

Demonstration of WDM weighted addition for principal component analysis

Alexander N. Tait,* John Chang, Bhavin J. Shastri, Mitchell A. Nahmias, and Paul R. Prucnal

Princeton University, Princeton, NJ 08544, USA

[*atait@princeton.edu](mailto:atait@princeton.edu)

Abstract: We consider an optical technique for performing tunable weighted addition using wavelength-division multiplexed (WDM) inputs, the enabling function of a recently proposed photonic spike processing architecture [J. Lightwave Technol., 32 (2014)]. WDM weighted addition provides important advantages to performance, integrability, and networking capability that were not possible in any past approaches to optical neurocomputing. In this letter, we report a WDM weighted addition prototype used to find the first principal component of a 1Gbps, 8-channel signal. Wideband, multivariate techniques have immediate relevance to modern radio systems, and photonic spike processing networks enabled by WDM could open new domains of information processing that bring unprecedented bandwidth and intelligence to problems in radio communications, ultrafast control, and scientific computing.

© 2015 Optical Society of America

OCIS codes: (070.1170) Analog optical signal processing; (200.4700) Optical neural systems; (060.5625) Radio frequency photonics; (200.3050) Information processing.

References and links

1. 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).
2. J. Hasler and H. B. Marr, "Finding a roadmap to achieve large neuromorphic hardware systems," Frontiers in Neuroscience **7** (2013).
3. 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).
4. S. Friedmann, N. Frémaux, J. Schemmel, W. Gerstner, and K. Meier, "Reward-based learning under hardware constraints - using a RISC processor embedded in a neuromorphic substrate," Frontiers in Neuroscience **7** (2013).
5. D. Liang, G. Roelkens, R. Baets, and J. E. Bowers, "Hybrid integrated platforms for silicon photonics," Materials **3**, 1782–1802 (2010).
6. G. Roelkens, L. Liu, D. Liang, R. Jones, A. Fang, B. Koch, and J. Bowers, "III-V/silicon photonics for on-chip and intra-chip optical interconnects," Laser Photonics Rev. **4**, 751–779 (2010).
7. D. A. B. Miller, "The role of optics in computing," Nat. Photonics **4**, 406–406 (2010).
8. J. Misra and I. Saha, "Artificial neural networks in hardware: a survey of two decades of progress," Neurocomputing **74**, 239–255 (2010).
9. 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** (2014).
10. 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).

11. 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).
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** (2013).
13. 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).
14. B. J. Shastri, M. A. Nahmias, A. N. Tait, B. Wu, and P. R. Prucnal, "Simpel: circuit model for photonic spike processing laser neurons," under review, *Opt. Express*. available arXiv:1409.7030 (2014).
15. M. C. Soriano, S. Ortín, D. Brunner, L. Larger, C. R. Mirasso, I. Fischer, and L. Pesquera, "Optoelectronic reservoir computing: tackling noise-induced performance degradation," *Opt. Express* **21**, 12–20 (2013).
16. T. V. Vaerenbergh, M. Fiers, P. Mechet, T. Spuesens, R. Kumar, G. Morthier, B. Schrauwen, J. Dambre, and P. Bienstman, "Cascadable excitability in microrings," *Opt. Express* **20**, 20292–20308 (2012).
17. D. Woods and T. J. Naughton, "Optical computing: photonic neural networks," *Nat. Phys.* **8**, 257–259 (2012).
18. B. J. Shastri, M. A. Nahmias, A. N. Tait, Y. Tian, B. Wu, and P. R. Prucnal, "Graphene excitable laser for photonic spike processing," in "Proc. IEEE Photonics Conf. (IPC)," (paper PD.4, Seattle, WA, USA, 2013), pp. 1–2.
19. S. Furber, D. Lester, L. Plana, J. Garside, E. Painkras, S. Temple, and A. Brown, "Overview of the SpiNNaker system architecture," *IEEE Trans. Comput.* **62**, 2454–2467 (2013).
20. J. Chang, J. Meister, and P. Prucnal, "Implementing a novel highly scalable adaptive photonic beamformer using 'blind' guided accelerated random search," *J. Lightwave Technol.* **32**, 3623–3629 (2014).
21. M. Chang, A. Tait, J. Chang, and P. Prucnal, "An integrated optical interference cancellation system," in "Wireless and Optical Communication Conference (WOCC), 2014 23rd," (2014), pp. 1–5.
22. J. W. Goodman, "Fan-in and fan-out with optical interconnections," *Opt. Acta* **32**, 1489–1496 (1985).
23. J. Schemmel, J. Fieries, and K. Meier, "Wafer-scale integration of analog neural networks," in "Neural Networks, 2008. IJCNN 2008. IEEE International Joint Conference on," (2008), pp. 431–438.
24. B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, "Photonic spike processing: ultrafast laser neurons and an integrated photonic network," *IEEE Pho. Soc. Newsletter* **28**, 4–11 (2014).
25. 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," (Springer Berlin Heidelberg, 2014), pp. 183–222.
26. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT, 2012).
27. S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nat. Phys.* **10**, 631–633 (2014).
28. M.-E. Baylor, "Analog optoelectronic independent component analysis for radio frequency signals," Ph.D. thesis, University of Colorado (2007).
29. E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *J. Neurosci.* **2**, 32–48 (1982).
30. E. Oja, "Simplified neuron model as a principal component analyzer," *J. Math. Biol.* **15**, 267–273 (1982).
31. J. Capmany, B. Ortega, and D. Pastor, "A tutorial on microwave photonic filters," *J. Lightwave Technol.* **24**, 201–229 (2006).
32. J. Chang, M. Fok, R. Corey, J. Meister, and P. Prucnal, "Highly scalable adaptive photonic beamformer using a single mode to multimode optical combiner," *IEEE. Microwave Wireless Compon. Lett.* **10**, 563–565 (2013).

1. Introduction

Unconventional computing techniques that are neuromorphic (i.e. biological neuron-inspired) have attracted renewed interest due, in part, to incipient plateaus in power dissipation and clock speed of conventional computers [2]. The conventional combination of von Neumann architecture, digital coding, and microelectronic implementation may never find equal in terms of procedural, calculation-based tasks; however, many applications (e.g. pattern analysis, optimization, learning) demand capabilities that far exceed the conventional roadmap [3, 4].

At the same time, photonic integrated circuit (PIC) manufacturing is undergoing a coming of age with silicon photonics technologies [5, 6]. While driven by a demand for intra-chip optical communication links, this manufacturability inherently affords new room for ideas in large-scale optical computing, even though digital and sequential optical logic still face fundamental barriers [7]. Neuromorphic approaches to optical computing have historically tended to focus on spatially-multiplexed (e.g. holographic) interconnects, yet suffered practical barriers in scalability, relevance, and manufacturability [8]. Some of these barriers could be avoided if unconventional photonic circuits could be made from mainstream device sets and focus on

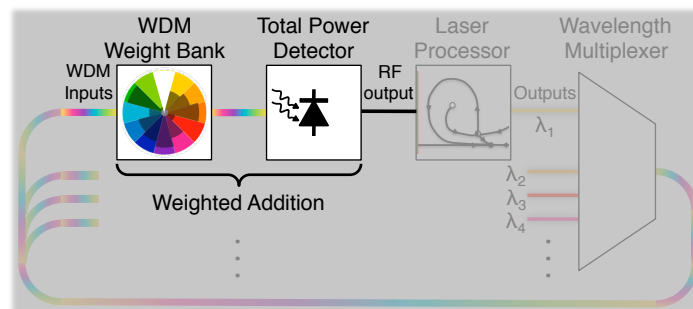


Fig. 1. WDM weighted addition in a broadcast-and-weight spike processing network, from [1]. Weighted addition is a central function in neural networks. The optical WDM version consists of a tunable spectral weight bank and power detector.

solving problems otherwise impossible with current and future electronics. A new generation of neuron-inspired optical systems has experienced a surge in interest [9–11].

Many modern photonic neuro-inspired processing approaches exploit similarities in laser and neuron dynamics (termed “spiking” dynamics) and aim for significant (~ 8 order) speed increases over electronic counterparts [12–18]. From an applications standpoint, the 10GHz bandwidth range is a fertile regime for new ideas in computing because of the increasing demand for radio frequency (RF) systems that are both wideband and intelligent (i.e. complex and adaptive). Microelectronic techniques for neuron-inspired processing at biological speeds have great difficulty extending to faster timescales, in part due to interconnection limitations [19]. So far, most research on photonic neurons has focused on single laser dynamics without proposing a solution for interconnecting multiple laser neurons. Neural networking has a prominent many-to-one aspect, which is accomplished through a reconfigurable linear operation called weighted addition.

A photonic architecture was recently proposed for neuron-inspired processing and networking using standard PIC components [1]. This proposal relied heavily on wavelength-division multiplexing (WDM) for both network routing and weighted addition computations (Fig. 1). WDM signals are weighted by a reconfigurable spectral filter and detected together in a single photodiode (PD), whose electronic output represents their sum. WDM total power detection effectively strips wavelength and channel information, a fact that, while counterproductive in optical communication, has found use in alternative contexts [20, 21] because it efficiently avoids undesirable coherent effects traditionally associated with optical summation through fan-in [22].

Compared to electronic counterparts, WDM weighted addition promises significantly improved interconnect performance – characterized by bandwidth and fan-in degree (i.e. the number of inputs to each node). Digital electronic implementations that commonly use time-division multiplexing (TDM) to accumulate summands face an undesirable tradeoff between fan-in and effective signal bandwidth [3, 19]. For this reason, they are largely constrained to operate on kHz timescales, with one notable exception [23]. In [1], WDM weighted addition was estimated to be capable of 34 channels at 10GHz, using high-Q microring resonator filters that are ubiquitous components of PIC platforms. With advances in neuromorphic engineering, PIC manufacturing, and laser dynamics research, WDM-based processing networks represent a promising approach to core problems in unconventional and optical computing [24, 25].

In this paper, we present an experimental demonstration of WDM weighted addition and use this prototype to perform principal component analysis (PCA) on 8 partially correlated 1Gbps

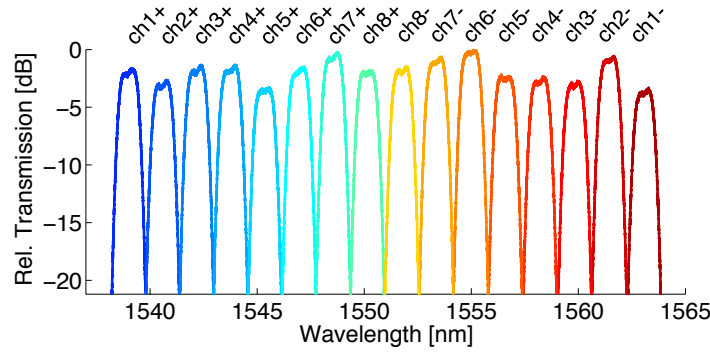


Fig. 2. Measured individual transmission spectra of the WDM weight bank $[W_{1...16}(\lambda)]$, where each weight is tuned to maximum transmission, with all others minimized. Inter-channel cross-talk is less than 20dB. Filters for (+) channels (blue-green) and those for (−) channels (red-yellow) are labeled as such because they act on complementarily modulated pairs of signals.

inputs. PCA is a very general technique for finding patterns in and reducing the dimensionality of multivariate data without *a priori* knowledge. The first PC output is the projection of the data onto their vector of greatest variance, which can be considered the most informative single basis. PCA and its variants are ubiquitous in machine learning [26, 27], cognitive radio [28], and computational neuroscience [29]. While the present work does not approach the theoretical performance limits of WDM-based PCA, it establishes an experimental proof-of-principle of multiwavelength statistical methods applied to RF photonic devices. Arrayed-antenna systems in particular present a challenge of digitizing many signals which are largely redundant. Statistical techniques for dimensionality reduction, implemented in the analog domain, could lead to a greatly reduced strain on digital signal processing requirements in wideband, multi-antenna RF systems.

2. Methods

The WDM weighted addition prototype (Fig. 3(b)) applies weights to a 16-wavelength signal using arrayed waveguide grating (AWG) (de)multiplexers and a bank of tunable optical attenuators (Enablence iVOA 1600). Figure 2 shows the power transmission spectra formed by zeroing all but one filter channel at a time, $W_{1...2d}(\lambda)$, where the number of effective channels is d (8 in this work). Two wavelengths, complementarily modulated, are required to represent each channel, in order to enable positive/negative weighting. The overall spectral response, H , depends on the weight vector, $\boldsymbol{\mu}$, such that

$$H(\lambda, \boldsymbol{\mu}) = \sum_{i=1}^d |\mu_i| W_{k[i]}(\lambda) \quad (1a)$$

$$k[i] = \begin{cases} i, & \text{if } \mu_i \geq 0 \\ 2d + 1 - i, & \text{if } \mu_i < 0 \end{cases} \quad (1b)$$

where $k[i]$ picks between wavelengths corresponding to positively and negatively modulated versions of x_i , depending on the sign of μ_i . The PD then responds to total optical power, yielding an electronic representation of the weighted sum: $m(t) = \boldsymbol{\mu} \cdot \mathbf{x}$.

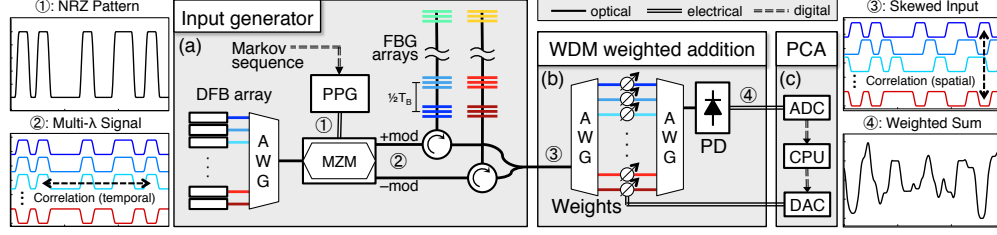


Fig. 3. Experimental setup. (a) Input generator, where DFB: distributed feedback laser; AWG: arrayed-waveguide grating multiplexer; PPG: pulse pattern generator; MZM: Mach-Zehnder modulator; FBG: fiber Bragg grating. (b) WDM weighted addition where circles are variable optical attenuators, and PD: photodiode, and (c) PCA algorithm, where ADC: analog-digital converter; CPU: central processor; DAC: digital-analog converter. Illustrations show signals at various points in the circuit: ① RF input to the system that exhibits temporal autocorrelation, ② Modulated WDM optical signals – complementary modulations are provided by the MZM, ③ λ -dependent delays transform initial temporal correlation into instantaneous spatial correlations, ④ electrical output of the PD representing the weighted sum of correlated channels.

The CPU receives samples of the PD output at 4GS/s and updates the attenuator tuning. During adaptation, the CPU converges to the first principal component by updating weights according to the well-known iterative Hebbian learning rule with normalization [29, 30]:

$$\Delta\mu_i(n) = \gamma \langle x_i(t) \cdot m(t, n) \rangle_t \quad (2a)$$

$$\mu_i^*(n) = \mu_i(n) + \Delta\mu_i(n) \quad (2b)$$

$$\mu_i(n+1) = \frac{\mu_i^*(n)}{\sum_{j=1}^d \mu_j^*(n)^2} \quad (2c)$$

where $\Delta\mu(n)$ is the update vector of epoch n , $x_i(t)$ is the stored input i , $m(t, n)$ is the measured output of the epoch, and γ is a positive constant. $\langle \cdot \rangle_t$ signifies an average over the samples of an epoch (16k samples in this experiment). Normalization is applied to enforce $\|\mu(n+1)\| = 1$. Higher-order PCs could be obtained with identical weighted addition circuits working in parallel. The second PC can be viewed as the first PC of the data projected into the orthogonal subspace of the lower PC, and so on. Therefore, normalization, Eq. (2c), would be replaced with orthonormalization. In vector notation,

$$\mu_1(n+1) = \frac{\mu_1^*}{\|\mu_1^*\|} \quad (3a)$$

⋮

$$\mu_k(n+1) = \frac{\mu_k^* - \sum_{h=1}^{k-1} \text{Pr}_{\mu_h}(\mu_k^*)}{\|\mu_k^* - \sum_{h=1}^{k-1} \text{Pr}_{\mu_h}(\mu_k^*)\|} \quad (3b)$$

where μ_k is the k th-order PC vector, and $\text{Pr}_b(a)$ is the vector projection of a onto b .

The input signals $x_i(t)$ needed in Eq. (2a) are not measured concurrently, but rather stored in memory. Prior to the adaptation phase, each input is measured sequentially by zeroing all but one weight at a time, thus presenting the transmission spectra from Fig. 2. This serial pre-measurement requires a trigger from the pattern source; however, it is economically scalable,

requiring only one detector and ADC regardless of the number of channels under test. Additionally, this approach guarantees that inputs $x_i(t)$ and outputs $m(t, n)$ are sampled in a common time basis, allowing for accurate calculation of input-output correlation, Eq. (2a), without over-all delay or fading calibrations.

PCA takes advantage of statistical redundancy between variables so is trivial with identical (i.e. perfectly correlated) inputs. To test the WDM weighted addition and PCA system, we constructed an input generation circuit (Fig. 3(a)) that affords continuous control of the *partial* correlations between multiwavelength signals, using a single pulse pattern generator (PPG) and Mach-Zehnder modulator (MZM). The MZM produces complementary modulations of a single 1Gbps non-return-to-zero (NRZ) signal onto 16 wavelength carriers. After modulation, fiber Bragg grating (FBG) arrays impart wavelength-dependent time-of-flight delays. Since the optical path to each FBG is different, channels become skewed by one bit period (1.0 ns) per channel. This time skew has the effect of transforming temporal autocorrelation of the original PPG signal to instantaneous inter-channel correlation. To easily parameterize temporal autocorrelation, we use a Markov chain model, wherein subsequent bits have a $(0.5 + \alpha)$ probability of being the same. When given a Markov chain, the FBG time skew yields a partially correlated, multiwavelength signal that is suitable for PCA.

Although the Markov process is digital, its α parameter provides tight control of the continuous covariant statistics between channels. Figure 4(a) shows a subset of partially correlated positive input channels and their negative complements on other wavelengths. In Figure 4(c), two of these representative channels are plotted against one another in order to visualize their time-averaged correlation. Since the ADC clock is not synchronized with the input pattern, it has some chance of sampling during a transition of the NRZ signal; however, the greater likelihood of sampling during stable times results in a visible 4-point constellation. Figure 4(d) indicates that the instantaneous analog correlation between multiwavelength inputs is proportionally controlled by α , even though α parameterizes a discrete stochastic process.

3. Results and discussion

Once a multiwavelength signal with controllable inter-channel correlations is generated, a PCA algorithm can converge repeatably to a well-defined first PC. Figure 4(b) shows the measured output of the WDM weighted addition circuit after PCA convergence. For this experiment, the iteration count was fixed at 40, although convergence typically occurred within 15 epochs, depending on algorithm parameters such as γ and epoch duration. The measured PC is compared to the PC calculated offline by a software-based non-iterative singular value decomposition (SVD) method. These signals are plotted against one another, showing time-averaged density in Fig. 4(e). The correlation of measured and calculated PCs are plotted versus α in Fig. 4(f). As should be expected, performance is worse and more variable for less-correlated signals around $\alpha = 0$ because the principal component basis becomes ill-defined when inputs are uncorrelated.

Non-idealities in the results are likely caused by electronic and optical amplifiers. Firstly, the minimum of the curve in Fig. 4(f) is biased away from $\alpha = 0$. This could be due to frequency-dependent fading in the RF amplifier following the PD, band limited at 1.3GHz. Bit sequences with $\alpha < 0$ have increased spectral power outside of this bandwidth, thereby experiencing greater distortion. Secondly, the expected dip in accuracy at $\alpha = 0$ does not reach 0, which could be due to impedance mismatches causing overshoot and ringing, which are visible in Fig. 4(a). These artifacts can introduce analog redundancies to otherwise uncorrelated signals, thereby spawning unintentional PCs. Finally, imperfect agreement between calculation and measurement for $\alpha \neq 0$ is likely caused by slow-timescale cross-saturation in an optical amplifier following the weight bank, which results in an artifactual weight-dependent gain to which PCA algorithms are sensitive.

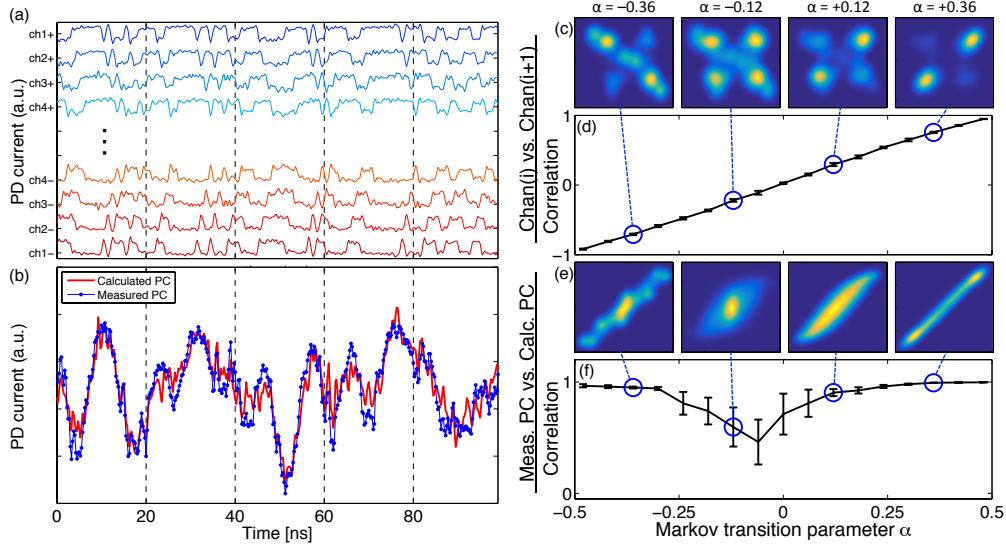


Fig. 4. (a-b) A 100ns time window of a typical epoch. (a) 4 positive and 4 negative inputs with channel-dependent delays. The generating pattern is a Markov sequence (length $2^{13} - 1$) with transition parameter $\alpha = +0.3$. (b) First principal component output as calculated by a software matrix decomposition-based PCA (red) compared to the measured output after convergence of the iterative algorithm, Eq. (2a), (blue). Both calculated and measured PCA algorithms are applied to the measured inputs. (c-f) Statistical analysis of performance over a range of generating Markov parameters. (c) Probability density plots of input channel (i) vs. adjacent channel (i+1), and (d) corresponding correlation values, showing repeatable, proportional dependence of inter-channel correlation on the Markov parameter α . (e) Probability density plots of measured vs. calculated PC outputs, and (f) corresponding correlation values, indicating the converged accuracy of WDM weighted addition. Error bars in (d) and (f) represent standard deviation over 10 different sequences generated with the same transition parameters.

Many techniques for RF photonic filtering [31], beamforming [32], and other applications can handle high-bandwidth analog signals, but most lack control algorithms that can tune system parameters fast enough to perform online analysis in changing environments. A further direction for research is decreasing epoch time using iterative unsupervised learning rules from computational neuroscience, such as Hebbian and its stable contemporaries [29,30]. Compared to matrix-based SVD algorithms, the simple pair-wise operations required for a bio-inspired PCA controller, as in Eq. (2a), are more feasible for a co-integrated microelectronic processor, or perhaps even other analog and/or optoelectronic hardware.

4. Conclusion

In this paper, we have presented an experimental prototype for WDM weighted addition on 8 effective channels at 1Gbps and assessed performance with a PCA task, which involved development of novel methodologies for generating partially correlated multiwavelength signals, which could scale to test future prototypes with more channels and higher bandwidths. In addition to improving performance, further work could focus on integration or on accelerating epoch updates. A theoretical analysis of the limits of weighted addition in optical and electronic implementations is also called for. Ultimately, high speed linear functions that are compatible

with photonic integration trends could constitute an important piece of future RF systems, either directly, or as an element of larger processing-networks, such as photonic spike processors.