MICRO-LASERS
FOR NEUROMORPHIC COMPUTING

//

Venkata Anirudh PAMMI[1], Sylvain BARBAY[1,*]

[1] Centre de Nanosciences et de Nanotechnologies, CNRS, Université Paris Saclay, Palaiseau, France

* sylvain.barbay@c2n.upsaclay.fr
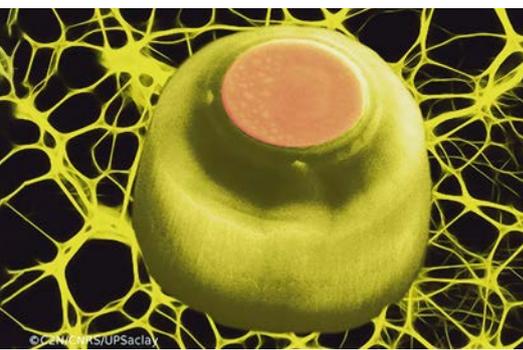
Spiking micro-lasers are interesting neuromorphic building blocks to realize all-optical spiking neural networks. Optical spike-based computing offers speed and parallelism of optical technologies combined with a sparse way of representing information in spikes, thus with a potential for efficient brain-inspired computing. This article reviews some of the latest advances in this field using single and coupled semiconductor excitable micro-lasers.

Neuromorphic photonics aims at drawing inspiration from the brain to design efficient computing hardware using a photonic substrate. The interest in this field stems from the potential gain in speed and parallelism due to the use of optical technologies while reducing the energy consumption compared to standard Von Neumann architectures. Amongst the numerous neuromorphic photonic platforms, excitable micro-lasers display many properties present in biological neurons and are thus appealing to implement fast and efficient brain-inspired functionalities. Starting from the building block, the optical neuron, one of the major objectives is to design photonic neural networks which are interconnected excitable nodes with controllable weights, implementing learning capabilities. These building blocks can also be used in new computing frameworks such as liquid state machines [1] and photonic processing accelerators, using spike-based computing strategies [2] replacing the binary logic paradigm: while in standard computers information is encoded in a voltage level representing either the binary value 0 or 1, in the brain information is encoded in 'spikes' which are short electrical pulses emitting by neurons. Spiking is a fundamental property of biological neurons which is a generic property found in many physical systems and is also called excitability (insert).

Excitability in optics can be obtained in different experimental systems. A particularly attracting configuration is the case of the semiconductor laser with saturable absorber (see insert). A laser with saturable absorber contains, in addition to a gain medium, an absorptive medium whose absorption saturates with the intracavity laser intensity.

This type of laser can be self-pulsing above the laser threshold and reveals an excitable behaviour below threshold; the rest state is then the laser off state. When the laser is subjected to perturbations above the excitable threshold, it emits calibrated optical pulses which can be of hundreds of picosecond duration or less. The difficulty is to integrate in a laser both the gain and the saturable absorber medium, because basically both require opposite pumping conditions. We have been able to design and fabricate such a semiconductor micro-laser, opening the way to optical neurons behaving similarly to biological ones, but with the important differences that optical neurons respond on timescales more than one million times shorter than their living counterpart [3].

Another important property of neurons is their absolute refractory period: neurons cannot spike at arbitrary high frequencies; there is

a minimal time difference between two spikes. There is also a relative refractory period which corresponds to an intermediate situation where a neuron produces an attenuated spike in response to a second stimulus after having spiked once. This means that neurons have a memory of their past state. In semiconductor optics, the relative refractory period is very short, less than one nanosecond. The refractory periods are fundamental for unidirectional transport of information in the axon, as will be explained below.

Because excitability is the result of a dynamical process, the neuron responds to any above threshold stimulus with the same calibrated pulse but with a different latency time, which depends on the input perturbation strength: the weaker the perturbation, the larger is the delay in the response. This mechanism introduces a natural coding scheme into spikes of input perturbation amplitudes, the temporal coding: information on the input amplitude of a stimulus is reflected into the temporal delay of the produced spike. This coding mechanism is an alternative coding scheme to the simpler rate coding scheme which encodes the input amplitude in the rate of spikes. However, it has been proven that this former coding can only explain the fast image recognition capability of the brain. It is therefore an important mechanism for any spike-based, biologically plausible computing architecture. In the optical neuron, spike latency ranging from few hundreds of picoseconds to several nanoseconds have been measured in response to short optical input stimuli.

Equipped with excitable response, refractory period and spike latency, the optical neuron can respond to stimuli and encode information into spikes. It can act as a nonlinear gate but it cannot compute on its own. The computing ability comes from a fourth important ingredient, the temporal summation. Neurons can integrate sub-threshold stimuli over time: they are 'leaky-integrators'. As a consequence, a neuron receiving delayed stimuli from different sources will be able to spike if the stimuli are close and not if they are far away: they can detect coincidences [4]. This simple property has huge consequences since it mathematically leads to the property of universal computing: networks of coincidence detector neurons can approximate any function. They constitute the third generation of neural networks and are believed to be more powerful in terms of the number of nodes needed for a specific computing with respect to other kinds of neural networks.

COMPUTING WITH COUPLED NEUROMORPHIC MICRO-LASERS

Even though a single optical neuron can realize simple computing tasks, coupling several such neurons is necessary to scale up the computational power [5]. A simple yet non-trivial coupling scheme is self-coupling or autapse, where the laser response is coupled back to itself *via* a time delayed feedback. This configuration is motivated from biology since autapses are found in the mammalian brain and it is hypothesized that they linked to the memory capacity. From a physical point of view, having such structures in neural networks can enable complex dynamics and present ways to store information.

An all-optical autapse is built by reflecting back the response of the laser

We have been able to design and fabricate such a semiconductor microlaser, opening the way to optical neurons behaving similarly to biological ones, but with the important differences that optical neurons respond on timescales more than one million times shorter than their living counterpart [3].

using *e.g.* a distant mirror placed several tens of centimeters away, forming an external cavity with a feedback time on the order of tens of nanoseconds. If the system is fed with a single optical perturbation above the excitable threshold, it will generate a train of spikes whose repetition period corresponds to the sum of spike latency and cavity round-trip time. Additional input perturbations can be sent to the system to add or remove spike trains and change the timing of existing spike trains. This demonstrates an all-optical buffer for spikes which can be controlled with single optical pulses. In the short term (~ tens to hundreds of external cavity round-trips) after the initial excitation, the individual spike trains seem independent and do not interact with each other. This implies that the information can be stored in

> Even though a single optical neuron can realize simple computing tasks, coupling several such neurons is necessary to scale up the computational power [5]. A simple yet non-trivial coupling scheme is self-coupling or autapse, where the laser response is coupled back to itself via a time delayed feedback.

the inter-spike intervals and can be preserved for hundreds of nanoseconds. This is sufficient to act as a working memory, *i.e.* an intermediate storage with limited capacity, a central concept in neuroscience. However, if the spike trains evolve for longer durations, we observe that the inter-spike intervals equalize: any time coded information is lost in the long term. This equalization can be seen as a computational process similar to that in an

associative memory. Associative or Content Addressable memory is a type of memory with a computing ability. It accepts an input which can be partial information and retrieves the stored memory closest to it. This is exactly what happens in the case of the optical autapse. If the system is fed with one time-equalized spike train, it responds with this state. If it is fed with a nearby state (corresponding for instance to a given, non periodic,

Excitable semiconductor microlaser: an excitable semiconductor micro-laser is depicted in a) with a SEM micrograph in b). It comprises two Distributed Bragg Reflectors and an active zone composed of quantum wells (for gain and saturable absorption) and is encapsulated in a SiN layer to prevent oxidation and for better heat management. The microlaser is optically pumped (not shown) and subjected to pulsed perturbations (P). In the excitable regime, the 'rest' state of the system corresponds to the laser off state (below laser threshold).
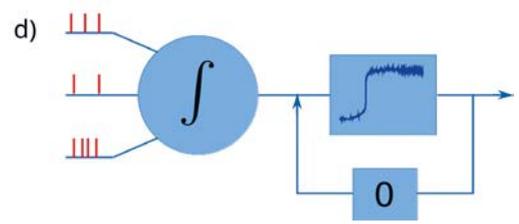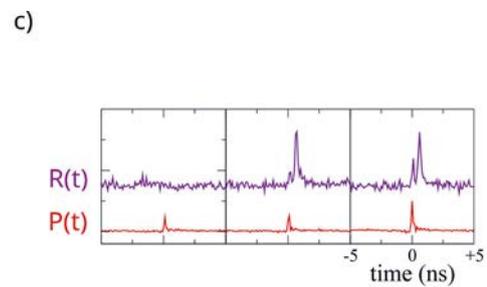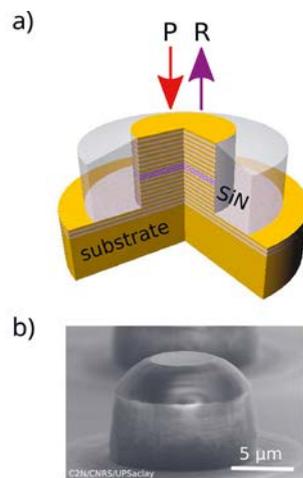
Small perturbations do not change the state of the neuron. However, when a perturbation above the so-called excitable threshold is sent to the micro-laser, an 'action potential' is generated in the form of a response pulse R (a spike). Importantly, as can be seen in c), the shape of the generated pulse does not depend on the input perturbation as long as it is above the excitable threshold: the neuron acts as a "all-or-none" system and has thus a highly nonlinear response. Other major properties of the neuron are represented schematically in the diagram d): the optical neuron can sum input stimuli in the form of spikes from different sources, the integrated signal is compared to a threshold and a spike is emitted if the threshold is reached. The neuron then recovers to its quiet state. The recovery is not instantaneous and takes some time: this is referred to as the refractory periods of the neuron. Also, the response spike is generated with a dynamical delay which depends on the integrated perturbation strength.
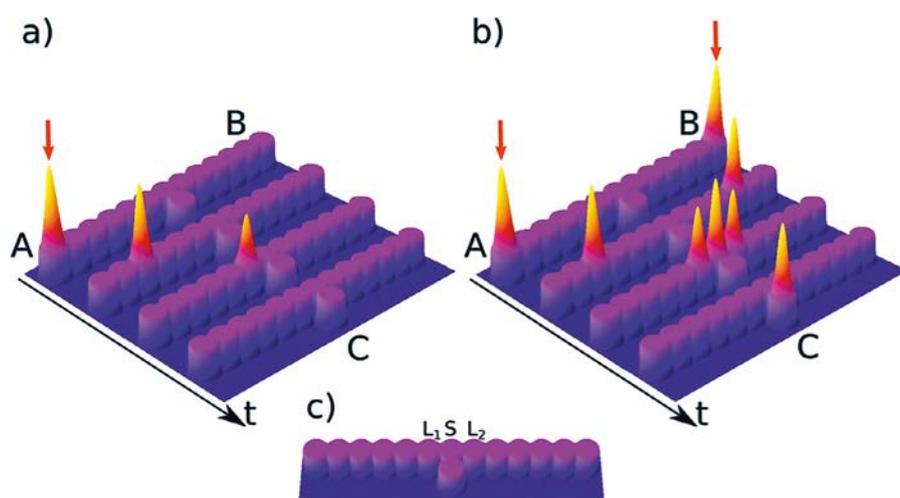
Figure 1.
Temporal AND logic gate. The micropillars are represented by small touching cylinders in c). All the nodes are in the excitable regime and receive the same level of pump except nodes L_1 and L_2 which are pumped less. In a) node A is initially excited by an external perturbation. The response spike propagates along the gate but fails to excite node S. Thus, no output is detected at the exit node C. In b) both nodes A and B receive an input at the same time. The excitations propagate towards each other and are summed in S resulting in a spike at the output node C.

timing configuration of spikes), it will converge to the closest state in memory. Notably, this functionality is implemented here thanks to a single neuron, while it usually needs many nodes and connections such as in Hopfield networks.

A second scheme consists of several evanescently coupled optical neurons. In this scheme the lasers are placed physically close together so that their emission modes overlap, resulting in the coupling of optical field from one laser to its neighbour. Suppose there is a chain of N coupled lasers in their rest state in the excitable regime (see Fig. 1). If a perturbation excites a spike in the first laser and if the coupling is large enough, this excitation will propagate along the chain much like in the axon in biological neurons, giving rise to a lossless saltatory propagation. The propagation is unidirectional: when one laser inside the chain emits a spike, the optical field leaks symmetrically in both its nearest neighbours. But one of the neighbours is inevitably in its refractory period and does not respond to this spike whereas the other node

does. Excitability is a highly nonlinear process and these propagating pulses have special properties. Unlike linear waves that propagate and interfere when they meet, excitable spikes usually collapse when they meet. Using this property and recalling that the excitable threshold depends on the pump amongst other things, a temporal AND gate can be built. The temporal AND gate shown in Fig. 1 only fires when it receives two (or more) inputs within a certain time window such that the sum of the individual inputs exceeds a threshold. The nodes L_1 and L_2

shown in Fig. 1c receive a lower level of pump as compared to the other nodes, allowing to realize this functionality. In Fig. 1a, only node A is excited by an input signal. This signal propagates till the node L_1 where the excitable response is lower because of the decreased pump. The lower response fails to excite the summation node S and consequently no response is detected at node C. However, if both nodes A and B are excited at the same time as in Fig. 1b, the node S sums the two sub-threshold excitations and emits a spike at node C. The same kind of coincidence detection circuitry is found in some animals (*e.g.* in the Barn owl) for determining the direction of sound stimuli. Elaborating on this effect and using the propagation delays enabled by the spike propagation, more advanced circuits can be designed allowing to distinguish between spike trains with different interspike intervals and detect a definite temporal pattern sequence.

In conclusion, spiking micro-lasers constitute good candidates to design fast and scalable neuromorphic circuits for spike-based processing. Coupled to artificial synapses such as built from phase change materials, spiking lasers have recently been shown to provide learning capabilities too [6]. Besides their interest in analog computing, spiking micro-lasers also enable to study biologically plausible architectures that can benefit from the advances in neurosciences. ●

REFERENCES

[1] W. Maass, in *Computability in Context*, edited by S.B. Cooper and A. Sorbi (Imperial College Press, London, 2011), pp. 275–296

[2] M.A. Nahmias *et al.*, IEEE J. Quantum Electron. **19**, 1800212 (2013)

[3] F. Selmi *et al.*, Phys. Rev. Lett. **112**, 183902 (2014)

[4] F. Selmi *et al.*, Opt. Lett. **40**, 5690 (2015)

[5] V.A. Pammi *et al.*, IEEE J. Sel. Top. Quantum Electron. **26**, 1500307 (2020)

[6] J. Feldmann *et al.*, Nature **569**, 208 (2019)