that exceed the Q-switch threshold will trigger large pulse discharges (see Fig. 5(a)).

We define the effective gain and absorption factors for each section as:

$$\tilde{g}(n_{g[j]}) = \Gamma \frac{v_g g_0}{n_0}(n_{g[j]} - n_0)$$

$$\tilde{\alpha}(n_{\alpha[j]}) = \Gamma \frac{v_g g_0}{n_0}(n_0 - n_{\alpha[j]})$$

where Γ denotes the confinement factor of the QW sections with the transverse mode, v_g is the waveguide group velocity, g_0 is the gain coefficient of the material, n_0 is the transparency density and n_g, n_α denote the carrier concentrations in the gain and absorber sections, respectively.

For laser j, we can describe the internal cavity dynamics using the following system of equations:

$$\frac{dN_{ph[j]}}{dt} = \left[\tilde{g}(n_{g[j]}) - \tilde{\alpha}(n_{\alpha[j]}) - \frac{1}{\tau_{ph}}\right] N_{ph[j]} + \tilde{g}(n_{g[j]})\frac{n_s}{V_g} \tag{10a}$$

$$\frac{dn_{g[j]}}{dt} = \frac{I_{g[j]} + i_{e[j]}(t)}{eV_g} - \frac{n_{g[j]}}{\tau_g} - \tilde{g}(n_{g[j]})\frac{N_{ph[j]}}{V_g} \tag{10b}$$

$$\frac{dn_{\alpha[j]}}{dt} = -\frac{n_{\alpha[j]}}{\tau_\alpha} + \tilde{\alpha}(n_{\alpha[j]})\frac{N_{ph[j]}}{V_\alpha} \tag{10c}$$

where $N_{ph}(t)$ is the total number of photons in the cavity, $n(t)$ is the number of carriers, n_0 is the transparency carrier density, V is the cavity volume, Γ is the confinement factor, τ is the carrier lifetime, τ_{ph} is the photon lifetime, n_s the spontaneous emission factor, and $i_e(t)$ represents the electrical modulation in the gain provided by the photodetector system, separate from the constant current bias I_g. Subscripts g and α identify the active and absorber regions. Parameter values are shown in Table 1 and based on those found in [31, 34]. The absorber is

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

assumed to have a faster lifetime, which helps improve the dynamics and consistency of the input signals [13]. The equations can be simplified to the undimensionalized set of equations through a set of variable substitutions [12, 38]. The results are shown in Fig. 5.

The output power of laser j can be computed via:

$$P_{out[j]}(t) = \frac{\eta_c}{\tau_{ph}} \frac{hc}{\lambda_j} N_{ph[j]}(t) \tag{11}$$

for output coupling efficiency η_c, photon lifetime τ_{ph}, Planck's constant h, speed of light c, wavelength λ_j, and photon number $N_{ph}(t)$. The resulting single-mode power spectrum of the laser takes the form $S_{out[j]}(\lambda,t) = P_{out[j]}(t)\delta(\lambda - \lambda_i)$, as mentioned in section 2.2. The signal $S_{out[j]}$ is coupled into the broadcast loop, becoming the input of other PNNs.

2.5. *Non-Idealities*

Although this analysis has considered the physics of the signal pathway to first order, there are many second order non-idealities that will require further study and consideration. Discussed here is a brief overview of the major sources that could lead to decreased performance. First, let us consider fan-in limitations. In addition to the fundamental transform-limited bandwidth $\delta\lambda_B$, there is also temperature-induced linewidth fluctuations $\delta\lambda_T$ and carrier-induced laser chirp $\delta\lambda_C$. If the filters have a linewidth of $\delta\lambda_F$, we must space filters by at least

$$(\Delta\lambda)^2 \approx (\delta\lambda_B)^2 + (\delta\lambda_T)^2 + (\delta\lambda_C)^2 + (\delta\lambda_F)^2 \tag{12}$$

to avoid significant interference. Closely spaced, overlapping filters can unintentionally allow the same signal to interfere with itself, resulting in coherent crosstalk. This manifests as phase differences $\delta\phi$ between each path becoming visible as amplitude noise δP with an SNR that depends on the magnitude difference between the filtered signals.

Therefore, for a given a channel spacing $\Delta\lambda$ and gain bandwidth G_λ (typically around $\sim$50 nm), the maximum channel fan-in capacity C_f of the system can be defined as:

$$C_f \approx \frac{G_\lambda}{\delta\lambda} \tag{13}$$

This puts a fairly hard limit on the fan-in of neurons. In the very best case, (i.e. bandwidth-limited channels with 10 ps pulses in a 50 nm gain bandwidth), we can expect a maximum of $\sim$200 channels. Although it is a hard limit, this number is still very high compared to what is possible in electronics, and still unmatchable at such enormous signal bandwidths (i.e. $\sim$40 GHz). In addition, networks using clever organizational strategies can include many more PNNs than the fan-in limit, as mentioned earlier in the manuscript.

We must also consider noise sources that can compromise the SNR of the signal as it travels through the system, which has not been studied here. There are many sources of noise to consider including amplified spontaneous emission (ASE) noise, thermal noise, and shot noise for both the optical and electrical signal. Since the laser pump current I_p is the largest signal in the pathway, we can expect that its shot noise contribution σ_s will dominate as input power is reduced. Nonetheless, all noise sources can play a contribution in reducing the signal SNR as it travels through the pathway and obfuscates the lasers ability to make a discrete decision based on its input. The effect of noise on this pathway remains to be studied, although it will not be significantly higher than what is already seen in directly modulated laser diodes and photodetectors.

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

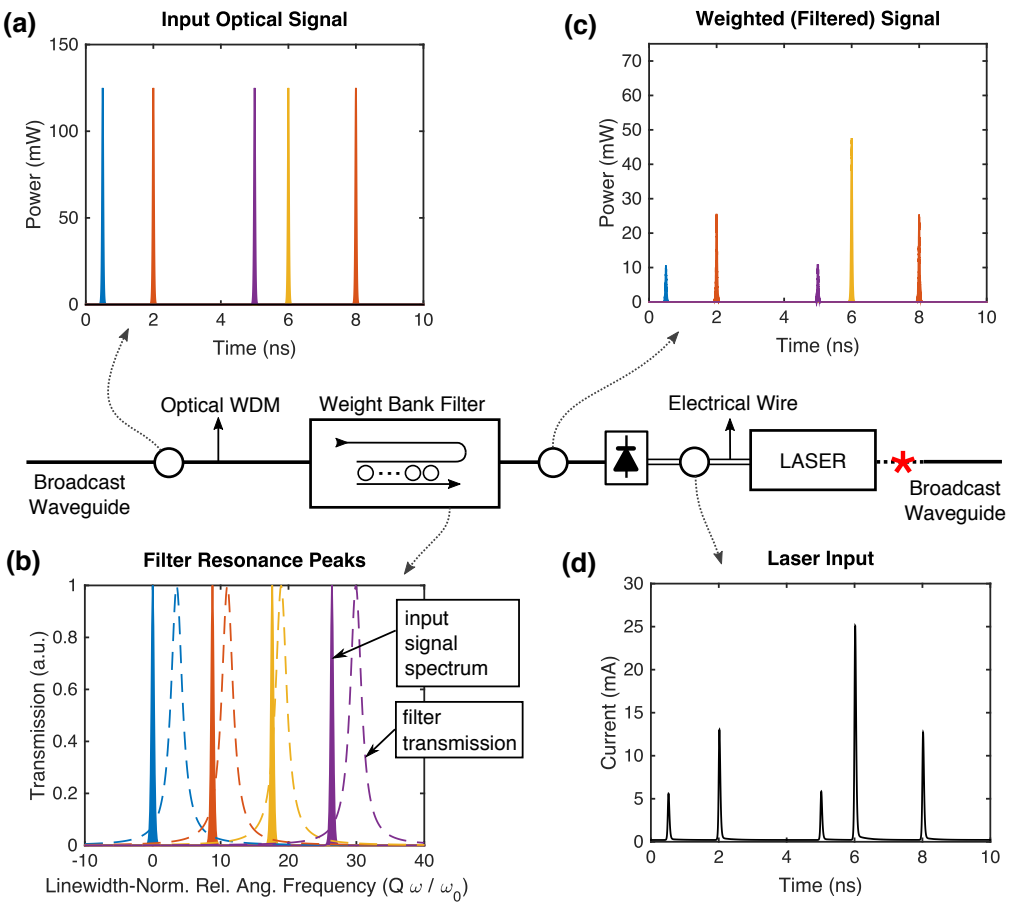

Fig. 4. Simulation of the response of the processing network node at each step. (a) Input WDM spike signals coming from the broadcast waveguide with FWHM = 40 ps. Trace color indicates the carrier wavelength of each pulse. (b) Transmission of spectrum of the weight bank filter (dotted lines) and input signal power spectra (solid curves). Input signals are assumed to be at the transform limit for 10 ps pulses. (c) WDM signals after transmission through the filter bank. Pulses on the same wavelength channel acquire the same weight and are then detected. (d) Electronic current signal, $i_e(t)$, that modulates the laser neuron after traversing the parasitic circuit model in Fig. 3. Pulses are low pass filtered to FWHM = 56 ps after traveling through device parasitics.

3. Discussion

The model developed in this paper—including the filter bank, electrical junction, and laser neuron—enables the exploration of both analog signal properties and the effect of physical design decisions on the behavior of networked PNNs. An example simulation of the complete model with WDM optical-in and optical-out is shown in Figs. 4 and 5. The front-end weight bank and photodetector in Figs. 4(a)–4(c) act to generate an electronic representation of the weighted sum of WDM inputs carrying FWHM=40 ps pulses. This signal that modulates the laser neuron is shown in Fig. 4(d), after traversing the parasitic circuit model from Fig. 3. While some pulse spreading is visible (FWHM=56 ps), pulse amplitude and timing information is clearly maintained when realistic parasitic values are used. Figure 5 illustrates the internal dy-

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

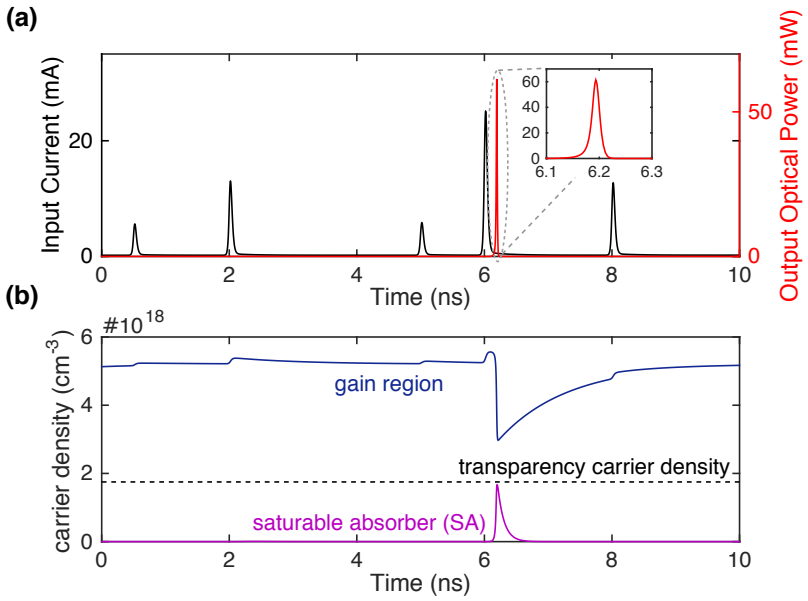

Fig. 5. Simulated internal dynamics of the laser neuron in response to the modulation signal in Fig 4.d. (a) input electrical modulation $i_e(t)$ (black line) causes the release of a large optical pulse near 6ns (red line). Inset shows the magnified output spike, which is approximately a $sech^2$ pulse with FWHM=16.5 ps. (b) Simulated carrier densities in both the gain (blue line) and saturable absorption (purple line) sections of the laser. The dashed line represents the transparency carrier density, n_0.

namics of the excitable laser neuron in response to this modulation signal. Small perturbations to the gain population produce no optical output, but a larger input causes saturation of the SA and release of a short pulse (FWHM=16.5 ps), even narrower than the original input pulses. This is a testament to the system's regenerative properties, which can continue to propagate pulses without their eventual degradation [12].

This simulation indicates that the PNN design meets a number of important scalability criteria: the dynamical processing model is both cascadable and restorative, the signal pathway can accept large fan-in and perform summation, elements are compatible with a high-volume multiplexing scheme, and the processing nodes are constructed from standard elements that can be implemented on PICs. Prior work on spiking laser neurons has largely focused on single laser dynamical phenomena, while suggesting the possibly of all-optical processing networks. No realistic approach for many-to-one interconnection with optically injected lasers has been proposed, and the well-known challenges of all-optical modulation (wavelength conversion, phase noise accumulation, two-beam beat interference) remain a question in these systems.

The electronic mediation in the PNN enables a cascadable node based on a spiking laser neuron, but does not bring the typical downsides of O/E/O conversion in digital communication systems because transduction is separated from receiving (i.e. sampling, quantization, retiming). The energy and cost of electronic conversion in optical systems results largely from high-speed clocked transistor circuitry [39] and the need to demultiplex WDM channels before conversion [40]. In our system, electronic transduction does not regenerate or receive signals, but instead exploits the electronic physics for intermediate analog processing. The conversion between optical and electronic domains curtails the propagation of optical phase noise and the need for direct wavelength conversion, eliminating two major barriers facing scalable optical

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015

computing [41]. At the same time, the temporary conversion to the electrical domain does not significantly degrade the signal, as this work has simulated. Every device in the primary signal pathway performs both physical and computational roles, resulting in a robust, ultrafast, and efficient signal pathway.

Utilization complementary physics along the signal pathway is not unlike this pathway in biological neurons, in which electrical action potentials are converted to chemical signals called neurotransmitters upon reaching a synapse (the junction between neurons). Chemical signaling is a relatively short-range process, but it introduces much more functionality compared to direct electrical modulation (e.g. both excitation and inhibition, a variety of synaptic timeconstants, and the ability to adapt the connection strength). The post-synaptic neuron body passively accumulates the effect of many synapses, and its internal spiking dynamics generate a single channel of electrical action potentials, which is suitable for long-distance transmission. This process is roughly paralleled by O/E/O conversion in the PNN: the balanced photodetector and analog electrical link enable cascadable summation of multiwavelength inputs, from which an excitable laser generates a new spiking optical signal. This functionality does not impact bandwidth performance because the electrical link can be very short-range, similar to how the neurotransmitter link is very short in the biological pathway.

4. Conclusion

We have investigated a device that can act as a node in an ultrafast photonic neural network, using a computational model that fully exploits the temporal resolution of optical signals. The architecture is physically enabled by three primary discoveries. First, a laser with gain and saturable absorption sections behave analogously to a biological neuron, but with dynamics and temporal resolutions on the order of nanoseconds and picoseconds, respectively. Secondly, a WDM broadcast-and-weight networking approach allows a large number of signals to be superimposed on a single waveguide where connection strengths are easily configured by tunable filter banks. Finally, a short receiver-less electronic link can carry a WDM fan-in signal without significant distortion or loss. By using the full model constructed here that takes into account many physical effects, we show that these optical subcircuits constitute a complete PNN capable of participating in a cascadable, scalability photonic neural network.

Although the PNN promises several unique properties advantageous for processing, there still remain aspects that require exploration as the number of elements increase—these include the reliability and configurability of WDM filter-based weighting, the effect of noise accumulation, the effect of transform-limited pulses and chirp on the interconnection network density, and perhaps most importantly, power consumption and the effect of temperature on a large number of active elements clustered together on a PIC. Nonetheless, this represents an important direction in the consideration of scalability for photonic neural networks. Further developments could pave the way for processing systems that could perform real-time, complex processing of ultrafast (GHz) signals in real-time, which has the potential to revolutionize radio cognition or the control of ultrafast physical phenomena.

Funding Information

The authors acknowledge the support of the National Science Foundation Graduate Research Fellowship Program (NSF GRFP) and the Banting Postdoctoral Fellowship administered through the Natural Sciences and Engineering Research Council of Canada.

#246538 Received 22 Jul 2015; revised 15 Sep 2015; accepted 22 Sep 2015; published 2 Oct 2015