
Project Acronym: **Fun-COMP**

Project Title: **Functionally scaled computing technology:** From novel devices to non-von Neumann architectures and algorithms for a connected intelligent world

**WP3**

**Neuromorphic computing with N-vN devices and networks**

**(WP Leader Thales)**

# Deliverable D3.2: Hardware demonstrator consisting of a photonic RC network incorporating phase-change elements and with self-learning capabilities

Deliverable ID: D3.2

Deliverable title: Hardware demonstrator consisting of a photonic RC network incorporating phase-change elements and with self-learning Hebbian capabilities

Revision level: FINAL

Partner(s) responsible: IMEC

Contributors: IMEC (Alessio Lugnan, Andrew Katumba, Peter Bienstman); UNEXE (Santiago G-C Carrillo, Emanuele Gemo, C David Wright); OXFORD (Samarth Aggarwal, Xuan Li, Harish Bhaskaran); WWU (Frank Brukerhoff Pluckermann, Wolfram Pernice)

Dissemination level: PU[1]

[1] CO: Confidential, only for members of the Fun-COMP consortium (including the Commission Services); PU: Public.

# Summary

In this deliverable we present the most relevant aspects and steps of Fun-COMP research on the development of a self-learning reservoir computing system based on an integrated photonic network containing phase change material cells (N-vN unit cells).

In Section 1 we discuss the reasoning behind our choice of the type of photonic network we planned to fabricate and investigate, together with the main functional requirements we aimed to achieve.

In Section 2 we focus on the main building block for our plastic reservoir network. In particular, we describe the development of a new numerical model to simulate the device in a modular and efficient way. Moreover, we show that our simulations predict that significant improvements in energy efficiency and contrast of memory operations can be achieved by suitably choosing the design parameters of the device, allowing us to build more scalable plastic photonic networks.

In Section 3 we discuss about system-level simulations of our plastic reservoir networks. In particular, we present a novel scalable training method to improve the performance of our reservoir system by means of plasticity.

In Section 4 we describe the design, fabrication and measurement of several versions of the main building block device for our reservoir network.

In Section 5 we describe the design and successful fabrication of several plastic reservoir network versions.

In Section 6 we summarize the main conclusions regarding the work presented in this deliverable.

# 1 RATIONALE FOR THE DEVELOPMENT OF THE PLASTIC PHOTONIC NETWORK FOR SELF-LEARNING RESERVOIR COMPUTING

In the past decade, the Fun-COMP partners IMEC have been investigating and developing *reservoir computing*[2][3] (RC) systems based on passive photonic networks on a silicon chip[4][5][6]. The reservoir in these systems is based on multiple input and output ports (consisting of grating couplers) connected by waveguides, splitters and combiners, so to form a network where the input optical signal interferes with multiple delayed versions of itself (an example is schematized in Fig. 1).

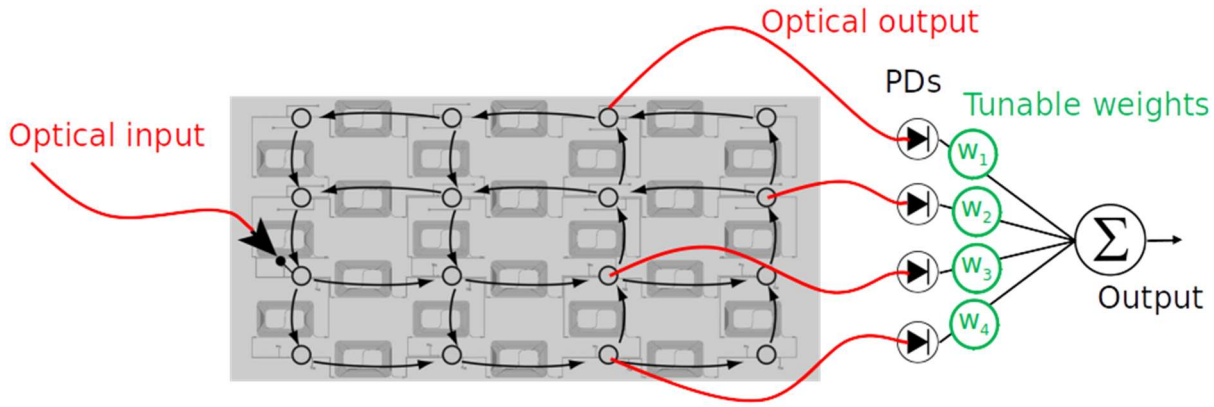


***Figure 1.*** Schematic of the passive integrated photonic RC system. Input optical signals can be inserted in the chip (grey area, black arrows and circles show the network topology) through grating couplers. The photonic circuit splits the input signal, delays the different splits (through long waveguides folded into spirals) and recombines the differently delayed versions of the input signal at the network nodes, so that they optically interfere. The resulting signal at some nodes is read out (optical outputs) by fast detectors that measure the time-dependent optical signal intensity. These intensity measures are directly fed into a ML linear classifier or regressor (one layer of neurons) that is trained to carry out a given task on the input signal.

The connections in this type of network are fixed and, because of fabrication errors, they cannot be predicted before fabrication. Therefore, at different output ports, different representations of the input signal are obtained, due to the different ways the original input signal interferes with itself at different nodes of the network. The signal collected at the different reservoir outputs is then fed into a linear *readout,* i.e. a linear regressor or classifier (that can as well be considered as a single layer of neurons of an artificial neural network), which is trained to carry out a chosen machine learning (ML) task on the input signal, such as bit sequence classification, Boolean operations with memory or

[2] Tanaka, Gouhei, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. "Recent advances in physical reservoir computing: A review." *Neural Networks* 115 (2019): 100-123.

[3] An introduction on the topic is provided in the public deliverable D3.1: "A new approach to reservoir computing".

[4] Vandoorne, Kristof, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. "Experimental demonstration of reservoir computing on a silicon photonics chip." *Nature communications* 5, no. 1 (2014): 1-6.

[5] Sackesyn, Stijn, Chonghuai Ma, Joni Dambre, and Peter Bienstman. "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation." *Optics Express* 29, no. 20 (2021): 30991-30997.

[6] Ma, Chonghuai, Joris Lambrecht, Floris Laporte, Xin Yin, Joni Dambre, and Peter Bienstman. "Comparing different nonlinearities in readout systems for optical neuromorphic computing networks." *Scientific Reports* 11, no. 1 (2021): 1-8.

telecom signal equalization. The role of the reservoir is to enhance the effective computational power of the trained readout, by increasing the dimensionality of the original input signal (e.g., in Fig. 1, one time-dependent input is transformed into four time-dependent representations) and by introducing memory into the system. In particular, in the discussed passive integrated type of photonic reservoir, memory is provided by optical delay lines (long waveguides folded into spirals) that connect the different splitters and combiners. The main advantages of employing a passive photonic circuit as reservoir is that it does not actively consume energy (although the input signal undergoes significant power losses while propagating through the network) and that it can operate at extremely high speed (limited only by the time of travel of the light through the circuit). On the other hand, a passive circuit can only perform linear optical operations and thus the required nonlinearity (to expand the dimensionality of the input signal) is only given by the detectors at the readout, which transform an optical signal (characterized by amplitude and phase) into an electric one, proportional to the optical intensity. The RC system therefore represents a shallow artificial neural network (ANN), as opposed to deep ANNs where nonlinear nodes are cascaded, forming multiple layers. Generally, a deep nonlinear architecture is more powerful in creating diverse representations of the input using a limited number of connections and nodes, which often translates into a higher computational power.

In Task 3.1 of the Fun-COMP project we aimed to enhance a photonic reservoir by introducing the all-optical non-volatile memory provided by the N-vN unit cells (i.e. waveguide segments covered by a thin film of phase-change material, namely here the $Ge_2Sb_2Te_5$ alloy, or GST for short). In particular, we want this non-volatile memory to be accessible by the same input signal that we want to process via our ML system, as opposed to an external tuning of the memory elements. That is, we want the input signal to reach the N-vN cells in the reservoir circuit with enough power to change the memory state (i.e. the solid-phase state of the GST thin films). This would allow our enhanced reservoir to plastically adapt its internal connections to its input: we would thus introduce the *plasticity* property into our ML system. Such a condition is key to achieve *self-learning*, i.e. the ability to improve the performance on a ML task without explicitly and externally tuning the parameters to be learned, such as the synaptic weights in the case of an ANN. Importantly, our brain learns by means of plasticity and self-learning, therefore we know that these properties can potentially enable us to overcome the scalability limitations of today's neuromorphic computing systems, which heavily rely on the non-biologically-plausible backpropagation training algorithm[7][8].

However, there are two main issues that prevent us to readily insert the N-vN unit cells into the described integrated passive photonic reservoir to obtain a plastic RC system:

- The first issue is that the passive integrated reservoir (e.g. see Fig. 1) operates much faster than the N-vN cell. Indeed, the maximum time scale of the first is given by the maximum length of the delay lines connecting the optical nodes, while the minimum time scale of the latter is given by how fast the solid-state phase of the GST on a silicon waveguide can be changed by a realistic optical input signal (the time scale is from tens to hundreds of nanoseconds[9]). In order to match the time scales of the two devices, way too long optical delay lines would be

[7] Whittington, James CR, and Rafal Bogacz. "Theories of error back-propagation in the brain." *Trends in cognitive sciences* 23, no. 3 (2019): 235-250.

[8] Taherkhani, Aboozar, Ammar Belatreche, Yuhua Li, Georgina Cosma, Liam P. Maguire, and T. Martin McGinnity. "A review of learning in biologically plausible spiking neural networks." *Neural Networks* 122 (2020): 253-272.

[9] Li, Xuan, Nathan Youngblood, Zengguang Cheng, Santiago Garcia-Cuevas Carrillo, Emanuele Gemo, Wolfram HP Pernice, C. David Wright, and Harish Bhaskaran. "Experimental investigation of silicon and silicon nitride platforms for phase-change photonic in-memory computing." *Optica* 7, no. 3 (2020): 218-225.

required in the passive reservoir, causing the optical connections to be too lossy and to occupy a too large chip area.

- The second issue is that significant optical power is always lost at the combiners of the optical nodes in the passive reservoir, due to interference. This makes it very challenging to introduce several N-vN cells (that also are significantly lossy) in a large passive reservoir, since enough optical power must reach all the N-vN cells and the RC readout. This considerably limits the scalability of a plastic passive reservoir.

It should be noted that both issues specifically originate from the requirement of plasticity and could otherwise be circumvented if we allowed ourselves to externally tune the memory state of the N-vN unit cells.

To solve these problems, we consider an integrated optical reservoir based on a network of silicon ring resonators (RRs, see schematic in Fig. 2) instead of the passive reservoir architecture.

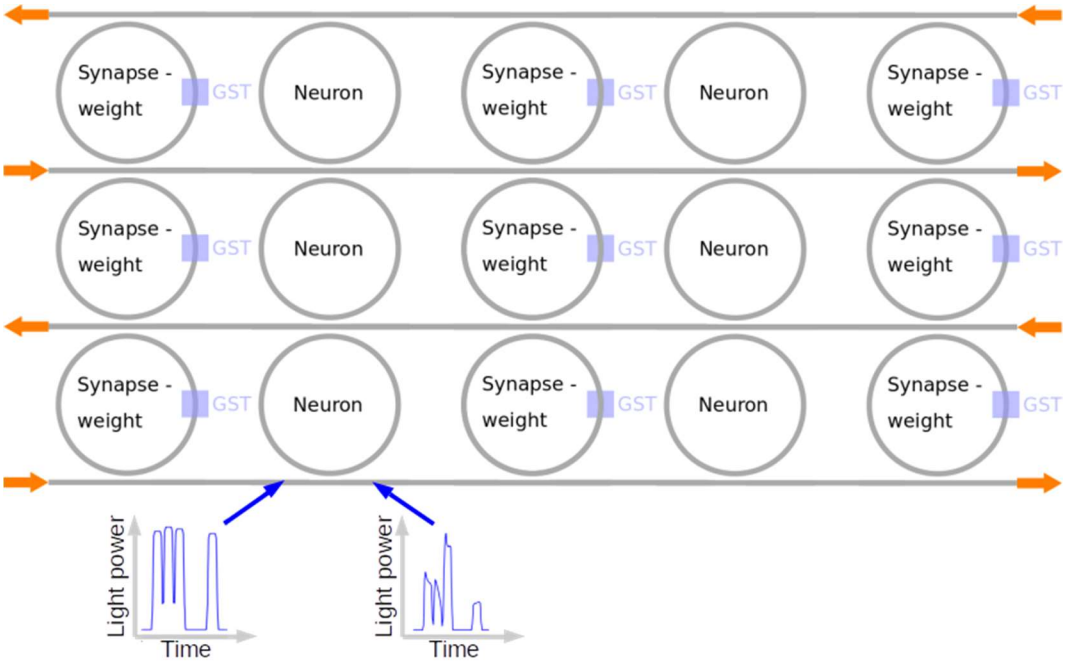


***Figure 2.*** Schematic of the proposed silicon RR optical network for integrated plastic RC with PCM. Several RRs are coupled by straight silicon waveguides. Those RRs that do not comprise a N-vN cell (GST layer) are used to provide volatile memory and nonlinearity via silicon nonlinear effects (thus they can be seen as neurons of a RNN). Those RRs combined with PCM cell, instead, provide all-optical non-volatile memory to the circuit, and thus plasticity (they can be seen as connections with plastic synaptic weights). The small waveform plots at the bottom exemplify how a RR can nonlinearly transform an input sequence of optical pulses.

The first of the aforementioned issues is then solved by employing silicon nonlinear effects (based on temperature and free carrier concentration of a silicon waveguide) to provide the network with the required volatile memory for RC operations, instead of the optical delay lines in the passive reservoir architecture. In particular, a strong enough light pulse exciting a silicon RR can increase the temperature and free carriers concentration in the ring waveguide, which can in turn modify the RR resonance properties and thus the way the light is transmitted by these devices. After the optical excitation of a RR, the consequent perturbation in temperature and free carrier concentration relaxes with characteristic times of, respectively, around 5ns and 100ns. Importantly, these are well comparable with the characteristic times of all-optical memory operations in a N-vN cell. In addition to provide volatile memory, the silicon RRs are also nonlinear nodes and can be considered as neurons of an ANN with memory, i.e. a recurrent neural network (RNN). The presence of multiple layers of

neurons, as previously discussed, is expected to provide a significant advantage in computational power w.r.t. the passive reservoir case.

The latter of the aforementioned issues is addressed by combining a N-vN unit cell with a silicon RR to form a plastic synapse with improved energy efficiency (for brevity, from now on we will refer to this device as the *plastic node*). Although such a device is structurally the same as what we call (in the context of Fun-COMP project) the *extended N-vN* unit cell, it is functionally different: instead of being used and optimized to mainly provide a means for wavelength division multiplexing, it is used and optimized to enhance energy efficiency and cascadability of non-volatile memory operations, that is in order to provide a more scalable source of plasticity. In particular, thanks to its resonant behaviour, the silicon RR presents an enhanced sensitivity to the state of the GST deposited on its waveguide. This allows us to obtain a sufficient output contrast (i.e. the difference in the RR optical output due to different memory states) from a significantly shorter GST cell, which in turn requires less optical energy to be switched from a solid-state phase to another. Further details can be found in our recently published article[10]. At the same time, using this type of plastic node coupled with other silicon RRs (e.g. as in Fig. 2) provides a way for the optical input power to reach and change the memory state of N-vN cells that are located far away and deep in the network. Indeed, the plastic node can be designed so that an optical pulse that changes its memory state can also shift the resonance wavelength away from the input wavelength (because of the aforementioned silicon nonlinear effects). This means that, for a certain time after its excitation, the node does not disrupt the passage of subsequent input pulse, which can therefore reach the next RRs in the series, and so on.

Finally, the configuration space of the proposed plastic reservoir is greatly expanded by the strong sensitivity of RRs to the wavelength of input light. Indeed, for very close wavelengths, e.g. with differences of just 0.01nm, the network can behave very differently. Thus, in principle we obtain many different network topologies corresponding to different input wavelengths. Moreover, thanks to the wavelength multiplexing capability of RRs (RRs have quasiperiodic resonance wavelengths in the frequency domain) these different network topologies can be connected to one another by means of the silicon nonlinear effects, thus potentially forming a network much larger than the physical one, without increasing the on-chip footprint. Indeed, two different resonant wavelengths can be coupled in a RR because both affect and are affected by the temperature and the free carrier concentration of the same ring waveguide.

# 2 SILICON RING RESONATOR WITH GST CELL: AN EFFICIENT AND SCALABLE PLASTIC NODE

In this section we will present the theoretical and numeric investigation of the plastic node introduced in the previous section. Such a plastic node consists of a silicon RR with a GST cell on its waveguide (extended N-vN configuration) that is specifically designed so that its resonance properties (dynamically affected by silicon nonlinear effects) improve the energy efficiency of memory operations and the cascadability of several plastic nodes in a network. Here we present the main results, while more details can be found in the article[10] that we have recently published together with the Fun-COMP EXETER partners.

[10] Alessio Lugnan, Santiago García-Cuevas Carrillo, C. David Wright, and Peter Bienstman, "Rigorous dynamic model of a silicon ring resonator with phase change material for a neuromorphic node," Opt. Express **30**, 25177-25194 (2022)

We developed a numerical model to simulate the dynamical behaviour of the plastic node, by combining two pre-existing numerical models:

- the compact behavioural model of the basic N-vN unit cell (developed by EXETER partners and presented in Deliverable D1.6 and in[11]) whose parameters are adapted to the silicon platform case;
- a nonlinear RR model based on the temporal couple mode theory (TCMT), which we adapted to the specific case of our plastic node by properly accounting for the high absorption and asymmetry in the ring due to the phase change material.

In particular, we were interested in developing a modular model that can easily be used as building block in larger circuit simulations, comprising several photonic devices. For this purpose, we also restructured the optical equations so that we could efficiently employ our model in a modular way within a commercial software for system-level photonics simulations (*Luceda Caphe*).

The main input and output quantities of the basic N-vN cell model are respectively the input optical power and the complex effective index of the waveguide covered with PCM (GST). Therefore, keeping in mind the configuration we want to simulate, the simplest way of combining this model with a RR model is to run the two models in parallel so that, at each time step, the RR provides the input optical signal to the GST waveguide segment, which in turn updates its complex effective index. In this case, a change in complex effective index of the ring waveguide segment simply results in a variation of the attenuation and of the phase shift of the light passing through it, respectively resulting in a change of width and position of the RR resonance peak in the frequency domain. Also note that the complex refractive index provided by the GST model depends both on the corresponding temperature at that location of the waveguide, and on the degree of amorphization of the GST layer.

## 2.1 ENERGY EFFICIENCY AND SPEED IMPROVEMENT OF MEMORY OPERATIONS

By accounting for silicon nonlinear effects, which shift the RR resonance wavelength during the excitation by pump pulses, our model allows to predict the optimal RR power coupling coefficients (i.e. how much each ring is coupled to the straight waveguides, see Fig. 2) and wavelength detuning (i.e. the difference between resonance wavelength and input laser wavelength) that maximizes energy efficiency of memory operations. Such a prediction can provide important qualitative insight, but also quantitative insight once accurate model parameters are experimentally obtained.

We run several simulations with different values of GST cell length, input pulse power, RR coupling coefficients and wavelength detuning. We found that a favourable GST cell length is 0.7µm. Setting the GST cell to this length, the minimum pulse power required to reach the maximum amorphization fraction is around 26mW with 10ns duration (see Fig. 3 (a) for the full parameter sweep results). Clearly, this type of information could not be obtained from simulations that do not account for the relevant nonlinear dynamics. For example, without nonlinear dynamics, the resonant wavelength of the RR would be fixed and therefore we would wrongly obtain that a null wavelength detuning optimizes the energy efficiency of memory operations.

Furthermore, employing the same optimal parameters and after an exhaustive exploration of different input pulses combinations, we found that full recrystallization (i.e. nullifying the maximum achievable amorphization) could be obtained by inserting four pulses of 10ns duration: the first of around 23mW

[11] Carrillo, Santiago García-Cuevas, Alessio Lugnan, Emanuele Gemo, Peter Bienstman, Wolfram HP Pernice, Harish Bhaskaran, and C. David Wright. "System-level simulation for integrated phase-change photonics." *Journal of Lightwave Technology* 39, no. 20 (2021): 6392-6402.

and the others of around 7mW. The aforementioned input pulses for amorphization and crystallization are shown in Fig. 3 (b), and the corresponding temporal evolution of the states in the developed model is represented in Fig. 3 (c). In order to obtain access to the steady state values without running the simulation for an unnecessary longer time, we forced the volatile variables back to their initial values after the amorphization and crystallization pulses (artificial cooldown), which has the same effect of waiting until the system reaches its equilibrium again.

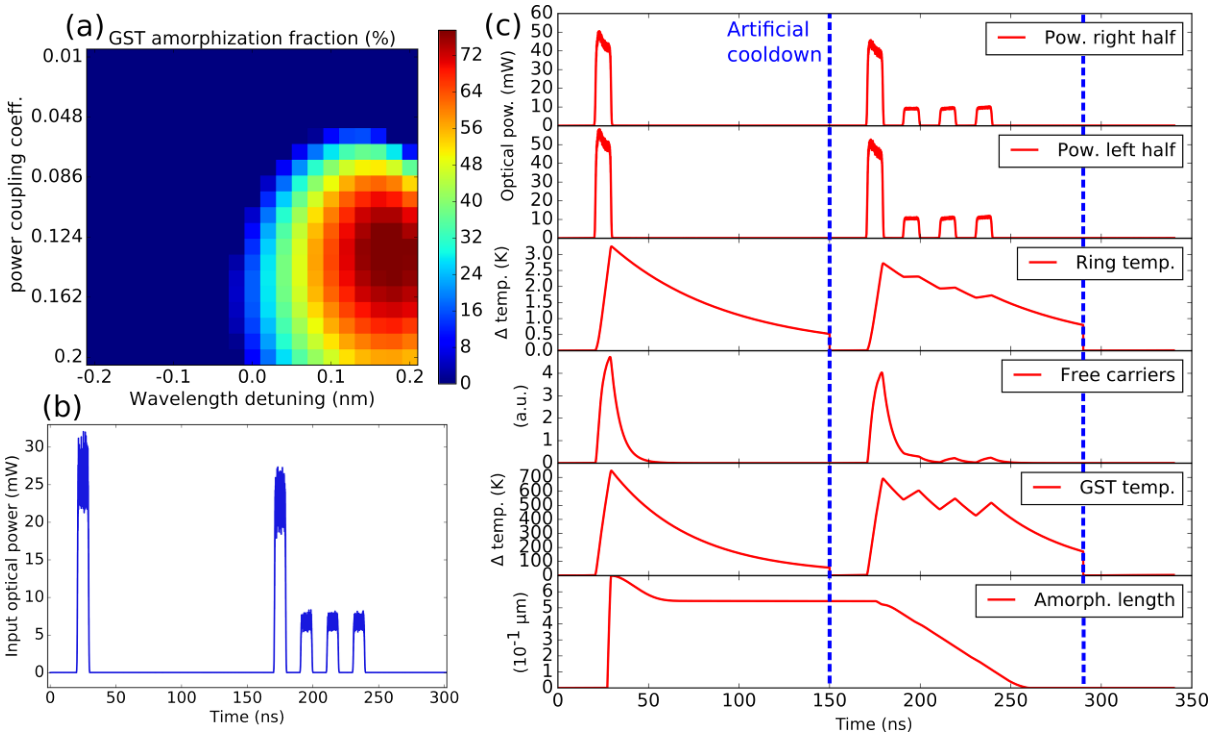


***Figure 3*** (adapted from[10]). (a) Colormap of the GST amorphization fraction as a function of resonance wavelength detuning and of the RR power coupling coefficient. An optical pulse of 10ns duration and around 26mW power was considered for amorphization. (b) Inserted pulse sequences for fast and energy efficient amorphization (1st pulse) and subsequent crystallization of the GST cell. (c) Corresponding time evolution of the model states during the insertion of the amorphization and crystallization pulse sequences, in (b).

We made a comparison between these simulation results and the results from the employed basic N-vN cell model (straight Si waveguide with 4μm GST cell, validated by fitting the experimental data reported in[9]). With such a comparison, we aim to show, at least qualitatively, the advantages enabled by the RR resonance in terms of speed and energy efficiency of memory operations. To start with, the maximum transmission contrast achieved with our plastic node model is around 15% considering a single input. To obtain such a contrast value, we considered an input amorphization pulse of 10ns conveying only around 0.26nJ energy, while full recrystallization is achieved within 70ns with less than 0.5nJ. In comparison, to achieve a similar contrast considering the basic N-vN cell model (based on straight silicon waveguide), a 100ns long amorphization pulse with energy of 1.6nJ is required. This shows that, thanks to the RR resonance, both speed and energy efficiency of memory operations are substantially improved. Moreover, to obtain the same amorphization length of around 543nm, the GST on waveguide model requires a pulse more than 50% longer (using the same peak optical power) and it only achieves less than 3% transmission contrast, compared to around 15% for the RR device case. This result indicates that the optical resonance improves memory operations by reducing the energy required to amorphize the GST and, at the same time, by enhancing the sensitivity to changes in GST material phase distribution. In addition, the same input pulse that causes the maximum contrast in the case of the simulated RR with GST, is not able to modify the GST crystallinity fraction

in the case of its straight waveguide counterpart (see Fig. S4 (c) in the supplemental document). Finally, full recrystallization in the straight waveguide simulation is achieved employing a double-step pulse (as in[9]) 530 ns long and with total energy of more than 3.9nJ. Also in this case, the RR enables a significant improvement in both speed and energy efficiency.

## 2.2 CONTRAST AND ENERGY EFFICIENCY OF READING THE MEMORY STATE

The observed large advantage in terms of speed and energy efficiency of memory operations is relevant only if the memory state can be easily read and used to control subsequent photonic operations. In order to take this aspect into account, in addition to calculating the *transmission relative contrast* (i.e. the difference in transmitted power after and before amorphization, divided by the latter value), we also consider the corresponding *transmission absolute contrast* (i.e. the same difference in transmitted power but divided by the input power), for each output port. The latter provides important information regarding the optical loss of the memory state reading, as it is relative to the total input power.

Moreover, we also monitor the *phase contrast* (i.e. the difference in optical phase of the output after and before amorphization, divided by π), which can be converted into transmission contrast through optical interference. In particular, if the memory device is used as a component of a coherent optical circuit, e.g. as non-volatile weight or as plastic synapse for neuromorphic computing, having a high enough phase contrast can be key, even if the transmission contrast is low. Finally, in order to measure the optical loss corresponding to the phase contrast, we also monitor the *transmission* of each output port after amorphization (given by the fraction of the output power w.r.t. the total input power).

We explored how these measures vary in different configurations by performing, as before, a parameter sweep in coupling coefficient and resonance wavelength detuning, but this time using constant low optical power as input, so as to leave the RR in the linear regime. At first, we considered a single input port and, considering the two extremes in the achievable crystalline fraction range of the GST cell, we obtained a maximum transmission relative contrast of around -15% (Fig. 4 (a)) at the *through* port (that is the output port on the straight waveguide where input is inserted), whose absolute value is very similar to the maximum value reported for a straight silicon waveguide with GST[9]. The *drop* port (that is the output port on the straight waveguide opposite to the input waveguide) shows a maximum value of around 21%, with the high contrast area in the parameter space not well overlapping with the *through* port case. But for both output ports, the transmission absolute contrast is always below 3.5% in absolute value (Fig. 4 (b)). This indicates that accessing the memory state in the linear regime is quite inefficient in terms of energy. Considering this input configuration, the phase contrast was negligible.

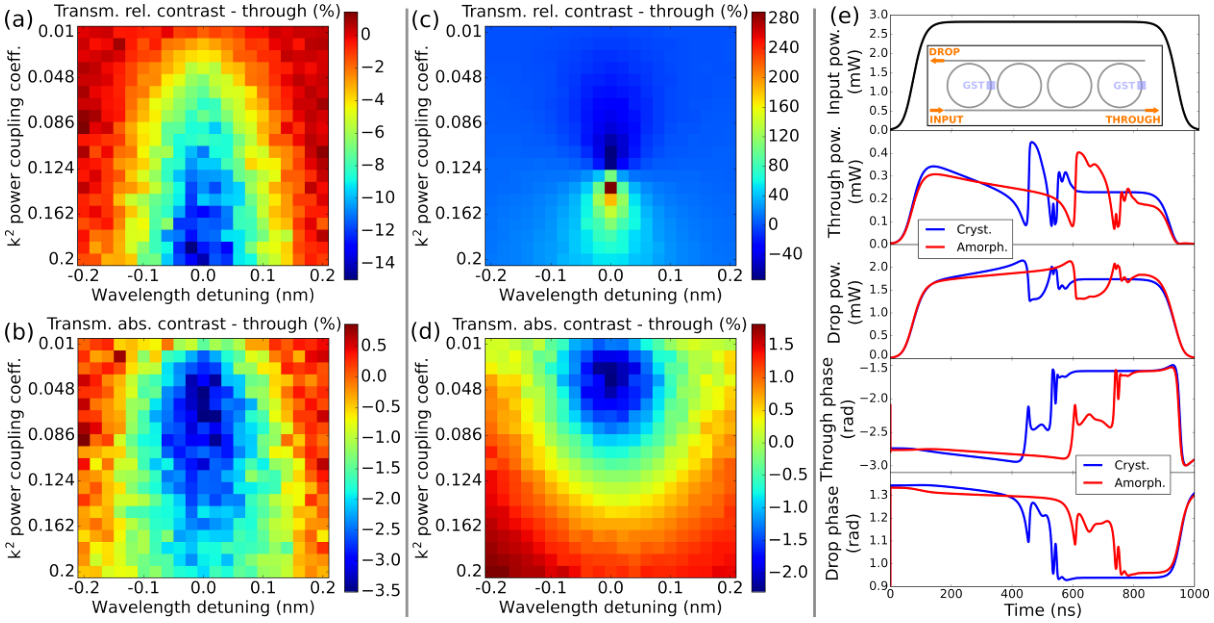


***Figure 4*** (adapted from[10]). (a), (b), (c), (d): Colormaps showing the results of the performed parameter sweeps in wavelength detuning and coupling coefficient, at the *through* port of the simulated RR with GST. (a) and (b): Transmission relative (and absolute, respectively) contrast when a single input port is excited. (c) and (d): The same quantities when both input ports are excited with same power and phase. The considered RR resonance has even azimuthal number. (e): Results of two time-dependent simulations, one for crystalline (red) and the other for amorphized (blue) GST cells, of the silicon photonic circuit schematized in the inset of the top plot. From top to bottom, plots showing input optical power, output power and optical phase at *through* and *drop* ports, as a function of time.

Nevertheless, we found that the achievable contrast values can be significantly increased by exploiting the enhanced sensitivity of interference effects in a RR. In particular, by inserting the same constant optical power in both input ports (on opposed straight waveguides), we achieved transmission relative contrasts up to 280% at *through* port (Fig. 4 (c)) and up to 350% at *drop* port. Still, we notice that the access to the memory state is not energy efficient, being the transmission absolute contrast not higher than 2.5% in absolute value (Fig. 4 (d)). Moreover, it should be stressed that when two inputs are considered, very different responses are obtained depending on whether the resonance azimuthal number is even or odd. In this case we considered an even azimuthal number, namely 164. This value is found by setting the resonance wavelength to 1550nm and the RR radius to about 15nm.

Another way to enhance the impact of the GST crystalline fraction on the response of a photonic circuit, is through the excitation of dynamics provided by nonlinear optical cavities. To show this for the first time, we employed our model as a building block of a small network of four RRs (Fig. 4 (e)), two of which with a 0.7µm GST cell. The execution of the 1µs simulation (more than 330,000 time steps) required less than 40s on a fairly standard desktop computer (processor: Intel(R) Xeon(R) CPU E5-2650L v3 @ 1.80GHz) employing up to around 0.5 GB of RAM. The two RRs without GST have a relatively small power coupling coefficient (around 4%), so that the power enhancement in the ring waveguide allows to trigger nonlinear effects without reaching input power levels that can change the GST memory state. Considering suitable design parameters for this novel device, we obtained strong contrast in the time-dependent network response, both in terms of output power and optical phase (Fig. 4 (e)). Importantly, it should be noticed that this is achieved with relatively low overall power loss. In particular, such an effect is relevant for building plastic recurrent NNs, where the network topology is expanded along the time dimension, and where we believe the potential of the modelled device can be exploited at its fullest.

# 3 SYSTEM-LEVEL SIMULATIONS: A SCALABLE ON-CHIP TRAINING APPROACH

In this section we propose and investigate, by means of numerical system-level simulations, a novel on-chip training approach based on plasticity, to optimize the non-volatile weights (given by the memory states of the aforementioned plastic nodes) of our plastic RC system. The proposed training method does not require to tune the network weights one by one nor to observe the internal states, but it only requires to update the input optical signal on the basis of the optical output of the reservoir. Therefore, it potentially overcomes the scalability limitations suffered by other common training approaches, such as the ones based on backpropagation. It should be stressed that, in order to provide accurate quantitative predictions, our plastic node model should be first fit and validated using direct experimental data. Instead, in our model we employ experimentally validated parameters for the basic N-vN unit cell model and the nonlinear silicon RR model separately. Nevertheless, the presented simulations provided important qualitative insight, useful to design the final hardware demonstrator for a self-learning photonic RC system.

We simulated several variations of photonic dynamical networks like the one in Fig. 2 by just connecting together many instances of our plastic node model (described in the previous Section), within the Luceda Caphe simulator framework. It is sufficient to set the GST cell length to zero to obtain a silicon RR building block without PCM cell. The parameter space corresponding to different possible versions of simulated networks is extremely large: we can sweep over the distance between the RRs w.r.t. their radius, the number of node rows and columns, the number and position of RRs with GST cell w.r.t. the ones without GST cell, the characteristics of single nodes such as coupling coefficient and GST cell length, and so on. For this reason, it was impossible to perform an exhaustive parameter exploration without limiting the number of RRs too much. Still, we explored different network versions trying to achieve a good balance between plasticity and nonlinearity (the RRs with GST cells have weaker nonlinear behaviour because the PCM cell introduces a significant power loss). For our numerical proof of concept, we selected the network version schematized in Fig. 5.

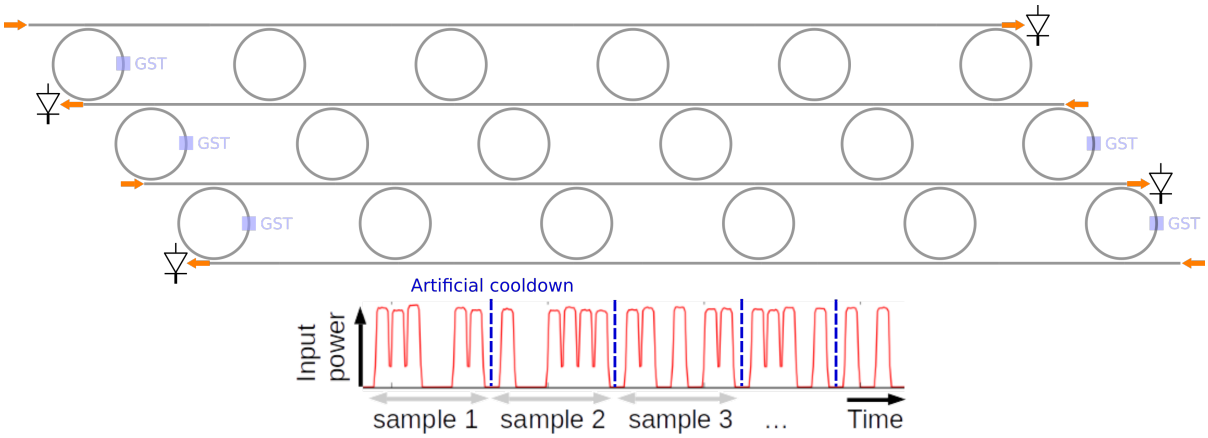


***Figure 5***. On top: schematic of the plastic reservoir network employed to obtain the presented system-level simulation results. Arrows pointing towards the straight waveguides indicate the input ports; arrows pointing outwards indicate the output ports, whose optical intensity is acquired by simulated detectors and sent to a linear classifier (logistic regression with optimized L2 regularization strength), in accordance with the RC paradigm. At the bottom: example of optical input power as a function of time. 15 different types (classes) of 5 optical pulse sequence are inserted with random order. An artificial cooldown between each input sequence simulates the case where the sequences are separated by a time interval that is long enough for the system to reach thermal equilibrium.

At the input ports of the simulated network we continuously insert 15 different types (classes) of 5 optical pulse sequences in random order, one after another. Thermal equilibrium is forced between

one sequence and the other (to avoid unnecessarily longer simulation times) simulating the case where a long time interval separates the sequences. In machine learning terms, each 5 pulse sequence is a class (therefore we tackle a 15-classes classification problem) and each sample fed into the linear readout classifier is a vector with a number of elements equal to the number of reservoir outputs (that is 4 in the case depicted in Fig. 5). Each element is the optical energy at the corresponding output port, acquired over a time interval corresponding to the last pulse of each sequence. This is just one of the many possible feature extraction method; we considered this one because, thanks to the volatile memory provided by silicon nonlinear effects in the RRs, information of how the pulses of a sequence affect the network should be available in how the network responses to the last sequence pulse. It should be stressed that the intensity of the inserted pulse sequences is high enough to trigger nonlinearity in RRs without GST, but not to modify the solid-state phase of the GST cells.

In order to train our simulated reservoir in a scalable way through plasticity, we employ what we call a *pumped version* of the sequence classes. E.g., the pumped version of class 1 is the same sequence corresponding to class 1, with the only difference that the last pulse has a higher peak power, potentially enough to change the memory states of the GST cells, triggering plasticity. In particular, the proposed reservoir training method consists of the following two steps:

1. *Readout training* step: this corresponds to the traditional training and performance evaluation of a RC system. In particular, all the sequence classes are inserted (e.g. see the bottom of Fig. 5) and the readout linear classifier is trained on these. Then, the classification error of the trained classifier is estimated for each individual class.
2. *Probe and pump* step (Fig. 6): we select one of the sequence class that scored worst (i.e. that was classified with the highest error) in the previous step and we insert a signal into the reservoir with the scope of improving the classification of the selected class by means of reservoir plasticity, without retraining the readout classifier. In particular, we repeatedly insert the selected sequence class (probe sequence) followed by its pumped version (pump sequence, see top of Fig. 6). For each inserted probe sequence, an online classification error estimation is performed, employing the linear readout trained in the previous step. The intensity of the last pulse of the following pump sequence (which determines how likely it is that the memory state of the GST cells are changed) is set to a value proportional the online estimated error. This way, intuitively, the state of the GST cells might be changed until the estimated error is reduced enough so that the pump is not anymore strong enough to make further changes. The parameters of this feedback loop between the error estimation and the pump intensity required accurate tuning to achieve convergence to a memory configuration corresponding to a classification improvement. For example, at the bottom of Fig. 6, we can see that the online error estimation for the inserted sequence class has decreased with time w.r.t. its initial value.

These two steps can be employed to improve the classification of a specific class sequence, or they can be repeated to try and improve the multi-class classification performance on multiple types of sequences.

In Fig. 7 we show an example application of the proposed training method for our plastic photonic reservoir. The first plot on the left shows the classification performances obtained after the first *readout training* step, before the memory states of the GST cells are changed. The sequence class with label 8 is selected for the first *probe and pump* step, which causes the GST amorphization fraction of the memory cells to change over time as shown in the plots on the right. Afterwards, the second *readout training* step is performed and we can notice (in the second plot on the left) that the error of

classifying the selected sequence class was significantly reduced thanks to the plastic reservoir training. We then choose to tackle the sequence class with label 4. After the second *probe and pump* step, we see (in the last plot on the left) that the error rate corresponding to the selected class was significantly reduced again. Importantly, the classification improvement (regarding class label 8) achieved by the first *probe and pump* step appears not to be significantly undermined. That is, in this example we do not see a strong *catastrophic forgetting* effect[12].

Finally, while these simulations provide interesting results and important insight into the problem of exploiting plasticity for on-chip scalable training of our plastic reservoir, it is important to stress again that the employed node models should be experimentally validated before we can obtain reliable quantitative results from these system-level simulations. In the next section we show some selected results of a direct experimental investigation regarding our plastic node. Unfortunately, in the context of this project, we did not have the time to repeat these system-level simulations with experimentally validated parameter.

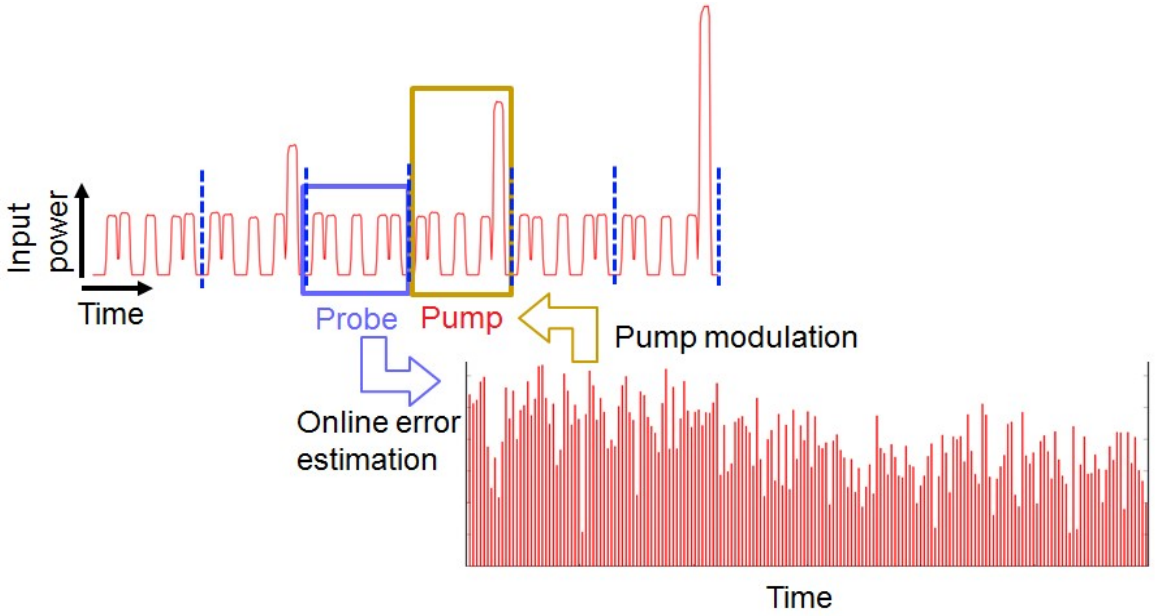


***Figure 6.*** Schematic of the *probe and pump* step of our method to train the photonic reservoir by means of plasticity.

[12]French, Robert M. "Catastrophic forgetting in connectionist networks." Trends in cognitive sciences 3.4 (1999): 128-135.

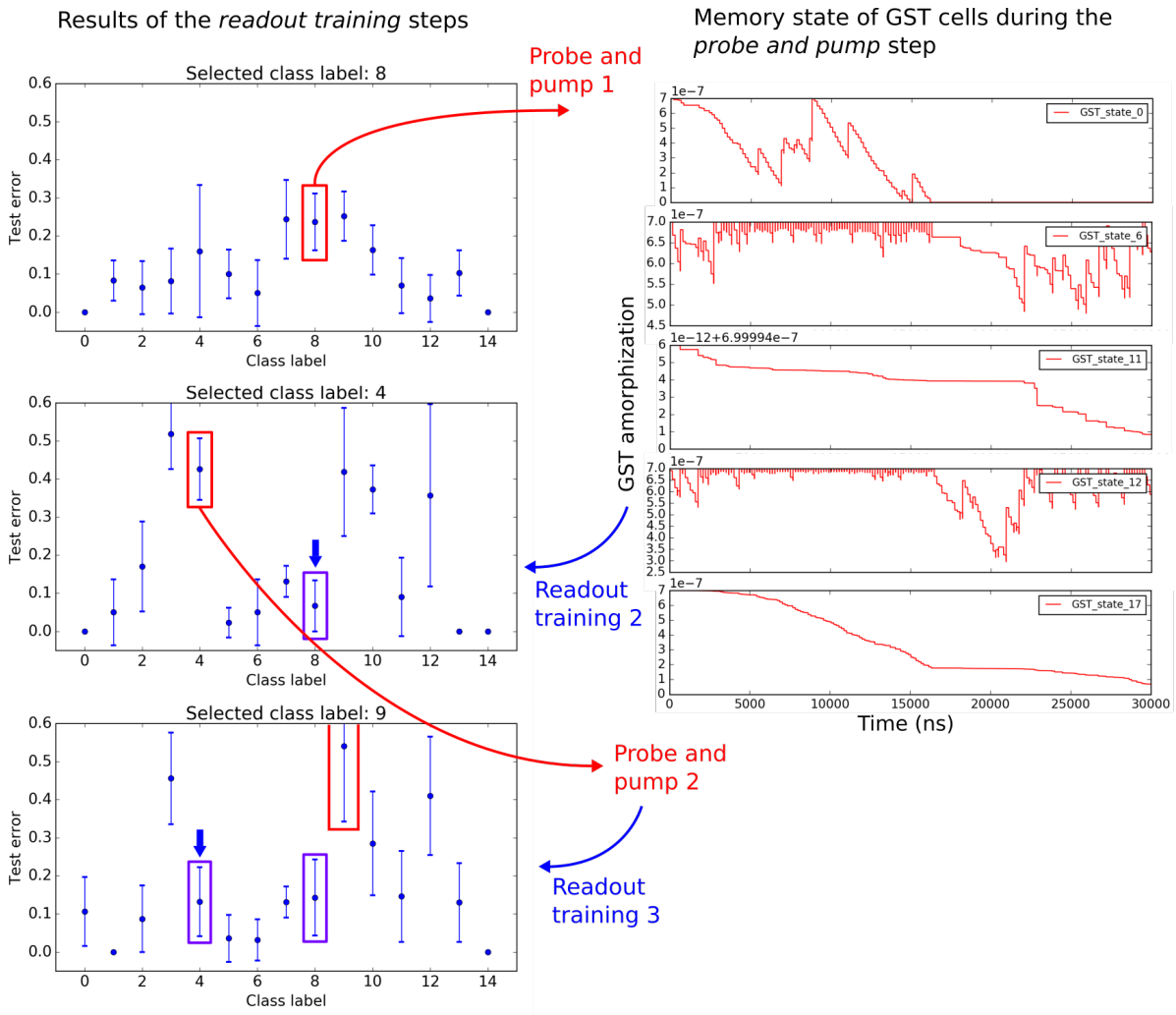


***Figure 7.*** Example application of the proposed plastic reservoir training method. On the left, plots showing the estimated classification performance (error rate) after subsequent *readout training* steps. On the right, the GST amorphization fraction (i.e. the non-volatile memory state) of the 5 GST cells in the simulated reservoir network (see Fig. 5), corresponding to the first *probe and pump* training step.

# 4 EXPERIMENTAL INVESTIGATION OF THE PLASTIC NODE DEVICE (SILICON RRS WITH A SHORT GST CELL)

Before designing and fabricating a photonic chip with plastic reservoir networks, we had to investigate the properties of the building blocks (that is, of the presented plastic node) we could realistically fabricate. In particular, an important requirement was to successfully deposit short enough GST cells on chip (GST deposition was performed by OXFORD partners), so to obtain plasticity without overly reducing the energy efficiency and scalability of the network. According to simulations, optimal GST lengths were under 1µm while the Fun-COMP partners usually employ GST lengths of more than 2µm. It should be stressed that fabricating good silicon RRs with short enough GST cells resulted quite challenging, taking several fabrication runs and requiring intense and efficient collaboration between the project partners involved (IMEC, OXFORD and MUENSTER). Moreover, the achievement was significantly delayed because of COVID19 regulations. Finally, we could obtain and measure satisfying building block devices via the following plan:

1. MUENSTER partners wrote the silicon waveguides on chip through electron-beam lithography.
2. OXFORD partners performed etching and GST deposition.
3. IMEC partners measured the fabricated devices.

An important requirement to obtain energy efficient plastic nodes is to fabricate RRs with suitable coupling coefficient, given a certain GST cell length. Indeed, to obtain critical coupling and therefore to minimize the energy cost of plasticity, the longer the GST cell the higher the required coupling coefficient (and thus the smaller the coupling gap, i.e. the distance between the straight waveguides and the ring waveguide). In practice, a 2D parameter sweep (over the GST length and the coupling gap) had to be fabricated, in order to determine the best combination between the two parameters (a view of our design is showed in Fig. 8). We also tried two different values for the RR radius, namely 7µm and 15µm (the latter showed better performances and was then selected for the plastic reservoir network design).

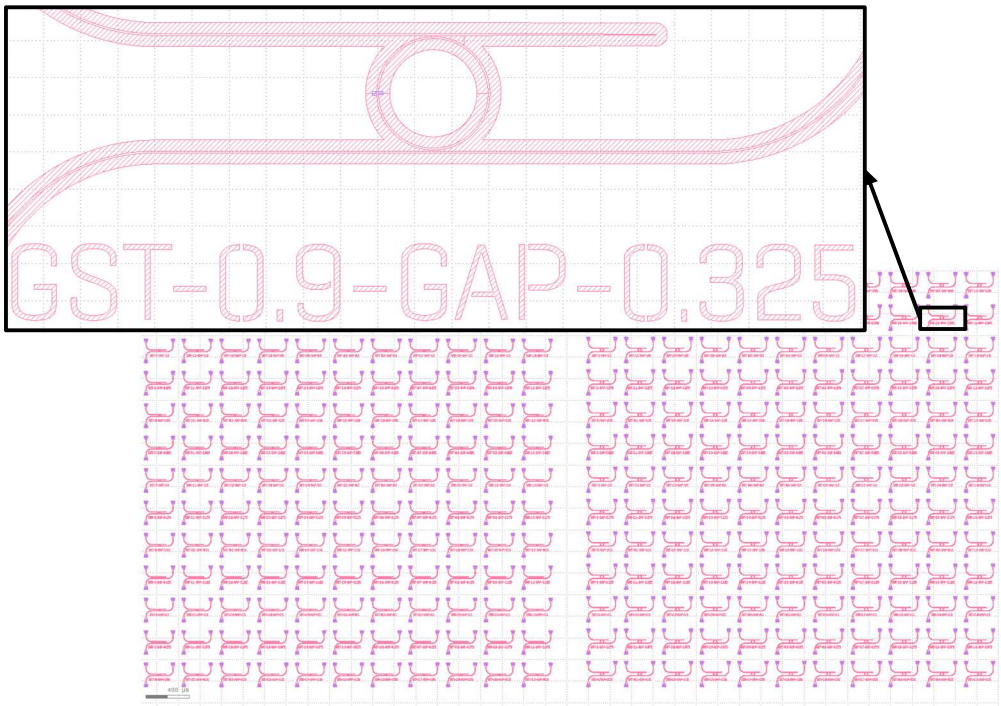


***Figure 8.*** View of our design comprising two 2D parameter sweep of our single building-block device (plastic node). The silicon RRs in the left block have a radius of 7µm, the ones in the right block have a radius of 15µm. In each block, along the vertical direction the RR coupling gap is changed, along the horizontal direction the GST cell length is changed.

Eventually, we could consistently fabricate devices with a GST cell length not shorter than 1µm. However, fortunately this provided satisfying energy efficiency and contrast of memory operations (significantly more than what our simulations predicted), given a suitable RR coupling gap.

In order to extract the most relevant information from the relatively large number of fabricated single devices, we performed the following two types of measurements:

- Automatic transmission spectra acquisition: by means of our automatic setup (it can automatically align input and output optical fibers to the grating couplers on the photonic chip) we measured the response of the RRs at both output ports to a low-power and constant input laser beam for different wavelengths around 1550nm (e.g., from 1530nm to 1570nm

with 0.01nm steps). The acquired spectra provided important information on the energy efficiency of the fabricated devices and on the impact of fabrication errors.

- High-power high-speed measurements: optical pulses (with durations from few μs to tens of μs and peak power around 10mW on-chip) were inserted into the plastic nodes to change the memory state of the GST cell. Then, spectra were acquired to check the impact of those changes on the linear transmission of the RRs. These measurements provided important insight into the energy efficiency and contrast of memory operations (considering both amorphization and recrystallization of the GST thin film).

More details regarding the outcome of these measurements are given in Deliverable D3.3.

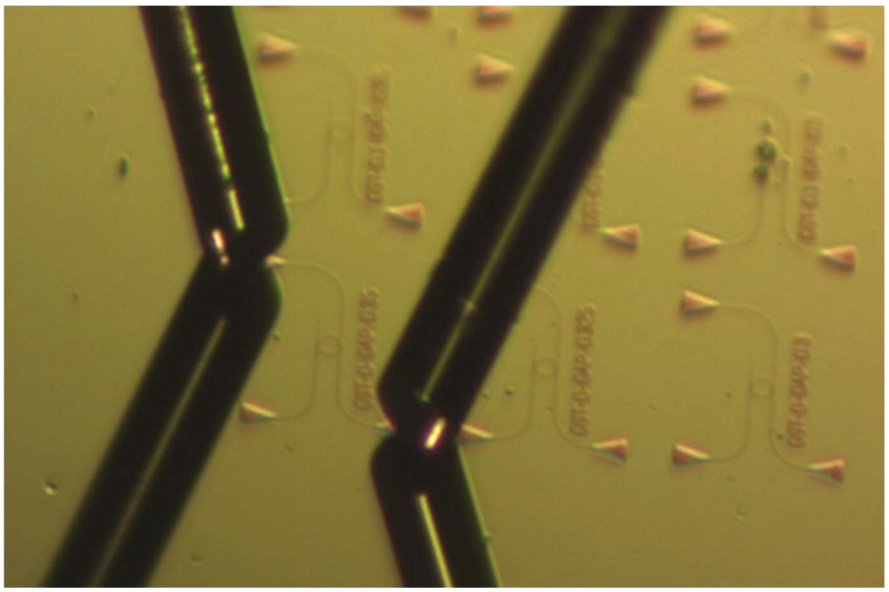

***Figure 9.*** Magnified picture of fabricated single plastic nodes on a silicon-on-insulator (SOI) chip. One device is being measured by coupling its grating couplers with two (input and output) optical fibers.

# 5 DESIGN AND FABRICATION OF PLASTIC RESERVOIR NETWORKS WITH SELF-LEARNING CAPABILITIES

Exploiting the insight regarding building-block parameters and network topology obtained through the simulations and measurements presented in the previous sections, we have designed several versions of the proposed photonic reservoirs, for two different approaches of fabricating silicon photonic integrated circuits:

- Fabrication through electron-beam lithography by the MUENSTER partners (a design overview is shown in Fig. 10), that is the same approach we used to fabricate the chips for single plastic nodes investigation, as presented in the previous section. This option allows to fabricate few chips relatively fast, but there is a significant chance that a fabrication step goes wrong, which can sometimes be found out only after measuring the chips. Moreover, the properties of nominally the same photonic component might change significantly from one chip to another, or even depending on the position on the chip area. Another downside is that the fabrication errors affecting waveguides and grating couplers limit the quality (Q-factor and finesse) of RRs and the efficiency of coupling the measurement setup with the photonic circuits, which in turn limits the scalability of viable reservoir networks.

- Fabrication by Imec's foundry[13] (a design overview is shown in Fig. 11), which is costly and can take several months (up to one year) but allows to obtain several tens of high quality silicon photonic chips. This is a highly reproducible and CMOS compatible fabrication, providing low-loss waveguides and highly efficient grating couplers, thus allowing to obtain more energy efficient and scalable silicon RRs networks. A downside regarding our specific case is that the OXFORD partners have to accurately etch windows through the chip's oxide overlayer in order to deposit GST on the waveguides. This is a difficult operation that requires many trials.

We chose to go for both options in parallel in order to increase the likelihood to obtain at least one chip with working plastic reservoir networks. Fortunately, we could achieve this via the first option (see picture of part of the integrated photonic circuits in Fig. 12), while the second fabrication option took significantly more time than expected (at the moment – i.e. June 2022 - we have just obtained some first chips without GST cells, but we still have to test them). In any case, while the chips obtained from the first fabrication option contain an important first hardware demonstrator for proof-of-principle measurements, the chips from the second option will allow us to investigate more scaled-up and advanced neuromorphic computing applications.

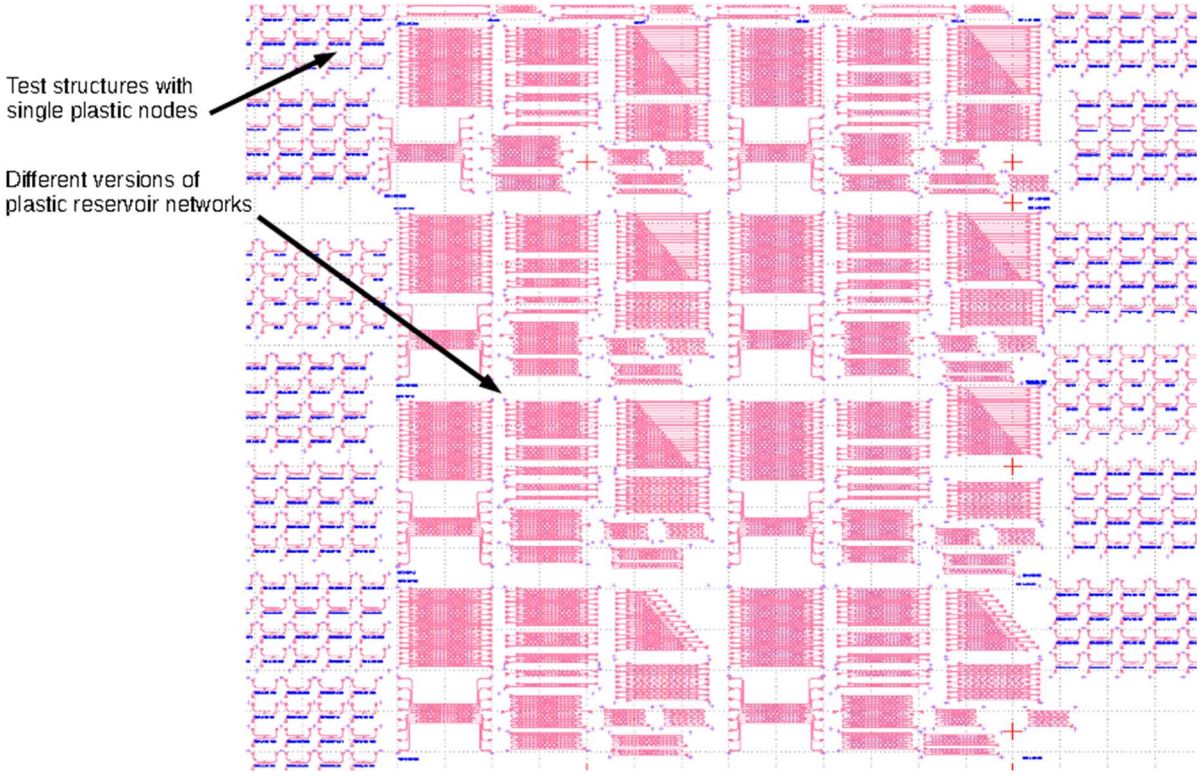


***Figure 10.*** Photonic integrated circuit design for fabrication through electron-beam lithography (first fabrication option). The design comprises single-device test structures (at the sides) and different versions of plastic reservoir networks, with variations in topologies, network dimensions, single-devices properties (such as RR coupling coefficients) and GST cells density.

[13] See for example https://europractice-ic.com/technologies/photonics/imec/

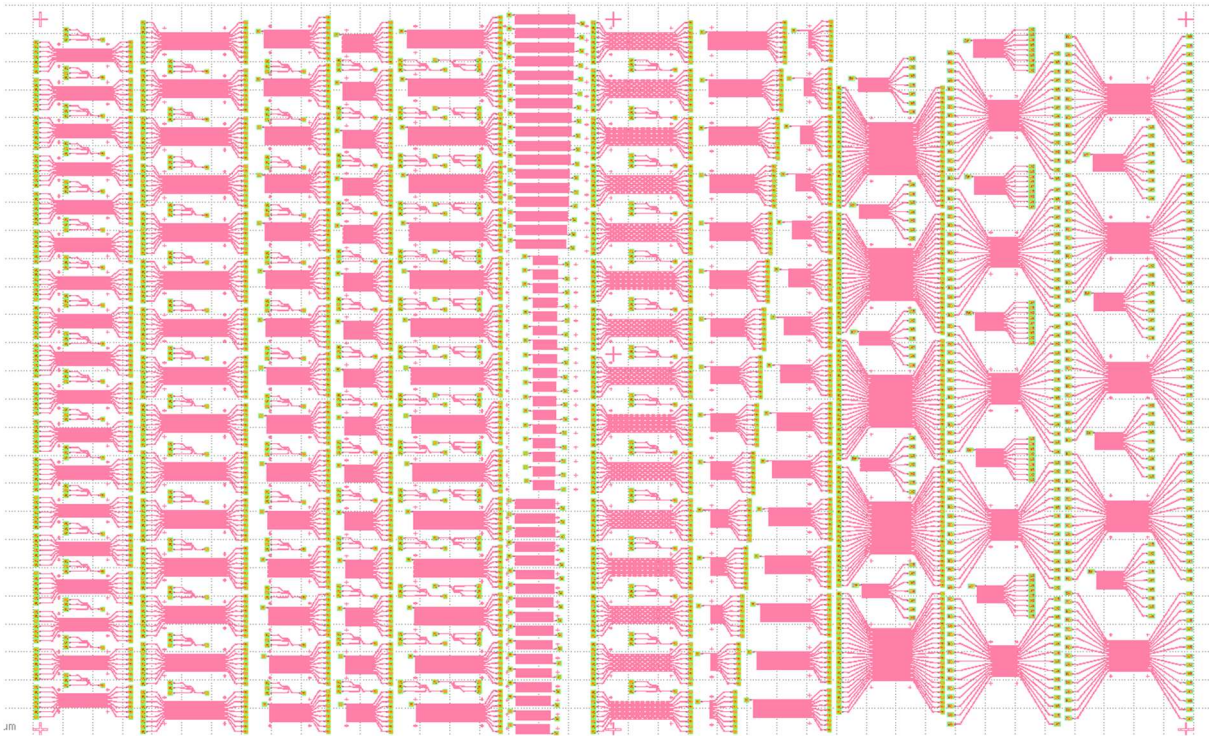

***Figure 11.*** Photonic integrated circuit design for fabrication by Imec's foundry (second fabrication option). The design comprises single-device test structures and different versions of plastic reservoir networks, with variations in topologies, network dimensions, single-devices properties (such as RR coupling coefficients) and GST cells density. In particular, the higher waveguide quality provided by this fabrication option allowed us sweep over a larger range of RR coupling coefficients.

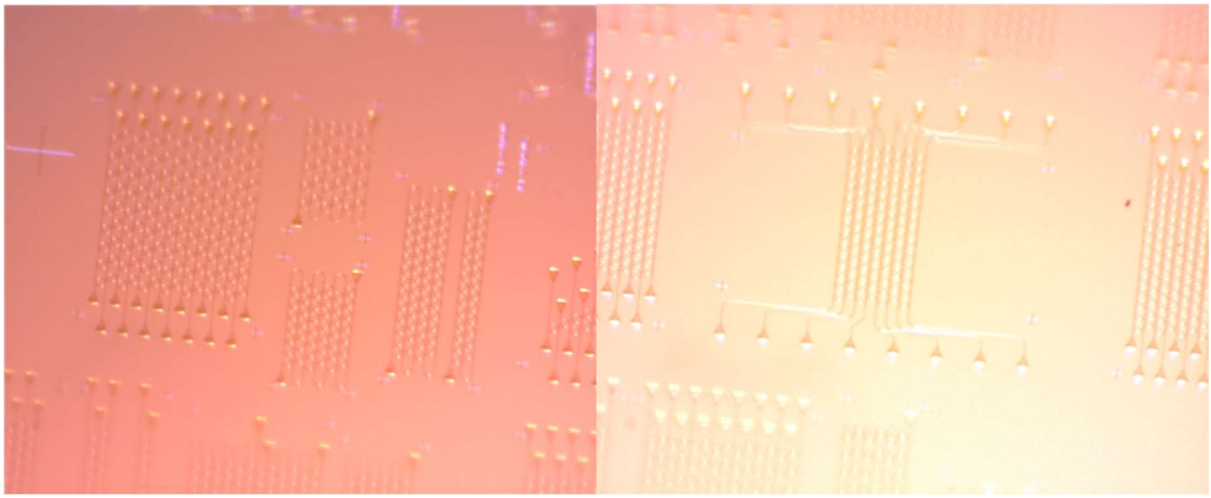

***Figure 12.*** Magnified picture of fabricated plastic reservoir networks on a silicon-on-insulator (SOI) chip.

In order to demonstrate self-learning RC, we inserted different classes of pulse sequences into different reservoir network versions (similarly as in the system-level simulations described in Section 3) and we first investigated how the output was modified by the solid-state phase of the GST cells that were plastically adapting to the input signal. Promisingly, we observed that the plastic non-volatile changes had a significant impact on the output signal, and that this impact was strongly dependent on the previously inserted sequences. Importantly, such a rich plastic behaviour was detected concurrently with a strong nonlinear distortion of the input signal, ascribed to the silicon nonlinear effects affecting the resonant behaviour of the RRs in the network. Therefore, as planned (see Section 1), our reservoir networks showed both volatile and non-volatile memory, together with nonlinearity.

Afterwards, in accordance with the RC paradigm, we trained a linear classifier to classify the input pulse sequences by employing the reservoir output. We observed that the classification performance was significantly improved by the reservoir presence and that such an improvement was strongly modified by the plastic adaptation of the network. The details regarding the employed reservoir networks, measurements and results are presented in Deliverable 3.3.

# 6 CONCLUSIONS

We have discussed the most relevant aspects and steps of our research on the development of a reservoir computing system based on an integrated photonic network, which, enhanced by N-vN unit cells, can plastically adapt to its input and has self-learning capabilities (i.e. the performance of a tackled machine learning task can be improved by the autonomous adaptation of the network, without the aid of an external learning algorithm that tunes the network parameters). In particular:

- We developed an efficient numerical model of a silicon ring resonator with a GST cell on its waveguide, which accounts for the nonlinear dynamics due to the silicon nonlinear effects affecting the ring resonance, and for the solid-state phase change of the GST thin film due to optical pulses (all-optical non-volatile memory operations).
- By employing the developed model, we showed via numerical simulations that this device (especially when short GST cells (≤ 1μm) are employed) is a promising candidate for a plastic node in a photonic reservoir. In particular, the combination of resonant behaviour and nonlinear effects allows to enhance the energy efficiency of memory operations and the scalability of networks based on this building block.
- By connecting multiple instances of the developed numerical model, we simulated networks comprising several silicon RRs (with and without GST cell) and showed that these can operate as photonic reservoir to carry out machine learning tasks such as classification of optical pulse sequences. Moreover, we showed that the classification performance of a given sequence class can be improved by means of a scalable training algorithm that exploits the reservoir network plasticity. Importantly, such a training method does not require external tunability of the network weights and observability of the internal states, as opposed to non-biologically-plausible training algorithms based on backpropagation.
- We designed, fabricated and measured several versions of the plastic node (silicon RRs with a GST cell) in order to experimentally investigate the energy efficiency and contrast of non-volatile memory operations, and to determine the most suitable design parameters for deployment in plastic reservoir networks.
- We designed several versions of plastic reservoir networks for two different silicon photonics fabrication options. One is performed by the project partners and takes a relatively short time, but provides less reproducible circuits and larger fabrication errors affecting the basic photonic components. The other takes a longer time, but provides several tens of high-quality photonic chips. Via the first fabrication option, we managed to fabricate working plastic reservoir networks. In particular, we found that the performances of the RC system in pulse sequences classification is strongly dependent on the non-volatile memory state of the GST cells, that can in turn be changed by the input signal. Therefore, our plastic reservoir networks meet the basic requirements for self-learning behaviour based on network plasticity.

More details regarding the measurements and experimental results are presented in Deliverable 3.3.