

Nanophotonic Cavity Based Synapse for Scalable Photonic Neural Networks

Aashu Jha , Chaoran Huang , Thomas Ferreira deLima , *Graduate Student Member, IEEE*, Hsuan-Tung Peng ,
Bhavin Shastri, *Senior Member, IEEE*, and Paul R. Prucnal, *Life Fellow, IEEE*

Abstract—The bandwidth and energy demands of neural networks has spurred tremendous interest in developing novel neuromorphic hardware, including photonic integrated circuits. Although an optical waveguide can accommodate hundreds of channels with THz bandwidth, the channel count of photonic systems is always bottlenecked by the devices within. In WDM-based photonic neural networks, the synapses, i.e. network interconnections, are typically realized by microring resonators (MRRs), where the WDM channel count (N) is bounded by the free-spectral range of the MRRs. For typical Si MRRs, we estimate $N \leq 30$ within the C-band. This not only restrains the aggregate throughput of the neural network but also makes applications with high input dimensions unfeasible. We experimentally demonstrate that photonic crystal nanobeam based synapses can be FSR-free within C-band, eliminating the bound on channel count. This increases data throughput as well as enables applications with high-dimensional inputs like natural language processing and high resolution image processing. In addition, the smaller physical footprint of photonic crystal nanobeam cavities offers higher tuning energy efficiency and a higher compute density than MRRs. Nanophotonic cavity based synapse thus offers a path towards realizing highly scalable photonic neural networks.

Index Terms—Photonic integrated circuits, photonic neural networks.

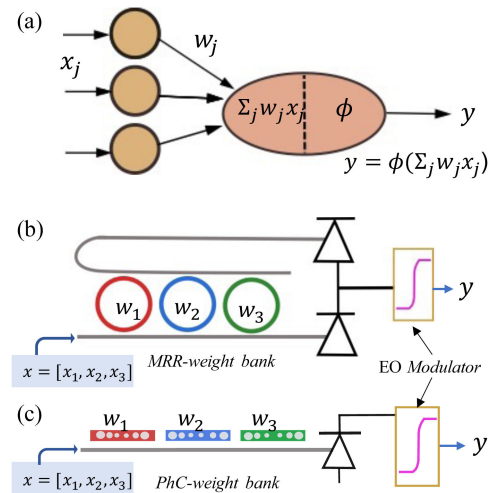


Fig. 1. (a) Model of the neuron showing linear weighting and summing of inputs, followed by a nonlinear function, ϕ . (b) WDM compatible MRR-based neuron with balanced photodetector and an electro-optic (EO) modulator, and (c) a PhC-based neuron with photodetector and an EO modulator. Photodetector carries out the summing operation while the EO modulator relays the nonlinear transformation.

I. INTRODUCTION

ARTIFICIAL intelligence (AI) enabled by neural networks has seen an explosive growth in the last decade [1]. Neural network algorithms are inherently parallel and thus inefficient on conventional hardware where processing is sequential, this challenge has motivated neuromorphic hardware engineering [2]. In neuromorphic electronics, bandwidth and interconnectivity have to be traded off [2]. Photonics offers the opportunity to simultaneously achieve high bandwidth with high interconnectivity, in tandem with low power consumption and low latency [3]. The interconnections between “neurons” in a neural network are known as synapses. Fig. 1(a) shows the functional model of a neuron: at each neuron, inputs from other neurons are *linearly* weighed by synapses, summed and nonlinearly transformed, generating an output. In photonics, one can theoretically parallelize hundreds of high speed information channels onto a single waveguide by encoding each channel on a different wavelength, a scheme known as wavelength-division multiplexing (WDM) [4]. Photonic implementations of synapses include coherent and WDM systems [5]–[8]. WDM maximally utilizes the high bandwidth of photonics, which then translates to significant network benefits in terms of information throughput.

Manuscript received February 22, 2022; revised May 19, 2022; accepted May 30, 2022. Date of publication June 3, 2022; date of current version June 29, 2022. This work was supported by the Office of Naval Research Award under Grants N00014-18-1-2297 and N00014-20-1-2664. (Corresponding author: Aashu Jha.)

Aashu Jha is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: aashuj@princeton.edu).

Chaoran Huang is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA, and also with the Department of Electronic Engineering at the Chinese University of Hong Kong, (e-mail: chaoranh@princeton.edu).

Thomas Ferreira deLima is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA, and also with the NEC Laboratories America, Princeton, NJ 08542 USA (e-mail: tlima@princeton.edu).

Hsuan-Tung Peng is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA, and also with the PsiQuantum, Palo Alto, CA 94304 USA (e-mail: hpeng@alumni.princeton.edu).

Bhavin Shastri is with the Department of Physics, Engineering Physics and Astronomy, Queens University, Kingston, ON K7L 3N6, Canada (e-mail: bhavin.shastri@queensu.ca).

Paul R. Prucnal is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: prucnal@princeton.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSTQE.2022.3179983>.

Digital Object Identifier 10.1109/JSTQE.2022.3179983

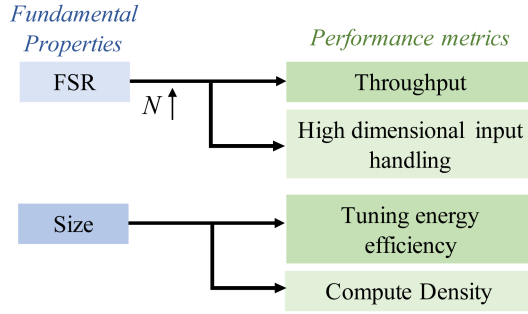


Fig. 2. Performance benefits of PhCs as synapses. Their fundamental properties, such as FSR-free operation and small physical footprint, translates to improvements in system performance metrics, including throughput, ability to handle high-dimensional input applications, higher tuning energy efficiency and high compute density. N refers to the WDM channel count.

Ref. [6] introduced the concept of a microring weight bank, shown in Fig. 1(b), which is a WDM-based synapse where an array of microring resonators (MRRs) selectively weighs an input vector of wavelength channels. Each MRR addresses a given wavelength channel where the transmission corresponding to that wavelength is the weight for the synapse. The optical output from the pass and drop ports of the weight bank are then summed and differentially detected by a pair of balanced photodetectors. The summed electrical signal from the PD then drives an electro-optical modulator with an optical pump to generate a nonlinearly transformed optical output [9].

Weighting at synapses represents the majority of operations in a neural network, the overall network performance is thus heavily dependent on the synapse properties, as encapsulated in Fig. 2. The free-spectral range (FSR) of an MRR limits the number of wavelength channels, N , i.e. the number of synapses in a network. We estimate $N \leq 30$ for silicon MRRs. The constraint on channel count inevitably limits the throughput, i.e. the amount of information being processed, as well constrains the input dimension making some applications unfeasible. Ref. [10] shows an approach to eliminate the FSR constraint by suppressing all but one microring resonance but the level of suppression varies with wavelength which isn't ideal. In practice, N is even lower due to effects like inter-channel crosstalk, and there have been efforts to mitigate it [11]. Besides spectral scale, the physical scale of the network also depends on the synapses. A smaller synapse improves the compute density, which is a measure of the number of operations done per unit time normalized per unit area [12]. The size of a photonic device used as a synapse determines the energy consumed while tuning the weight – the smaller the device, the higher the efficiency. While MRRs are relatively compact, as compared to interferometers [5], you can only make it so much smaller without introducing too much bending loss, typically MRRs have radii larger than $5 \mu\text{m}$. To make the synapse smaller entails a novel synapse.

Here, we present a novel WDM synapse – a photonic crystal (PhC) weight bank, as illustrated in Fig. 1(c). Photonic crystals (PhCs) are nanophotonic devices with periodic dielectric function that allows to manipulate the propagation of light. By introducing defects or breaking the periodicity in some way, localized modes for some light frequency can be formed [13]

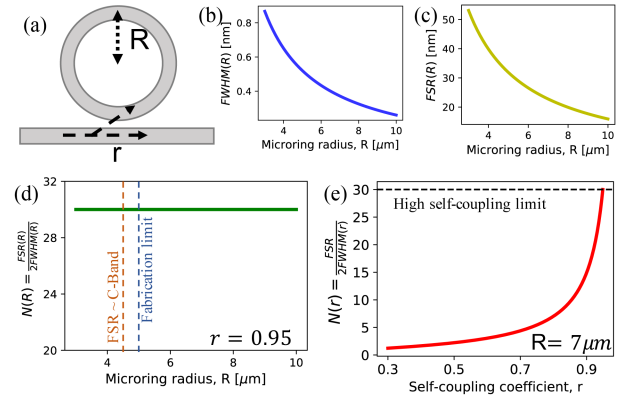


Fig. 3. WDM channel count limitation of microring resonator based WDM systems, (a) Illustration of a microring of radius R and coupling coefficient, r , (b) full-width half-maximum (FWHM) of MRR resonance as a function of radius R , (c) free-spectral range (FSR) of MRR as a function of radius R , (d) WDM channel count N , calculated as $\frac{FSR(R)}{2FWHM(R)}$ as a function of R , (e) WDM channel count N as a function of coupling coefficient r .

forming the basis of engineering optical resonators. They are also widely popular for nonlinear photonic applications like all-optical switching [14], [15], electro-optic modulation [16], [17]. The simplest form of PhCs is a one-dimensional PhC, also known in literature as nanobeam cavities. Their one-dimensional nature means they can easily be implemented in photonic integrated circuits where light can be evanescently coupled through a bus waveguide. PhC nanobeam based devices have previously been used as wavelength filters [18]. Here, we propose a PhC-based synapse, which replaces the MRRs in the current weight bank architecture, for two reasons. Firstly, we show that PhCs can be FSR-free, which significantly improves the WDM channel count, thereby enhancing the system throughput and enabling neural network applications with high-dimensional inputs. Secondly, we show that the physical size reduction with PhCs improves the overall tuning energy efficiency. Our results show that photonic crystals have the potential to enable the most scalable next-generation photonic neural networks with high throughput, energy efficiency and compute density.

II. SCALING CHALLENGE IN MRR-BASED SYNAPSES

Before proceeding to discussing PhCs, we would like to quantitatively analyse the scaling challenge in MRR-based synapses. The channel count in an MRR-based WDM link is proportional to the ratio of the free-spectral range (FSR) to the full-width half-maximum (FWHM) of the resonance [6], [19]. The design parameters in an MRR available to tune its FSR and FWHM are the radius, R and the gap between the microring and the coupling waveguide which determines the self-coupling coefficient, r , (refer Fig. 3(a)). The range of each parameter is constrained by foundry design rules due to their fabrication tolerances. To illustrate the channel count limitation fundamental to MRRs, we show how R and r affect the channel count, N in Figs. 3(b)–(e). Figs. 3(b) and (c) show that FWHM and FSR decrease similarly with increasing R . This results in $N = \frac{FSR}{2FWHM}$ being a constant function of R , as shown in Fig. 3(d). In other words, one cannot change the allowable channel count by changing

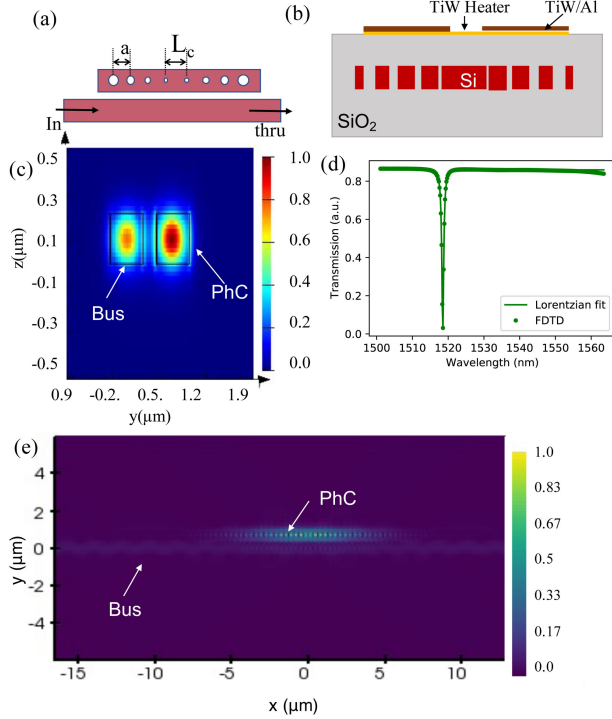


Fig. 4. (a) Design schematic of the PhC nanobeam cavity with horizontally symmetric tapered holes with periodicity a , and cavity length L_c . Only 4 holes on either sides are shown for simplicity of illustration. (b) Illustration of the PhC cross-section. (c) Transverse mode profile of a bus waveguide and PhC cross-section at the horizontal symmetry plane showing evanescent coupling into the PhC. (d) Simulated transmission spectrum of the PhC cavity, revealing a single Lorentzian resonance feature and thus its FSR-free property. (e) Simulated 2D electric field intensity profile showing the resonant cavity mode confined in the PhC.

the MRR radius. Alternatively, one can vary the self-coupling coefficient, r . Fig. 3(e) shows that N increases with r . However, one cannot indefinitely increase r —in the high self-coupling limit ($r = 0.95$), N approaches to 30. This illustrates that MRR-based WDM links run into fundamental limits and cannot be scaled indefinitely.

III. PHC DESIGN

An illustrative top-down view of the PhC nanobeam synapse is shown in Fig. 4(a). It comprises a one-dimensional lattice of cylindrical holes etched in a single-mode silicon waveguide, as evident in the side-view of the cavity in Fig. 4(b). The 1D nature eases both the design and fabrication constraints relative to higher-dimensional PhC cavities. The PhC is horizontally symmetric with 40 linearly tapered holes with radii in the range $[0.11, 0.04] \mu\text{m}$ and constant hole spacing of $a = 0.33 \mu\text{m}$ on either side. The combination of linear radii tapering and a large hole count ensures the phase velocity is preserved within each segment of the nanobeam, minimizing losses in the cavity [20], [21]. The resonance wavelength of the PhC can be varied by changing the cavity length, L_c , which refers to the separation between the two central holes. Light is coupled into the PhC through a bus waveguide as shown in Fig. 4(a). The gap between the waveguide and the PhC cavity was chosen to be $0.2 \mu\text{m}$,

which optimized the cavity quality factor and is close to parameters in [22]. The width and thickness of both the bus and nanobeam waveguides are 500 nm and 220 nm respectively.

We simulated the optical field distribution and transmission of the PhC shown in Fig. 4(a). Fig. 4(c) shows the optical mode at the horizontal symmetry plane of the PhC, calculated by a finite-difference eigenmode (FDE) solver on Lumerical MODE. It illustrates the evanescent coupling of light from the bus waveguide into the photonic crystal cavity. We then used a variational finite difference time domain solver in Lumerical, to simulate the transmission spectrum of the PhC across a wavelength range of 100 nm . Fig. 4(d) shows the transmission spectrum of the PhC, showing a single resonance at $1.519 \mu\text{m}$. This illustrates the FSR-free nature of the PhC within the wavelength range. Fig. 4(e) shows the resonant mode within the cavity, occupying $0.16 \mu\text{m}^3 \approx 0.59(\lambda/n)^3$, which is at least an order of magnitude lower than typical microrings. Time apodization was used so that only the resonant mode profile was captured.

For application as synapses, the PhC resonance wavelength needs to be tunable. Similar to MRR-based synapses, tuning the resonance wavelength tunes the transmission value at a given wavelength. In neural network terms, the transmission is the “weight” associated to a synapse. We employ a simple thermal tuning mechanism using metal heater deposited atop the photonic cavity, as shown in Fig. 4(b). The heat from the metal heater results in a local temperature increase which tunes the local index, due to the thermo-optic effect, and thereby tunes the cavity resonance. There are many other approaches to index tuning in silicon waveguides [23] which have been employed in PhCs such as carrier-injection and depletion modulation using PN junctions [24], as well as phase-change materials (PCMs) [25].

IV. EXPERIMENTAL CHARACTERIZATION

To establish the performance merits of PhCs relative to MRRs, we fabricated devices with the design presented above and experimentally characterized their resonance properties. We experimentally verify that PhCs can be FSR-free within operational wavelength range and also show they have higher tuning efficiency than MRRs. We also discuss the design strategy for designing PhC-based WDM photonic synapses, shown in Fig. 1(c).

Fabrication was carried out on a commercial process at Applied Nanotools, Canada. The devices were fabricated on a silicon-on-insulator (SOI) wafer with $3 \mu\text{m}$ oxide cladding atop silicon device layer. The silicon PhC, shown in the SEM image in Fig. 5(b), was patterned via 100 keV electron beam lithography as the feature size entails high resolution fabrication. Fig. 5(a) shows the metal layers deposited atop for thermal tuning. A 200 nm thick layer of high-resistance titanium-tungsten alloy (TiW), with bulk resistivity (ρ) of $0.61 \mu\Omega\text{-m}$, is used as the heater. A low resistance routing layer composed of tungsten/aluminum bi-layer is used to form contact to the heater.

The schematic of the experimental setup is shown in Fig. 6. The photonic chip was placed on a temperature controlled copper mount. The optical coupling into the on-chip waveguide is done by a Vgroove fiber array (not shown) incident on on-chip TE

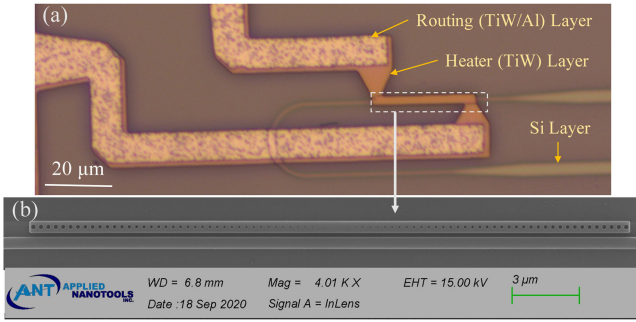


Fig. 5. (a) Optical micrograph of a fabricated photonic crystal nanobeam cavity, showing the Si device layer as well as the routing and heater layers (taken with a Zeiss AxioScope microscope) (b) Zoomed-in view of an all-pass photonic crystal nanobeam filter test structure captured with an SEM (Image taken by Applied Nanotools, Canada).

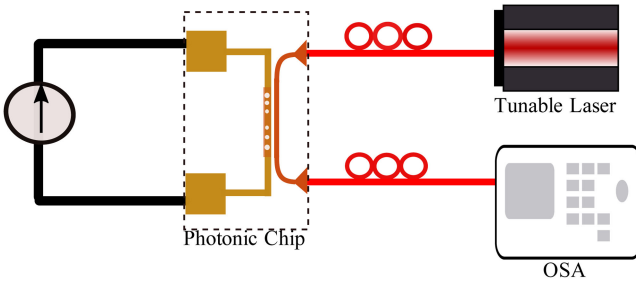


Fig. 6. Experimental setup for PhC characterization: electrical control for resonance tuning by a Keithley current source, electrical and optical coupling with the photonic chip is via electrical probes and V-groove fiber array respectively (not shown). Optical source is from a tunable laser and the output is collected by an optical spectrum analyzer (OSA).

grating couplers. The optical input to the PhC is from a tunable laser (TL) source, and the optical output is fed to an optical spectrum analyzer (OSA). The round-trip optical insertion loss of the grating couplers was 16 dB. However, recent advancements in fiber-to-chip coupling strategy should reduce it below 1 dB [26]. For electrical control of the metal heater, an off-chip current source is used, electrical coupling to the on-chip contact pads is through electrical probes (Picoprobe). The current source is remotely controlled using a lab automation software package, Lightlab [27].

Fig. 7 shows a schematic illustration of the experiment – input light of varying wavelengths is sent to the PhC and the output is measured. The measured transmission is thus the ratio of the output to the input at each optical frequency in the range 1525–1565 nm (spanning the entire C-band). From the measured transmission spectrum of a PhC cavity shown, we see a single resonance feature at around 1530.5 nm, with a Q-factor of 12,000 and extinction ratio of ≈ 14 dB. Having only a single resonance indicates the cavity is free-spectral range (FSR)-free. In contrast, for an MRR, a series of recurring resonances spaced an FSR (tens of nm) apart is expected. The wide-band envelope observed in the transmission spectrum can be attributed to the spectrum of the grating coupler, and the oscillation to the air gap at the fiber-chip interface.

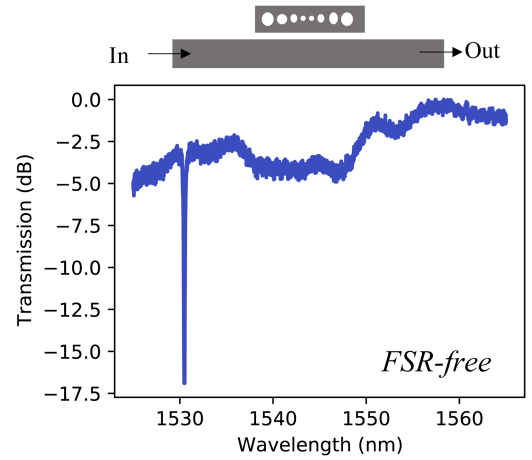


Fig. 7. (top) Schematic illustration of the experiment showing the device input and output. (bottom) Measured transmission spectrum of a PhC cavity showing a single resonance within the C-band, verifying its FSR-free property.

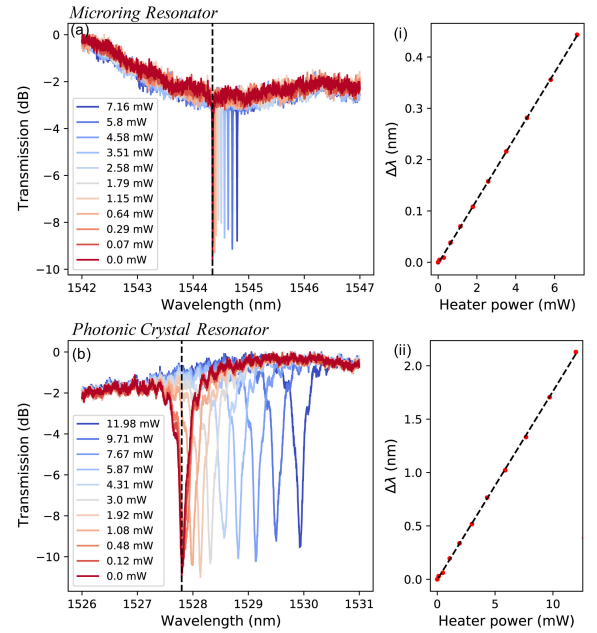
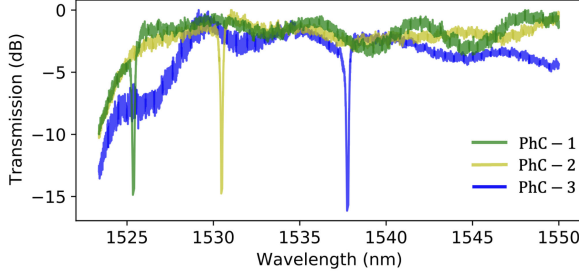


Fig. 8. Comparison of thermal tuning efficiency between an MRR and a PhC cavity test structures fabricated on the same chip. (a), (b) show the measured transmission spectra, and (i), (ii) show the corresponding resonance shifts as a function of heater powers for an MRR and a PhC respectively.

We then compare the experimentally measured tuning efficiencies of a PhC and an MRR fabricated on the same photonic die with identical routing and heater metal layers. The MRR design parameters are: 8 μm radius and 300 nm gap between bus and MRR. Tuning efficiency here is defined as the ratio of measured shift of the resonance wavelength ($\Delta\lambda$) to the applied heating power (P_{heat}) calculated as $P_{\text{heat}} = I^2 \times R$ where I is the applied current and R is the resistance measured by the Keithley source meter. Figs. 8(a) and (b) show the change in the transmission spectra with thermal tuning for an MRR and a PhC respectively. Fig. 8(i),(ii) show the corresponding resonance shifts as a function of heater power. Using linear regression,



	L_c	λ_0	Q	ER
PhC-1	0.25	1525.4	11,140	11.9
PhC-2	0.3	1530.5	13,250	13.2
PhC-3	0.35	1537.8	10,900	12.2

Fig. 9. (top) Measured transmission spectra 3 PhCs of varying cavity lengths. (bottom) Properties of the PhCs, including the cavity length L_c , resonance wavelength λ_0 , quality factor Q and extinction ratio ER .

we calculate tuning efficiencies of 0.18 nm/mW for the PhC and 0.04 nm/mW for the MRR. The efficiencies are expected to be an underestimate as R is the resistance of the entire circuit. However, it is adequate for a relative comparison, where we find the PhC has a 4.5 times higher thermal tuning efficiency relative to the MRR. Index-tuning efficiency increase here arises from the cavity mode volume reduction in a PhC which translates to an enhanced light-matter interaction relative to an MRR. Ref. [18] demonstrated a three times better tuning efficiency in a PhC nanobeam filter compared to an MRR; our results qualitatively agree.

A. A PhC Weight Bank

Fig. 9 illustrates the PhC weight bank design – it is a series of PhCs coupled to a common bus waveguide. Each PhC of the weight bank is assigned a different wavelength channel, which means the resonances of the PhCs need to varied such that they collectively span the operational wavelength range. One way to design PhCs with varying resonances is to vary the cavity length, L_c (refer Fig. 4). We fabricated PhCs with varying L_c and measured their transmission spectra and their resonance wavelengths. Fig. 9 shows the measured transmission spectra of three PhC filters with varying L_c (0.25 μm , 0.3 μm and 0.35 μm). We find that varying L_c by 100 nm results in a resonance wavelength shift of 12.4 nm (as reported in Fig. 9). Superposing the transmission spectra of the three PhCs can approximate the spectra of their corresponding weight bank. The fabrication limit of the foundry was reported to be about 90 nm, which means varying L_c cannot be a reliable design strategy for finer tuning of the resonance wavelength. For finer tuning, active resonance tuning mechanisms, such as thermal tuning, or post-fabrication trimming [28] will be needed.

V. PERFORMANCE BENEFITS OF PHC-BASED SYNAPSES

Here we discuss the system performance metrics where PhC-based synapses outperform their MRR counterparts. Fig. 2 encapsulates how the fundamental synapse properties lead to overall performance metric improvements. Firstly, as shown

TABLE I
PERFORMANCE METRIC COMPARISON BETWEEN PHCS AND MRRS

Performance metrics	Devices		Source
	PhC	MRR	
Energy per synapse [pJ/bit]	0.012	0.06	[18], [29]
Compute density [PMAC/smm ²]	0.77	0.1	Calculated
Throughput (in C-band) [Tbps]	7.8	1.8	Calculated
Input dimension (in C-band)	130	30	Calculated

in previous sections, PhCs can be designed to be FSR-free, which increases the channel count in WDM-based weight banks. More channels, in turn, means higher information throughput and amenability to high-dimensional inputs. Secondly, PhC nanobeam cavities, as discussed prior, occupy the same area as a single-mode waveguide and have a higher optical confinement. This size reduction brings about benefits in terms of improved tuning energy efficiency and compute density. A quantitative comparison between PhCs and MRRs across these metrics can be found in Table I.

First, we examine how eliminating the FSR limitation in PhC increases the WDM channel count relative to MRRs within the C-band. In this comparison, we estimate the theoretical channel count in the limit as: $N = \frac{FSR}{2FWHM}$ (refer Fig. 3). Due to the FSR-free nature of the PhC filters in the C-band, channels can span the entire C-band. We assume identical Q factor of 10^4 , which is typical for silicon MRRs and also within the range of that of our fabricated PhCs as reported in Fig. 9. Identical Q implies identical FWHM, which means N now depends solely on the FSR. Assume a typical MRR FSR of 12 nm and we set the FSR proxy of PhC to be 35 nm (entire C-band). We find that $N = 130$ for PhCs, which is 4 times higher than in an MRR weight bank (refer Fig. 3). If we don't constrain to the C-band, the system operational wavelength range will only be limited by the bandwidth of other components like off-chip amplifiers, grating couplers, etc. Also, here we assume identical Qs, however by tweaking the mirror holes in the PhC cavity, Q factors up to 10^7 have been shown [20] which would result in $> 10^4$ channels just within C-band. We do note that a higher Q means a longer cavity photon lifetime (τ) (by the relation $\tau = \frac{2Q}{\omega_0}$ where ω_0 is the cavity resonance frequency) which can adversely affect other metrics like processing speed. We note that any thermal crosstalk is not considered here, and given the all-pass nature, coherent crosstalk effects, present in add-drop MRRs [11], are absent.

A direct implication of improved channel count is seen in throughput. Throughput is a primary incentive for using WDM-based photonic synapses. It can be estimated as the signal speed (limited by the synapse bandwidth) times the number of WDM channels. For Q of 10^4 , typical for silicon MRRs, the maximum signal rate would be 60 Gbps. Then the maximum system throughput with the maximum channel count of 30 would be 1.8 Tbps. In PhCs, owing to increased channel count, the throughput can scale up to 7.2 Tbps for C-band.

For neural networks, the ability to handle high-dimensional input is crucial, especially for applications like image classification (with large pixel count) [30] or natural language processing (NLP) where a large vector space is required to

learn associations between words [31]. Increased channel count using PhCs directly translates to photonic neural networks being amenable to applications with high-dimensional inputs. While dimensionality reduction techniques are often used in neural network models to ease the computational cost, training time and prevent overfitting, the dimensions are still beyond what MRR-based photonic systems would be able to handle. Without a scalable solution like PhCs, applications with high input dimensions would not be feasible. Another consideration for scalability is the transmission loss. For PhC nanobeams with transmission $\alpha = 97\%$ [32] in an 8-bit MAC operation, the minimal optical power required by a photodetector would be $\frac{N_e \cdot e}{\Delta \tau \cdot R}$ where $N_e = 2^8$ electrons, $R = 0.9$. With an input power of 10 dBm, we estimate the channel count to be 350, which is larger than the limit of 130 within the C-band.

Energy consumption is often cited as the impetus to novel neural network hardware, in particular photonics [2]. It is also estimated that the majority of energy consumption in a neural network stems from the energy costs of synapses, since they scale as n^2 whereas the nonlinear unit scales as n for n neurons [12]. Energy consumption in synapses is the energy associated with tuning and holding the weight value. We experimentally demonstrated that a PhC had about 4.5 times higher thermal tuning efficiency than an MRR (refer Fig. 8). Energy consumption per weighting operation, in units of J/bits, can be calculated as the power consumed divided by the signal speed. With higher tuning efficiency, we can expect proportionately lower energy consumption per weighting operation. The overall energy reduction would then scale with the size of the neural network. The energy consumption values for a single PhC-based synapse and an MRR-based synapse, based on reported thermal tuning efficiency values in prior literature and a bit rate of 10 G, are shown in Table I. Energy efficiency for any index tuning in a photonic cavity depends on the interaction between light and the medium. For photonic cavities, this relationship can be quantified by the Purcell factor, F_p , defined as

$$F_p \approx \lambda_0^3 \frac{Q}{V} \quad (1)$$

where λ_0 is the cavity resonant wavelength, Q is the cavity quality factor and V is the cavity mode volume. Let Δ_p be the ratio of the tuning efficiency of a PhC to that of an MRR. We can now calculate Δ_p as:

$$\Delta_p = \frac{F_{p,\text{PhC}}}{F_{p,\text{MRR}}} = \frac{Q_{\text{PhC}}}{Q_{\text{MRR}}} \frac{V_{\text{MRR}}}{V_{\text{PhC}}}. \quad (2)$$

If we can assume $Q_{\text{PhC}} = Q_{\text{MRR}}$, the tuning efficiency of PhC is proportionately reduced with the cavity mode volume relative to an MRR. Since we had $Q_{\text{PhC}} = \frac{1}{20} Q_{\text{MRR}}$ in our experiment, the tuning efficiency improvement was not too significant. PhC cavities with $Q \approx 10^7$ have been demonstrated before [20], meaning PhCs should offer orders of magnitude higher tuning efficiency relative to MRRs. Another point to note is that this energy efficiency improvement translates to any index tuning mechanism, for example carrier modulation [33].

Finally, we compare MRRs and PhCs against another important metric, compute density. It is defined as the numbers of

operations done per unit time per unit MAC area, where MAC refers to the multiply-and-accumulate operation in a synapse unit [12]. In simple terms, it is a measure how physically compact can a processor be. It is also useful as a standard metric to compare different types of neural hardware (photonic, digital and analog electronics), which vary greatly in size and multiplexing techniques. In [12], the authors estimated that sub- λ photonics can achieve 100 times higher compute density than MRR-based architectures. For a typical MRR of radius $7 \mu\text{m}$ [19] and a sampling speed of 10 GS/s, its compute density would be 100 tera MAC/smm². For the PhC filter presented here, with footprint of $13 \mu\text{m}^2$ and the same sampling speed, the compute density is 770 tera MAC/smm², about 8 times higher than an MRR. It can be further improved by reducing the PhC footprint, nanobeams of only $2 \mu\text{m}$ have been demonstrated [24]. The compute density would eventually be limited by the inter-channel thermal crosstalk. However, clever design strategies can be employed to alleviate this issue. Firstly, since the optical mode is confined to mostly the central section of the PhC (refer to Fig. 4(e)), the heater needs to be only as long as the optical mode confinement length. Shorter heaters would then minimize inter-channel thermal crosstalk. Secondly, isolation trenches can disrupt the thermal conduction pathway and reduce the likelihood of crosstalk [34].

While our results are promising, further investigation will be necessary to enable photonic crystals as synapses in photonic neural networks. For example, it is worthwhile to look into ways to increase the PhC quality factor which will improve the tuning energy efficiency. Next, we can test the synapse architecture in Fig. 1(c), specifically, demonstrate multi-wavelength weighting by PhCs, summing by a photodetector and a nonlinear transformation by an electro-optic modulator. Also, the current PhC synapse architecture, as shown in Fig. 1(c), only allows for positive weights, unlike the add-drop MRR synapse architecture which permits both positive and negative weights. Positive weights suffice for typical artificial neural networks, but spiking networks require negative weights for inhibition [3]. Future work can explore improved PhC synapse architecture that can allow for negative weights in spiking networks. Finally, the tuning speed of thermal heaters is limited by the heater-waveguide physical separation to 100 kHz [35] which limits the weight reconfiguration speed. For higher reconfiguration speeds, electro-optic modulation schemes using carrier-based tuning on PhCs will need to be explored [24].

VI. CONCLUSION

High bandwidth of photonics is what makes it incredibly powerful, both for signal communication and processing. To take advantage of the bandwidth, photonic systems often implement wavelength division multiplexing (WDM). Prior work on WDM-based photonic neural hardware uses microring resonators as the synapses, which fundamentally have an upper limit on the channel count, that prohibits network scale. We propose a photonic crystal nanobeam cavity based synapse as the most scalable synapse solution for photonic neural networks. We have experimentally demonstrated superior performance merits

of PhCs pertinent to synaptic weighting including: first, an FSR-free operation, which results in 4 times higher channel density within the C-band and second, a 4.5 times higher weight tuning efficiency resulting from its smaller mode volume. Increased channel count improves the system throughput as well as makes the hardware amenable to applications with high-dimensional input like natural language processing. Another benefit from the smaller footprint of PhC is in compute density, a measure of processing done per unit physical area, which we estimate to be about 8 times higher compute density for our current design. Further work will be dedicated to cavity design optimization, and development of calibration and control procedures for practical applications of such weight banks.

REFERENCES

- [1] D. Amodei and D. Hernandez, "AI and compute," May 2018 [Online]. Available: <https://openai.com/blog/ai-and-compute>
- [2] B. J. Shastri Prucnal *et al.*, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, no. 2, pp. 102–114, 2021.
- [3] P. R. Prucnal and B. J. Shastri, *Neuromorphic Photonics*. Boca Raton, FL, USA: CRC Press, 2017.
- [4] H. Ishio, J. Minowa, and K. Nosu, "Review and status of wavelength-division-multiplexing technology and its application," *J. Lightw. Technol.*, vol. 2, no. 4, pp. 448–463, 1984.
- [5] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, 2017.
- [6] A. N. Tait *et al.*, "Microring weight banks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 22, no. 6, pp. 312–325, Nov./Dec. 2016.
- [7] J. Feldmann *et al.*, "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, no. 7840, pp. 52–58, 2021.
- [8] X. Cao *et al.*, "Mesh-structure-enabled programmable multitask photonic signal processor on a silicon chip," *ACS Photon.*, vol. 7, no. 10, pp. 2658–2675, 2020.
- [9] A. N. Tait *et al.*, "Silicon photonic modulator neuron," *Phys. Rev. Appl.*, vol. 11, no. 6, 2019, Art. no. 064043.
- [10] N. Eid *et al.*, "FSR-free silicon-on-insulator microring resonator based filter with bent contra-directional couplers," *Opt. Exp.*, vol. 24, no. 25, pp. 29009–29021, 2016.
- [11] A. N. Tait *et al.*, "Two-pole microring weight banks," *Opt. Lett.*, vol. 43, no. 10, pp. 2276–2279, 2018.
- [12] M. A. Nahmias *et al.*, "Photonic multiply-accumulate operations for neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan./Feb. 2020, Art. no. 7701518.
- [13] J. D. Joannopoulos, R. Meade, and J. N. Winn, "Photonic crystals," *Molding Flow Light*, vol. 2, pp. 122–134, 2008.
- [14] X. Yang, C. Husko, C. W. Wong, M. Yu, and D.-L. Kwong, "Observation of femtojoule optical bistability involving fano resonances in high- q /v m silicon photonic crystal nanocavities," *Appl. Phys. Lett.*, vol. 91, no. 5, 2007, Art. no. 051113.
- [15] T. Tanabe, M. Notomi, S. Mitsugi, A. Shinya, and E. Kuramochi, "All-optical switches on a silicon chip realized using photonic crystal nanocavities," *Appl. Phys. Lett.*, vol. 87, no. 15, 2005, Art. no. 151112.
- [16] J. Zhang *et al.*, "Generating fano resonances in a single-waveguide silicon nanobeam cavity for efficient electro-optical modulation," *ACS Photon.*, vol. 5, no. 11, pp. 4229–4237, 2018.
- [17] C. Gui, Y. Zhang, J. Du, J. Xia, and J. Wang, "Experimental demonstration of analog signal transmission in a silicon photonic crystal L3 resonator," *Opt. Exp.*, vol. 23, no. 11, pp. 13916–13923, 2015.
- [18] H. Zhou *et al.*, "Compact, submilliwatt, 2×2 silicon thermo-optic switch based on photonic crystal nanobeam cavities," *Photon. Res.*, vol. 5, no. 2, pp. 108–112, 2017.
- [19] A. N. Tait, T. F. De Lima, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Multi-channel control for microring weight banks," *Opt. Exp.*, vol. 24, no. 8, pp. 8895–8906, 2016.
- [20] P. B. Deotare, M. W. McCutcheon, I. W. Frank, M. Khan, and M. Lončar, "High quality factor photonic crystal nanobeam cavities," *Appl. Phys. Lett.*, vol. 94, no. 12, 2009, Art. no. 121106.
- [21] Q. Quan, P. B. Deotare, and M. Loncar, "Photonic crystal nanobeam cavity strongly coupled to the feeding waveguide," *Appl. Phys. Lett.*, vol. 96, no. 20, 2010, Art. no. 203102.
- [22] G. Dong, W. Deng, J. Hou, L. Chen, and X. Zhang, "Ultra-compact multi-channel all-optical switches with improved switching dynamic characteristics," *Opt. Exp.*, vol. 26, no. 20, pp. 25630–25644, 2018.
- [23] T. F. deLima *et al.*, "Design automation of photonic resonator weights," *Nanophotonics*, pp. 1–18, 2022, doi: [10.1515/nanoph-2022-0049](https://doi.org/10.1515/nanoph-2022-0049).
- [24] J. Hendrickson, R. Soref, J. Sweet, and W. Buchwald, "Ultrasensitive silicon photonic-crystal nanobeam electro-optical modulator: Design and simulation," *Opt. Exp.*, vol. 22, no. 3, pp. 3271–3283, 2014.
- [25] L. Zhou *et al.*, "Miniature silicon nanobeam resonator tuned by GST phase change material," in *Proc. Conf. Lasers Electro- Opt. Pacific Rim*, 2018, pp. 1–2.
- [26] M. R. Billah *et al.*, "Hybrid integration of silicon photonics circuits and InP lasers by photonic wire bonding," *Optica*, vol. 5, no. 7, pp. 876–883, 2018.
- [27] T. de Lima and A. Tait, "Lightlab – laboratory instrumentation and automation," 2018. [Online]. Available: <https://github.com/lightwave-lab/lightlab>
- [28] M. M. Milosevic *et al.*, "Ion implantation in silicon for trimming the operating wavelength of ring resonators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 4, Jul. 2018, Art. no. 8200107.
- [29] Q. Li *et al.*, "Single microring-based 2×2 silicon photonic crossbar switches," *IEEE Photon. Technol. Lett.*, vol. 27, no. 18, pp. 1981–1984, Sep. 2015.
- [30] V. Bangari *et al.*, "Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs)," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan./Feb. 2020, Art. no. 7701213.
- [31] Y. Li and T. Yang, "Word embedding for understanding natural language: A survey," in *Guide to Big Data Applications*. Berlin, Germany: Springer, 2018, pp. 83–104.
- [32] Q. Quan and M. Loncar, "Deterministic design of wavelength scale, ultra-high Q photonic crystal nanobeam cavities," *Opt. Exp.*, vol. 19, no. 19, pp. 18529–18542, 2011.
- [33] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators," *Opt. Exp.*, vol. 15, no. 2, pp. 430–436, 2007.
- [34] P. Dong *et al.*, "Low power and compact reconfigurable multiplexing devices based on silicon microring resonators," *Opt. Exp.*, vol. 18, no. 10, pp. 9852–9858, 2010.
- [35] P. B. Deotare, L. C. Kogos, I. Bulu, and M. Lončar, "Photonic crystal nanobeam cavities for tunable filter and router applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 19, no. 2, Mar./Apr. 2013, Art. no. 3600210.

Aashu Jha received the B.S. degree in physics and math from Bates College, Lewiston, ME, USA in 2017, and the M.A. degree in electrical engineering in 2019 from Princeton University, Princeton, NJ, USA, where she is currently working toward the Ph.D. degree advised by Prof. Paul R. Prucnal with Princeton University. Her research interests include nonlinear optics, optical signal processing, neuromorphic photonics, and photonic integrated circuits.

Chaoran Huang received the B.Eng. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 2016. Before joining CUHK, she was a Postdoctoral Research Fellow with Princeton University, Princeton NJ, USA, in 2017 to 2021. She has authored more than 40 peer-reviewed research papers, three book chapters, and one U.S. patent. Her research interests include silicon photonics, photonic integrated circuits, and nonlinear optics, serving applications including neuromorphic computing, computing security, real-time information processing, and optical interconnects. Her current research interests include developing neuromorphic photonic platforms, including novel devices, photonic integrated circuits, and complementary algorithms, for high-performance AI computing. She has been the TPC Member of several international conferences and is a frequent reviewer for different journals in IEEE, OSA, and the Nature Publishing Group. She was the recipient of 2019 Rising Stars Women in Engineering Asia, and was nominated by Princeton University to compete for Blavatnik Regional Awards for Young Scientists.



Thomas Ferreira deLima received the B.Sc. and the Ingénieur Polytechnicien master's degrees from Ecole Polytechnique, Palaiseau, France, in 2016 with a focus on physics for optics and nanosciences. He received the Ph.D. degree in electrical engineering in 2022 with the Lightwave Communications Research Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. He is currently a Researcher with the NEC Laboratories America, Inc., Princeton, NJ, USA. His research interests

include: real-time integrated photonic systems, nonlinear signal processing with photonic devices, spike-timing based processing, ultrafast cognitive computing, and dynamical light-matter neuro-inspired learning and computing.

Hsuan-Tung Peng received the B.S. degree in physics from National Taiwan University, Taipei, Taiwan, in 2015, and the M.A. degree in electrical engineering in 2018 from Princeton University, Princeton, NJ, USA, where he is currently working toward the Ph.D. degree. His research interests include neuromorphic photonics, photonic integrated circuits, and optical signal processing.

Bhavin Shastri (Senior Member, IEEE) received the Honors B.Eng. (with distinction), M.Eng., and Ph.D. degrees in electrical engineering (photonics) from McGill University, Montreal, QC, Canada, in 2005, 2007, and 2012, respectively. He is currently an Assistant Professor of engineering physics with Queen's University, Kingston, ON, Canada, and a Faculty Affiliate with the Vector Institute for Artificial Intelligence, Canada. He was an NSERC and Banting Postdoctoral Fellow (2012–2016) and an Associate Research Scholar (2016–2018) with Princeton University. He has authored or coauthored more than 70 journal articles and 100 conference proceedings, seven book chapters, and given more than 60 invited talks and lectures including five keynotes and three tutorials. His research interests include silicon photonics, photonic integrated circuits, neuromorphic computing, and machine learning. He is a co-author of the book, *Neuromorphic Photonics* (Taylor Francis, CRC Press, 2017), a term he helped to coin. Dr. Shastri was the recipient of the 2022 SPIE Early Career Achievement Award and the 2020 IUPAP Young Scientist Prize in Optics for his pioneering contributions to neuromorphic photonics from the ICO. He is a Senior Member of Optica (formerly OSA), 2014 Banting Postdoctoral Fellowship from the Government of Canada, the 2012 D. W. Ambridge Prize for the top graduating Ph.D. student at McGill, an IEEE Photonics Society 2011 Graduate Student Fellowship, a 2011 NSERC Postdoctoral Fellowship, a 2011 SPIE Scholarship in Optics and Photonics, a 2008 NSERC Alexander Graham Bell Canada Graduate Scholarship, including the Best Student Paper Awards at the 2014 IEEE Photonics Conference, 2010 IEEE Midwest Symposium on Circuits and Systems, the 2004 IEEE Computer Society Lance Stafford Larson Outstanding Student Award, and the 2003 IEEE Canada Life Member Award.

Paul R. Prucnal (Life Fellow, IEEE) received the A.B. degree (graduating summa cum laude) in mathematics and physics from Bowdoin College, Brunswick, ME, USA, and the M.S., M.Phil. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, USA. After his Doctorate, he joined the Faculty with Columbia University, New York, NY, USA, where, he was a Member of the Columbia Radiation Laboratory, he performed groundbreaking work in OCDMA and self-routed photonic switching. In 1988, he joined the Faculty with Princeton University, Princeton, NJ, USA. He has authored or coauthored more than 350 journal articles and book chapters and holds 28 U.S. patents. His research on optical CDMA initiated a new research field in which more than 1000 papers have since been published, exploring applications ranging from information security to communication speed and bandwidth. In 1993, he invented the Terahertz Optical Asymmetric Demultiplexer, the first optical switch capable of processing terabit per second pulse trains. He is the Author of the book *Neuromorphic Photonics* and the Editor of the book *Optical Code Division Multiple Access: Fundamentals and Applications*. He was the Area Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He is a Fellow of the Optical Society of America and the National Academy of Inventors, and a Member of honor societies including Phi Beta Kappa and Sigma Xi. He was the recipient of the 1990 Rudolf Kingslake Medal for his paper titled Self-routing photonic switching with optically-processed control, recipient of the Gold Medal from the Faculty of Mathematics, Physics and Informatics at the Comenius University, for leadership in the field of Optics 2006 and has received multiple teaching awards at Princeton, including the E- Council Lifetime Achievement Award for Excellence in Teaching, the School of Engineering and Applied Science Distinguished Teacher Award, The President's Award for Distinguished Teaching. He has been instrumental in founding the field of Neuromorphic Photonics and developing the photonic neuron, a high-speed optical computing device modeled on neural networks, as well as integrated optical circuits to improve wireless signal quality by cancelling radio interference.