Subwavelength neuromorphic nanophotonic integrated circuits for spike-based computing: challenges and prospects

B. Romeira*[a], J. B. Nieder[a], B. Jacob[a], R. M. R. Adão[a], F. Camarneiro[a], J. Arturo Alanis[b],M. Hejda[b], A. Hurtado[b], J. Lourenço[a,c], D. Castro Alves[c], J. M. L. Figueiredo[c], I. Ortega-Piwonka[d], J. Javaloyes[d]
[a]Ultrafast Bio- and Nanophotonics Group, INL – International Iberian Nanotechnology Laboratory, Av. Mestre José Veiga n/a, 4715-330 Braga, Portugal; [b]Institute of Photonics, University of Strathclyde, Technology and Innovation Centre, 99 George St., G1 1 RD, Glasgow, UK; [c]Centra-Ciências, Departamento de Física da Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1740-016 Lisboa, Portugal; [d]Institute of Applied Computing and Community Code (IAC-3), Departament de Física, Universitat de les Illes Balears, Cra. de Valldemossa, km 7.5, E-07122 Palma de Mallorca, Spain

ABSTRACT

Event-activated biological-inspired subwavelength (sub-λ) optical neural networks are of paramount importance for energy-efficient and high-bandwidth artificial intelligence (AI) systems. Despite the significant advances to build active optical artificial neurons using for example phase-change materials, lasers, photodetectors, and modulators, miniaturized integrated sources and detectors suited for few-photon spike-based operation and of interest for neuromorphic optical computing are still lacking. In this invited paper we outline the main challenges, opportunities, and recent results towards the development of interconnected neuromorphic nanoscale light-emitting diodes (nanoLEDs) as key-enabling artificial spiking neuron circuits in photonic neural networks. This method of spike generation in neuromorphic nanoLEDs paves the way for sub-λ incoherent neural circuits for fast and efficient asynchronous brain-inspired computation.

Keywords: Nanophotonics, nanoLEDs, neuromorphic computing, optical interconnects, resonant tunneling diodes, spiking neural networks

1. INTRODUCTION

Artificial intelligence (AI) systems using computing algorithms of deep learning neural networks are emerging rapidly.[1] Despite the recent advancements within the field, the power budget involved in running deep learning neural networks in conventional computers is growing exponentially.[2] A type of bioinspired neural network is spiking neural networks (SNNs),[3] that uses artificial neuron units that exchange information via spikes. A SNN uses the timing of the spikes to process information and each artificial neuron is typically only active when it receives or emits spikes. Such behavior reduces the required energy for operating the neural network. Neuromorphic hardware exhibiting a spiking behavior can be implemented using electronics,[4] but then typically operates a low (kHz) speeds and requires several picojoule/spike.

A different type of upcoming neural networks includes photonic neural networks,[5] and in order for such networks to function as a SNN, spiking photonic neurons have been implemented using graphene excitable lasers, distributed feedback lasers, or vertical-cavity surface-emitting lasers, to name a few. However, the footprint of such lasers is still too large for compact and efficient SNNs. An alternative is to use a compact light-emitting diode (LED), e.g. a nanoscale LED.[6,7] Noteworthy, the development of a single, miniaturized light-emitting source for spike-based operation remains an ongoing, significant challenge. In this invited paper, we discuss the challenges and opportunities of a new architecture based on neuromorphic resonant tunneling-assisted nanoscale LEDs (nanoLEDs) to achieve a small footprint, fast speed (multi-gigahertz), and low energy consumption (<100 fJ/bit), as needed for future highly-dense interconnected optical neural networks. We propose their integration with 3D flexible interconnections for scalable systems. Lastly, we analyze their performance for information processing using optically interconnected spiking neural network models. Architectures as the one proposed here will significantly boost the transmission and processing capabilities of spike-based optoelectronic chips, offering exciting complementary solutions to electronics hardware in neuromorphic computing for AI systems.

*bruno.romeira@inl.int;

This is a peer reviewed, accepted author manuscript of the following research paper: Romeira, B., Nieder, J. B., Jacob, B., Adão, R. M. R., Camarneiro, F., Alanis, J. A., Hejda, M., Hurtado, A., Lourenço, J., Castro Alves, D., Figueiredo, J. M. L., Ortega-Piwonka, I., & Javaloyes, J. (2021). Subwavelength neuromorphic nanophotonic integrated circuits for spike-based computing: challenges and prospects. In Proceedings Volume 11804, Emerging Topics in Artificial Intelligence (ETAI) 2021 (Vol. 11804). (Emerging Topics in Artificial Intelligence; Vol. 11804). SPIE. https://doi.org/10.1117/12.2591852

2. EXCITABILITY IN NEUROMORPHIC RESONANT TUNNELING DIODES

Neurons exhibit excitability,[8] the dynamical property that is key to biologically inspired artificial intelligence. Seeking a better architecture that supports spikes as information carriers, in this work we look at resonant tunneling diodes as excitable neuromorphic spike generators.[9] These nonlinear quantum nanoelectronic elements can reach terahertz frequencies and may be integrated with photonics[10–12] (e.g. LEDs, lasers and photodetectors) for all-optical data transmission. Their speed stems from the nanometric size (~10 nm) of the semiconductor layer of the RTD in the epitaxial growth direction. This active layer consists of a double barrier quantum well (DBQW) nanostructure (typically using GaAs/AlAs or InGaAs/AlAs compound semiconductors). This provides a current-voltage with pronounced negative differential conductance (NDC), which is exploited for the firing of spiking and bursting signals. Figure 1 shows a numerically simulated example of spikes randomly fired when perturbing with additive white noise the circuit biased at the peak and valley regions (see more details in[9] for a comprehensive analysis on the spiking generation in RTDs).

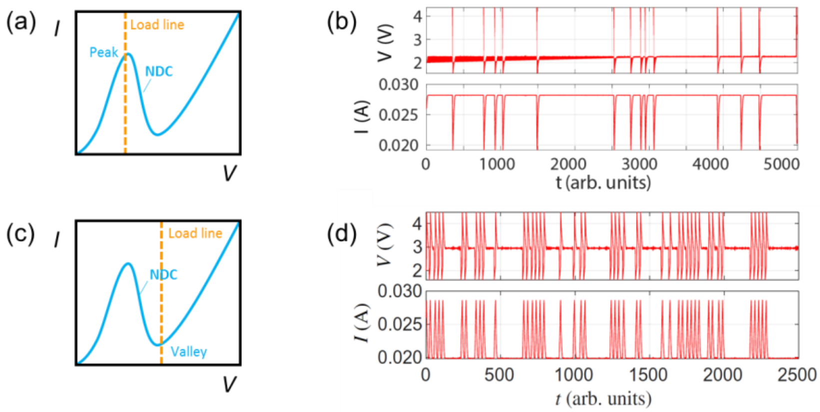

Figure 1. Current-voltage characteristics (left) and numerical traces of output pulses (right) randomly fired by perturbing with additive white noise. The system is biased in the peak region in (a) and in the valley region in (c). The input bias voltage and input noise intensity are (a) $V_0 = 2.26$ V, $\eta = 0.009$ V, (b) (c) $V_0 = 2.94$ V, $\eta = 0.014$ V (Adapted from[9]).

3. NANOPHOTONIC INTEGRATED NEURONS

Our nanophotonic integrated neurons exploit the unique physical properties of active semiconductor DBQW nanostructures – resonant tunneling diodes – embedded in metal-nanocavities for a new class of miniaturized optical artificial neurons. Figure 2(a) shows a schematic of the nanophotonic spiking neuron unit consisting of a DBQW monolithic integrated in a nanopillar LED. The DBQW enables control of the injection of electrons into the active region of the LED. This provides a nanoLED with a unique voltage-controlled NDC, red solid line in Fig 2(b), which is markedly different from the "linear" current-voltage characteristic of conventional LEDs. This enables extremely low-energy activation of all-or-nothing spiking responses in both the optical and electrical domains. The metal-nanocavity offers strong light–matter interaction at the nanoscale, leading to faster and efficient light emission,[13,14] a key feature for high-bandwidth optical spiking. Figure 2(c) shows the resonant tunneling-LED circuit used to simulate the static and dynamic properties of nanoLEDs and their potential as efficient spiking sources (see [15] for full details).

Figure 2(d) compares the operation of neuromorphic LEDs with standard current driven (non-spiking) micro- and nanoLED sources in terms of electrical and optical energy per spike (see[15] for a full discussion). Clearly, standard microLEDs (blue traces) are limited to modulation bandwidths <<1 GHz and require >100 fJ/bit. Remarkably, nanoLEDs (red traces) are suited for operation in the range 10–100 fJ/bit at multi-gigahertz speeds. However, there is a compromise between electrical energy and optical energy per bit produced. Noteworthy, for a neuromorphic nanoLED case (stars symbols), the electrical and optical energy per emitted spike is set by the intersection between the dashed-dot horizontal orange line (i.e. the refractory time, ~650 ps, of the analyzed spiking nanoLED) and the diagonal traces. In summary, neuromorphic nanoLEDs show prospects of spike generation at multi-gigahertz speeds which can be achieved upon receiving exceptionally low (sub-10 mV) synaptic-like activation signals (lower than biological voltages of 100 mV), and with remarkably low energy consumption, i.e. 10–100 fJ per emitted spike.

Noteworthy, our approach can potentially provide optically activated neuromorphic nanoLEDs using the photosensitive properties of resonant tunneling structures. Indeed, RTD-based photodetectors have been demonstrated,[16,17] by exploiting the properties of the tunneling current which is extremely sensitive to changes in the local electrostatic potential. This enables highly-sensitive detection (10^4 A/W) of photogenerated minority charge carriers and could pave the way to nanophotonic integrated photosensitive LEDs for fully optically interconnected artificial neurons.

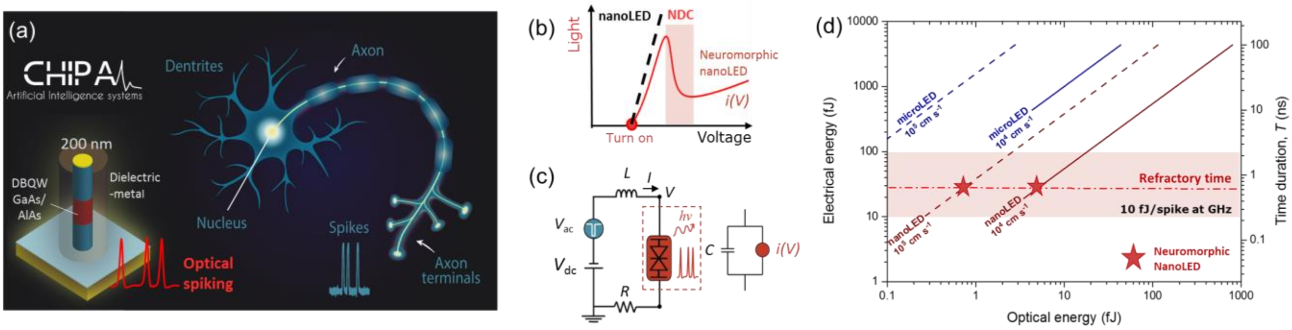

Figure 2. Schematic of the nanophotonic artificial neuron in a nanopillar metal-dielectric architecture. Also shown is the schematic of a biological neuron. (b) Comparison of the I-V curves between a "standard" nanoLED (dashed black line) and a neuromorphic nanoLED (solid red line). (c) Schematic of the nanoLED circuit modeled by a nonlinear voltage-controlled current source, $i(V)$, in parallel with the equivalent capacitance, C. (d) Electrical energy as a function of the optical energy per bit. The diagonal lines are the values for the micro- and nanoLEDs in a non-spiking regime. The dashed lines represent the case of a large surface recombination while the solid lines represent the best case of a low surface recombination. The dashed-dot horizontal line intersecting the diagonal traces indicates the refractory time and gives the electrical and optical energy per spike (indicated by the stars) of the neuromorphic nanoLED (adapted from[15]).

4. SYNAPTIC 3D PHOTONIC INTERCONNECTS

Achieving a versatile on-chip interconnection between nanophotonic devices is still a great challenge. The use of 3D waveguides was recently proposed,[18] and direct laser writing via two-photon polymerization (TPP) provides an attractive 3D microfabrication alternative to conventional optical assembly and packaging techniques. We propose the use of 3D waveguide designs for light coupling from and to the nanoLED neuromorphic sources analyzed in the previous section, Fig. 2(a), thus enabling flexible 3D photonic interconnections in spiking neural networks, Fig. 3(a). Our approach uses a custom-build TPP fabrication system consisting of a femtosecond laser centered at 795 nm wavelength, coupled to a 40× objective (NA=0.75). In our initial 3D microfabrication tests, a laser power of 18 mW and writing speed of 75 µm/s was used. The polymer is drop-casted on a glass coverslip, pre-baked at 80 °C for 10 min., post-baked at 130 °C for 20 min., and developed in OrmoDev for 12 min. In Fig. 3(b) waveguide structures with different supports and gaps are shown. The waveguide elevation varies between 40 and 60 µm depending on the support with a square section of 10×12 µm^2, Fig. 3(c). A closer view of (d) a 33 µm long support and of (e) a suspended waveguide is also shown. Further efforts of co-integration of such platform with nanophotonic integrated neurons is ongoing and will be of paramount importance to achieve fully scalable interconnected systems.

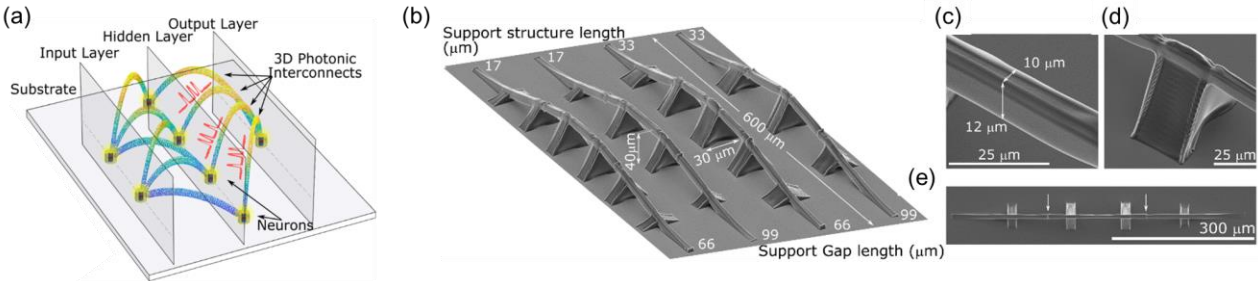

Figure 3. (a) Schematic example of a 3D interconnected spiking neural network. Scanning electron microscope images of fabricated 3D TPP architectures. b) Tilt view of structures with different supports. c) Top view of selected 3D waveguide. (d) Detail view of a 33 µm long support and of (e) a suspended waveguide (adapted from[19]).

5. SPIKE INFORMATION PROCESSING

In this last section, we investigate and analyze the proposed neuromorphic system and discuss a feasible network implementation of interconnected pre- and post-synaptic artificial neuron nodes. We analyze a representative network consisting of a 5-to-1 feedforward (two-layer) spiking neural network architecture, Fig. 4. Using physical models for each node (see sections 2 and 3), we numerically analyze the potential of the network to classify spatial 5-bit pulse patterns encoded in time. The training of such network, using a supervised learning scheme that employs a spike-timing dependent learning rule,[20] reveals that during the inference phase, 94%+ accuracy for spatial pulse pattern recognition can be achieved. After the training phase, the network can perform inference for recognition of the selected spatiotemporal 5-bit pattern. These results are the first theoretical demonstration of RTD-based neuromorphic nanophotonic spike information processing, revealing the feasibility of delivering successful operation in pattern recognition tasks by using multiple interconnected devices in the form of a photonic feed-forward spiking neural network.

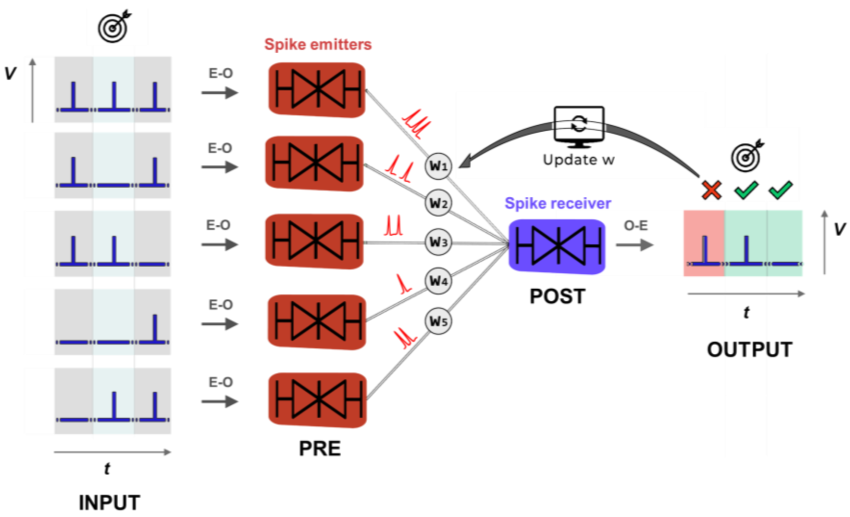

Figure 4. Simulated network architecture diagram, illustrating how patterns of input electrical pulses (in blue) enter the nanoscale spike emitter nodes (PRE) and are propagated as optical signals to the downstream spike receiver node (POST) using weighted, W, connection. The output state of the downstream node is compared to the label, and if there is a mismatch between label and output state, the weights are updated. The desired pattern is highlighted with the target icon.

6. CONCLUSIONS

In this work an architecture that supports spikes as information carriers was discussed. It consists of resonant tunneling diodes as excitable neuromorphic spike generators integrated with nanoLEDs for all-optical data transmission. This architecture and method of spike generation in neuromorphic nanoLED devices combined with techniques for complex interconnectivity using 3D interconnects, paves the way for sub-λ incoherent spiking neural circuits for optically interconnected photonic spiking neural networks and event-based asynchronous brain-inspired computation.

7. ACKNOWLEDGMENTS

The authors acknowledge funding support from the European Commission (Grant No. 828841 ChipAI-H2020-FETOPEN-2018-2020), the UKRI Turing AI Acceleration Fellowships Programme (Grant No. EP/V025198/1), and the Office of Naval Research Global (Grant No. ONRGNICOPN62909-18-1-2027).

REFERENCES

[1] LeCun, Y., Bengio, Y. and Hinton, G., "Deep learning," Nature **521**, 436 (2015).
[2] Xu, X., Ding, Y., Hu, S. X., Niemier, M., Cong, J., Hu, Y. and Shi, Y., "Scaling for edge inference of deep

neural networks," Nat. Electron. **1**(4), 216–222 (2018).
[3] Wu, Y., Deng, L., Li, G., Zhu, J. and Shi, L., "Spatio-Temporal Backpropagation for Training High-Performance Spiking Neural Networks," Front. Neurosci. **12**, 331 (2018).
[4] Merolla, P. A., Arthur, J. V, Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S. K., Appuswamy, R., Taba, B., Amir, A., Flickner, M. D., Risk, W. P., Manohar, R., et al., "A million spiking-neuron integrated circuit with a scalable communication network and interface," Science (80-.). **345**(6197), 668–673 (2014).
[5] Shastri, B. J., Tait, A. N., Ferreira de Lima, T., Pernice, W. H. P., Bhaskaran, H., Wright, C. D. and Prucnal, P. R., "Photonics for artificial intelligence and neuromorphic computing," Nat. Photonics **15**(2), 102–114 (2021).
[6] Romeira, B., Borme, J., Fonseca, H., Gaspar, J. and Nieder, J. B., "Efficient light extraction in subwavelength GaAs/AlGaAs nanopillars for nanoscale light-emitting devices," Opt. Express **28**(22) (2020).
[7] Dolores-Calzadilla, V., Romeira, B., Pagliano, F., Birindelli, S., Higuera-Rodriguez, A., van Veldhoven, P. J., Smit, M. K., Fiore, A. and Heiss, D., "Waveguide-coupled nanopillar metal-cavity light-emitting diodes on silicon," Nat. Commun. **8**, 14323 (2017).
[8] Izhikevich, E. M., "Neural excitability, spiking and bursting," Int. J. Bifurc. Chaos **10**(06), 1171–1266 (2000).
[9] Ortega-Piwonka, I., Piro, O., Figueiredo, J., Romeira, B. and Javaloyes, J., "Bursting and Excitability in Neuromorphic Resonant Tunneling Diodes," Phys. Rev. Appl. **15**(3), 34017 (2021).
[10] Romeira, B., Avo, R., Figueiredo, J. M. L., Barland, S. and Javaloyes, J., "Regenerative memory in time-delayed neuromorphic photonic resonators," Sci. Rep. **6** (2016).
[11] Romeira, B., Figueiredo, J. M. L. and Javaloyes, J., "Delay dynamics of neuromorphic optoelectronic nanoscale resonators: Perspectives and applications," Chaos **27**(11) (2017).
[12] Ironside, C., Romeira, B. and Figueiredo, J., [Resonant Tunneling Diode Photonics], Morgan & Claypool Publishers (2019).
[13] Romeira, B. and Fiore, A., "Physical Limits of NanoLEDs and Nanolasers for Optical Communications," Proc. IEEE **108**(5), 735–748 (2020).
[14] Romeira, B. and Fiore, A., "Purcell Effect in the Stimulated and Spontaneous Emission Rates of Nanoscale Semiconductor Lasers," IEEE J. Quantum Electron. **54**(2) (2018).
[15] Romeira, B., Figueiredo, J. M. L. and Javaloyes, J., "NanoLEDs for energy-efficient and gigahertz-speed spike-based sub-λ neuromorphic nanophotonic computing," Nanophotonics(0), 20200177 (2020).
[16] Zhang, W., Al-Khalidi, A., Figueiredo, J., Al-Taai, Q. R. A., Wasige, E. and Hadfield, R. H., "Analysis of Excitability in Resonant Tunneling Diode-Photodetectors," Nanomaterials **11**(6) (2021).
[17] Pfenning, A., Jurkat, J., Naranjo, A., Köck, D., Hartmann, F. and Höfling, S., "Resonant tunneling diode photon number resolving single-photon detectors," Infrared Remote Sens. Instrum. XXVII **11128**, M. Strojnik and G. E. Arnold, Eds., 47–56, SPIE (2019).
[18] Moughames, J., Porte, X., Thiel, M., Ulliac, G., Larger, L., Jacquot, M., Kadic, M. and Brunner, D., "Three-dimensional waveguide interconnects for scalable integration of photonic neural networks," Optica **7**(6), 640–646 (2020).
[19] Adao, R. M. R., Romeira, B., Nieder, J. B., "Design and Fabrication of 3D Interconnects for Photonic Neuronal Networks Using Two-Photon Polimerization," Lasers Electro-Optics (CLEO), (Optical Soc. Am. Washington, DC) (2021).
[20] Wang, W., Pedretti, G., Milo, V., Carboni, R., Calderoni, A., Ramaswamy, N., Spinelli, A. S. and Ielmini, D., "Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses," Sci. Adv. **4**(9) (2018).