POST-DIGITAL - European Training Network on Post-Digital Computing [GA860360]

Document Details

Title	D2.2 Report on the characterization of the photonic- and electronic- based analogue computing implementations
Deliverable number	D2.2
Deliverable Type	Report (public)
Deliverable title	Report on the characterization of the photonic- and electronic- based analogue computing implementations
Work Package	WP2 – Implementation and characterization
Description	Document describing the characterization of the different approaches to implement analogue computing in the POST-DIGITAL European Training Network and the level of progress of the ESRs
Deliverable due date	30/09/2022
Actual date of submission	29/09/2022
Lead beneficiary	CSIC
Version number	V1.0
Status	

Dissemination level

PU	Public	X
CO	Confidential, only for members of the consortium (including Commission Services	

Project Details

Grant Agreement	860360
Project Acronym	POST-DIGITAL
Project Title	POST-DIGITAL - European Training Network on Post-Digital Computing
Call Identifier	H2020-MSCA-ITN-2019
Project Website	https://postdigital.astonphotonics.uk/
Start of the Project	1 April 2020
Project Duration	48 months

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860360.

EXECUTIVE SUMMARY

This deliverable is the second of three deliverables in the POST-DIGITAL project describing the progress in building photonic- and electronic-based analogue computing devices. The present deliverable reports on the characterization of the hardware implementations developed within the consortium. Using photonic substrates, designs based in spatial, temporal or spatial multiplexing have been developed and tested. This report also includes the characterization of an alternative approach based on an electronic re-configurable neuromorphic chip that uses analogue processing with spiking dynamics. Along this deliverable, the level of progress of the ESRs is documented, and further steps towards reaching the final proof-of-concept results are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ACRONYMS

ANN	Artificial Neural Network
API	Application Programming Interface
AWG	Arbitrary Wave Generator
ESR	Early Stage Researcher
ETN	European Training Network
FPGA	Field Programmable Gate Array
HAR	Human Action Recognition
NMSE	Normalized Mean Square Error
NSM	Neural State Machine
POST-DIGITAL	Project ‘POST-DIGITAL - European Training Network on Post-Digital Computing’ EC GA 860360
PR	Power Ratio
RC	Reservoir Computing
SOA	Semiconductor Optical Amplifier
SRAM	Static Random-Access Memory
VCSEL	Vertical-Cavity Surface-Emitting Laser
WTA	Winner Take All

1 INTRODUCTION

This report covers the progress in the characterization of hardware implementations of analogue computing devices based on photonic and electronic platforms. The ESRs of the POST-DIGITAL ETN are exploring several hardware platforms and information processing strategies in the context of analogue computing. In the photonic domain, this report covers the advances in the exploitation of the corresponding physical substrates by multiplexing techniques, either in space (ESR3), time (ESR4, ESR5, ESR15) or frequency (ESR12). In the electronic domain, ESR1 and ESR2 have characterized the properties of the electronic mixed-signal neuromorphic hardware DynapSE2.

ESRs are therefore extending their skill set, including now different techniques to characterize hardware implementations. This report shows that the ESRs are already making a measurable impact in the field of post-digital computing. After the initial designs reported in Deliverable 2.1, the current Deliverable shows the details of an in-depth experimental characterization that will lead to the ultimate goal of high-performance applications.

List of contributors (alphabetical order): Steven Abreu (ESR1, Groningen), Mirko Goldmann (ESR5, CSIC), Alessandro Lupo (ESR12, ULB), Enrico Picco (ESR4, ULB), Guillaume Pourcel (ESR2, Groningen), Anas Skalli (ESR3, UBFC), Elger Vlieg (ESR15, IBM). Document compiled by Miguel C. Soriano, IFISC (CSIC-UIB).

2 DESIGN OF PHOTONIC- AND ELECTRONIC- BASED ANALOGUE COMPUTING DEVICES

There is a rapid growth in the use of artificial neural networks (ANN) to process data and solve problems in areas where traditional algorithms fail to surpass the capabilities of the human brain. The drawbacks of using conventional digital hardware when running ANN methods has led to physically implement them in analogue systems [1]. In the following, the activities of the different ESRs involved in the characterization of photonic (section 2.1) and electronic (section 2.2) based analogue computing devices are detailed. A list of references is provided in Section 2.3.

2.1 PHOTONIC SYSTEMS

As detailed in Deliverable 2.1, the different photonic designs follow complementary strategies that exploit multiplexing in either space (section 2.1.1), time (section 2.1.2) or frequency (section 2.1.3).

In the case of spatial multiplexing, this deliverable covers the work of ESR3 on the characterization of multi-mode optical systems as artificial neural networks implementations.

In the case of temporal multiplexing, ESR4 reports on the characterization of an opto-electronic Reservoir Computer based on a delay Loop and an FPGA. ESR5 and ESR15 are closely working on the characterization of delay-based photonic reservoir computers for high-speed operation.

Finally, ESR12 has characterized a photonic system that uses frequency multiplexing for reservoir computing.

2.1.1 Space multiplexing

In this section, the characterization of a promising approach to implement artificial neural networks and reservoir computing systems using spatially-multiplexed optical systems is presented.

The work of ESR3 has been focused on the characterization of the large area vertical cavity surface emitting laser (VCSEL) based ANN introduced in Deliverable 2.1. In particular, ESR3 studied how the performance varies with different physical parameters, namely, injection wavelength, injection power, and bias current. Furthermore, these physical parameters were linked to the general computational measures of consistency and dimensionality. Along such an experimental characterization, a general method of gauging dimensionality in high dimensional nonlinear systems subject to noise was introduced. This methodology could be applied to many systems in the context of neuromorphic computing. The fundamental characterization of the VCSEL-based approach paves the way for the use of similar photonic devices as building blocks for more complex hardware ANN platforms.

Figure 1 presents several aspects of the detailed characterization of the VCSEL-based ANN approach, which help in understanding the behavior and the computational performance of the analog computing approach developed by ESR3. Figure 1(a) shows how the VCSEL reacts to an injection drive laser at different injection wavelengths λ_{inj}. At around $\lambda_{inj} =$ 918.9 nm, the VCSEL locks to the injection laser and its free running modes are suppressed by around 10 dB. At this point, the VCSEL's own emission wavelength is shifted to that of the injection laser in a phenomenon called injection locking. Figure 1(b) shows how injection locking impacts performance. The injection wavelength is swept and the dependency of the NMSE on λ_{inj} is shown for different power ratios (PR) between the VCSEL and the injection for a 3-bit header recognition task. The VCSEL was biased 50% above its threshold and its emission power was 3.6 mW. There is a clear trend showing optimal performance around the resonance wavelength shown by the red dotted line. In addition, performance increases with higher PRs, yet it starts degrading after a PR higher than 1.

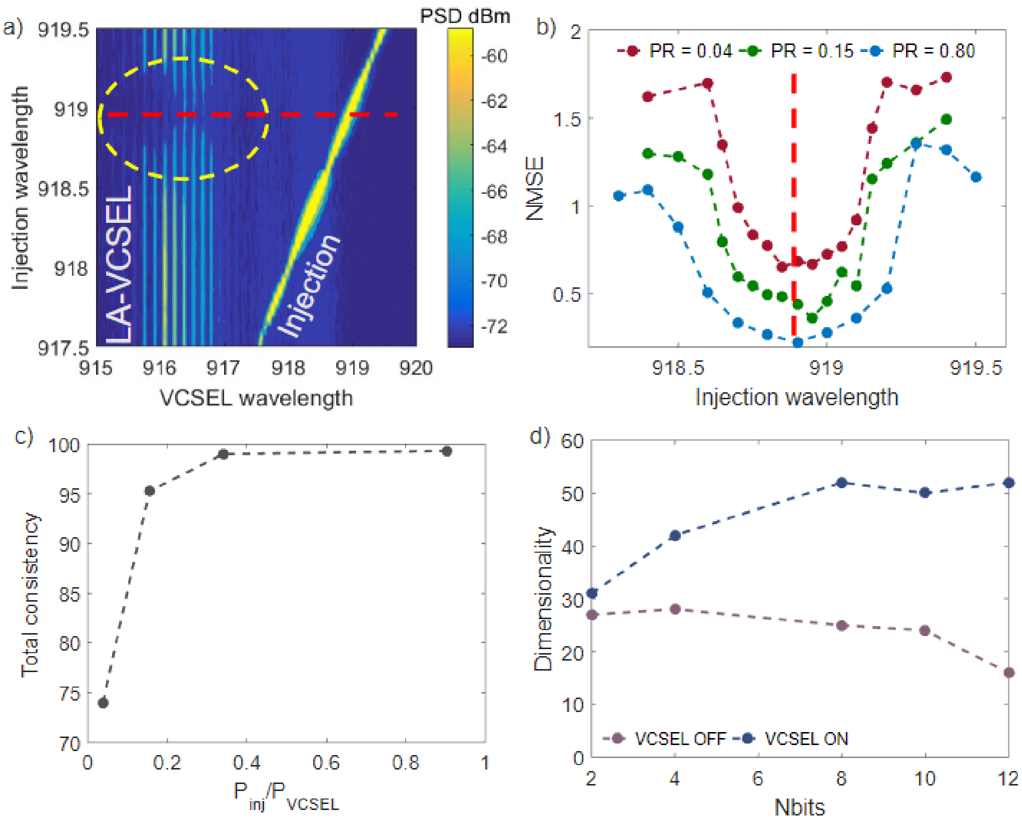

Figure 1 a) Injection locking of the VCSEL by an external drive laser. b) Performance (NMSE) for different injection wavelengths and powers. c) System consistency. d) Dimensionality with the VCSEL ON and OFF.

In addition to the characterization of the impact of the injection wavelength on the performance for a 3-bit header recognition task, consistency and dimensionality were also studied. Consistency is the ability of a given physical system to respond similarly when subjected to the same drive signal or input information multiple times. It is a crucial property when studying dynamical systems, especially when considering these dynamical systems as potential hardware candidates for neural network implementations. Indeed, a system that is not consistent to some degree would not be able to learn any task. Crucially, consistency serves as an upper bound for a given system's performance. Indeed, a system that is 95% consistent cannot achieve lower than 5% error for a continuous function approximation task. In turn, dimensionality is connected to the computational power of the device. A higher dimensional system will in principle be able to solve more challenging tasks. Figure 1(c) shows how the consistency saturates at a PR around 0.35 to nearly 100%. The high consistency value that was measured is promising as it shows the highly robust nature of the device. Finally, Fig. 1(d) shows how the VCSEL expands the dimensionality of the input data (ON vs OFF) for a photonic network consisting on 350 nodes/neurons. More info can be found at:

- Anas Skalli, Xavier Porte, Nasibeh Haghighi, Stephan Reitzenstein, James A. Lott, and Daniel Brunner, "Computational metrics and parameters of an injection-locked large area semiconductor laser for neural network computing [Invited]," Opt. Mater. Express 12, 2793-2804 (2022).

2.1.2 Time multiplexing

In this section, the characterization of delay-based photonic and opto-electronic approaches for reservoir computing is reported. This entails the work of ESR4, ESR5, and ESR15.

2.1.2.1 Delay-based reservoir computing based on a semiconductor laser subject to optical feedback

ESR5, in collaboration with ESR15 during his secondment at IFISC (CSIC), has numerically and experimentally characterized a photonic delay-based reservoir based on a semiconductor laser set-up with the objective to increase the processing speed of the system, using chaotic time series forecasting as a benchmark task. In a newly suggested time-multiplexed approach, the time per input is reduced by using a short temporal virtual node separation and by optimizing the number of inputs per reservoir delay.

The photonic system contains an injection laser in continuous wave mode that drives another semiconductor laser referred to as the reservoir laser. The injection signal is modulated using an arbitrary wave generator (AWG) and a Mach-Zehnder modulator to introduce the time-multiplexed signal of the input data. The reservoir laser features delayed self-feedback that generates a recurrence in the setup. The delay of the experimental setup at IFISC has a fixed length of 24.5ns, which places the system within the long delay limit. To determine optimal computing regions, this photonic setup can be modeled using the Lang-Kobayashi equations while attempting to match the time scales to the experimental characteristics. Accordingly, the delay length used in all numerical simulations is set to the one in the experimental setup.

In previous attempts, the time per input was fixed to the delay time and hence the processing speed was given via the inverse of the delay time. With the delay time given above this yields an information processing speed of around 40MHz. The goal of the current characterization has been to reduce the input time below the delay time while attempting to increase the performance of predictions on the Mackey-Glass delay system and Santa Fe time series. To that end, the temporal node separation of the virtual nodes is reduced to 12ps, which is the lowest separation that can be achieved with the experimentally used AWG. At the same time, the number of nodes is optimized to improve the prediction accuracy. Together with the fixed node separation, the number of nodes defines the input time and therefore the processing speed. Reducing the number of nodes, in turn, might render the input time shorter than the delay time.

Numerical results show that for nine steps ahead prediction of the Mackey-Glass task, the optimal number of virtual nodes is around 200. This leads to an input time of 2.4ns and corresponds to 10 inputs per delay time. For the Santa Fe task, the optimal number of inputs per delay lies around 2.5. For the Mackey-Glass task, the processing speed reaches around 0.4GHz while for the Santa Fe task it reaches 0.1 GHz. Furthermore, the prediction error in the Mackey-Glass time series prediction could be reduced by a factor of 1/5 compared to previous results of the same setup. For both tasks, to find the optimal performance, the strength of the feedback and the frequency detuning between the injection and reservoir laser were optimized. These numerical predictions have been verified in experimental measurements, with a detailed characterization being carried out by Irene Estebanez (IFISC).

2.1.2.2 Opto-electronic reservoir computing based on a delay loop

The opto-electronic setup proposed in Deliverable 2.1 (section 2.1.2.2) was improved and fully characterized by testing the performance for a Human Action Recognition (HAR) Task [1]. Figure 2 shows the experimental system, which is similar to the one shown previously, except for the analog readout that is discarded.

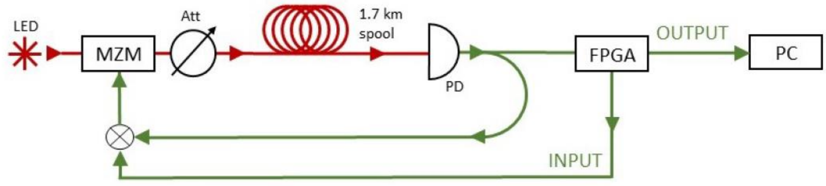

Figure 2 Scheme of the experimental setup. The optical and electronic components are shown in red and green, respectively. The reservoir layer consists of an incoherent light source, a Mach-Zehnder intensity modulator (MZM), an optical attenuator (Att), an approximately 1.7-km fibre spool, a feedback photodiode (PD) and a resistive combiner.

The dataset used by ESR4 to characterize the opto-electronic set-up is the popular KTH dataset [2], which consists in recordings of 25 actors performing 6 different actions (walking, jogging, running, boxing, hand waving, and hand clapping). The output layer consists then of 6 output nodes corresponding to the 6 action classes. Each node is trained with the ridge regression algorithm as a binary classifier with a “winner-takes-all” approach: in this way the output action class for a given video sequence is predicted. Figure 3 compares the best experimental results obtained by ESR4 with the literature: with 200 nodes a classification accuracy of 90.0% is reported. This performance is comparable to the state-of-the-art rates 90.02% - 96.0% achieved with far more complex and demanding architectures implemented on noiseless digital processor and also comparable to other photonic RC implementations, which actually employ reservoirs of thousands of nodes. These results thus show that a challenging computer vision task can be efficiently solved with a simple photonic system. In the near future, ESR4 will work towards increasing the number of nodes, and check whether the performance increases even further.

Publication	Classifier	Network Size	Accuracy
Jhuang [3]	SVM	19200	96%
Shu *et al.* [4]	SNN	4800	95.3%
Jahagirdar *et al.* [5]	KNN	19200	91.83%
Antonik *et al.* [6]	Photonic RC	16384	91.3%
Ji *et al.* [7]	3DCCN	295458	90.02%
This work	**Photonic RC**	**200**	**90%**
Schuldt *et al.* [2]	SVM	19200	71.83%

Figure 3 Performance of previous state-of-the-art digital approaches compared to ESR4’s best experimental results on HAR task. SVM : Support Vector Machine ; SNN : Spiking NN; KNN : K-Nearest Neighbors ; 3DCCN : 3-D Cellular Neural Network.

2.1.3 Frequency multiplexing

ESR12 is working on a detailed characterization of frequency multiplexing schemes for optical information processing. Both Reservoir Computers and Extreme Learning Machines have been implemented in fiber-optics systems where information is encoded in the lines of a frequency comb [https://doi.org/10.1364/OL.451087, https://doi.org/10.1364/OE.433535]. The schematic of the RC experiment is presented in Fig. 4.

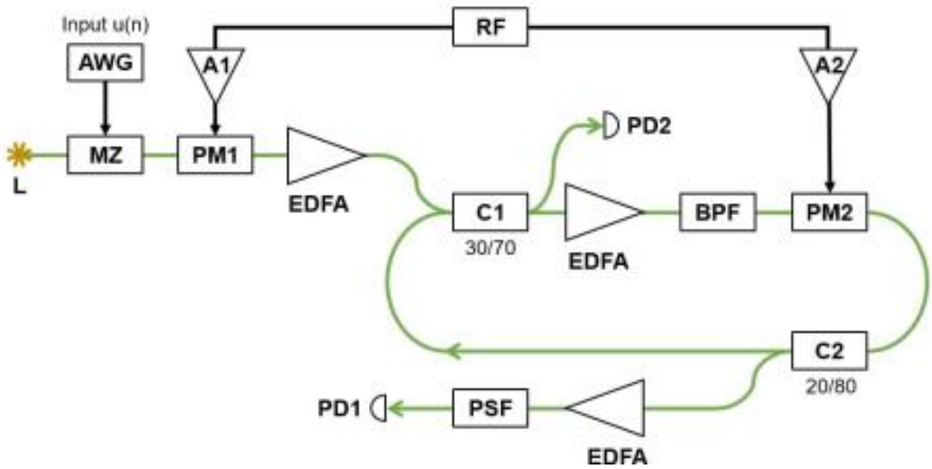

Figure 4 Schematic of the Reservoir Computing experiment. Black lines: electrical connections. Green lines: polarization maintaining optical fiber. L: laser. AWG: Arbitrary Waveform Generator. MZ: Mach-Zehnder modulator. PM1 and PM2: Phase Modulators. EDFA: Erbium Doped Fiber Amplifier. A1 and A2: RF Amplifiers. C1 and C2: Couplers. BPF: Band Pass Filter. PSF: Programmable Spectral Filter. PD1: Readout Photodiode. PD2: Ancillary Photodiode used for the stabilization of the loop.

The possibility of executing multiple parallel operations on the Extreme Learning Machine system, exploiting multiple non-interfering frequency combs, has also been demonstrated [https://doi.org/10.3390/app12010214]. A similar parallelization scheme is under study for the Reservoir Computing system. In the latter, the parallel computations may be completely independent, or could be used to implement a Deep Reservoir Computing scheme, where information gets processed by multiple reservoirs, thus increasing the information processing capabilities. The main identified limits in the Reservoir Computing scheme are related to the presence of a 15-meter-long fiber loop, which constitutes the recurrence path of the system. A long loop implies slow data elaboration rate (in particular, the rate should match the round-trip time of the light propagating in the cavity, thus 15 m are equivalent to approximately 20 MHz of processing rate). Moreover, a long cavity makes the system strongly sensitive to acoustic noise and thermal drifts, thus making necessary the presence of multiple stabilization mechanism. Current experimental limitations could be overcome by a photonic integrated chip, which can contain a shorter and more stable loop. Experimental tests on an integrated loop (approximately 40 mm long) have been started. In collaboration with ESR11, two more photonic chips have been designed, which will be used to realize an optical output layer mainly based on SOA wavelength converters. Chips are expected to be delivered by the end of 2022.

2.2 Electronic Designs

ESR1 and ESR2 spent a secondment at IniLabs, during which they worked with the electronic mixed-signal spiking neuromorphic processor DynapSE2. ESR1 developed an application programming interface (API) for Python to communicate with the chip through an FPGA. This API is built on top of the samna C++ library that the chip manufacturer (SynSense) provides. Using the Python API, it is possible to abstractly specify the desired spiking neural network architecture. The space of possible architectures is large compared to conventional deep learning neural networks.

The analog parameters and biases determine the analog behavior of the neural network, i.e. the gains, currents, refractory periods, thresholds of the neurons as well as the time constants, delays, weights and pulse extensions of the synapses. The neuron latches configure the neuron type (adaptive exponential integrate-and-fire or thresholded integrate-and-fire), whether or not the homeostasis and/or adaptation circuit is used, and can also be configured so as to merge 4 neurons together in order to use the neuron circuits as synapses in order to increase the possible fan-in of the network. The synapse configurations determine which of four dendrite models is used (AMPA, NMDA, GABA, or shunting), whether or not short-term potentiation is used, and others. The content-addressable memory (CAM) stores the synaptic connections, i.e. which pre-neuron each post-neuron listens to. The SRAM memory contains information about which chips and cores the firing events of a neuron are sent to.

The API designed by ESR1 uploads a high-level network specification to the chip and attempts to satisfy all device constraints such as maximum fan-in, and constraints about core-wide parameters (for example, the neuron type can only be set for an entire core of 256 neurons). If such a configuration is not possible, the API will return a warning before uploading the configuration to the chip. This was then used to speed up the exploration of computational primitives on the chip. A generic winner-take-all (WTA) circuit was implemented that can be tuned to operate as linear analog gain, non-linear gain control, non-linear selection, or signal restoration. The WTA circuit was extended to a neural state machine (NSM), though the implementation of the NSM remained unstable. A neural oscillator was implemented using two recurrently connected excitatory and inhibitory populations. Finally, reservoirs of spiking neurons as well as feed-forward layers of spiking neurons were also implemented.

Unfortunately, ESR1 and ESR2 no longer have access to the DynapSE2 chip from IniLabs and are still waiting for the delivery of the DynapSE2 chip from the next fabrication batch. Due to some packaging problems, the delivery date of this chip is not known at the moment. After some months of simulations studies, ESR1 is taking an initiative to obtain access to Intel's Loihi 2 cloud system in order to move simulations to this device setup. Loihi 2 offers a higher degree of programmability and although it is fully digital, it offers the benefits of asynchronicity and massive parallelism as well as increased reliability due to the higher production volume and digital design.

2.3 References

[1] E. Picco, P. Antonik and S. Massar, "Time-Multiplexed Photonic Reservoir Computer for Recognition of Filmed Human Actions", In: SFO-La Société Française d'Optique, Optique Nice 2022, July 2022, Session JNOG.

[2] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions : a local SVM approach, " in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004, pp. 32–36, Vol. 3.

[3] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.

[4] N. Shu, Q. Tang, and H. Liu, "A bio-inspired approach modelling spiking neural networks of visual cortex for human action recognition," 2014 International Joint Conference on Neural Networks (IJCNN), 2014.

[5] A. S. Jahagirdar and M. S. Nagmode, "Silhouette-Based Human Action Recognition by Embedding HOG and PCA Features," in Intelligent Computing and Information and Communication, 2018, pp. 363–371.

[6] P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Human action recognition with a large-scale brain inspired photonic computer," Nature Machine Intelligence, vol. 1, pp. 530–537, 2019.

[7] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221–231, 2013.