Reservoir computing in lithium niobate on insulator platforms

Mohab Sameh Mohamed Abdalla

Doctorat

Opérée au sein de :

École centrale de Lyon

Discipline : Électronique, Micro et Nanoélectronique, Optique et Laser

Soutenue publiquement le XX/12/2024, par :

Mohab Sameh Abdalla Mohamed

Reservoir Computing in Lithium Niobate on Insulator Platforms

---

Damien Querlioz Directeur de Recherche CNRS, C2N – Université Paris Saclay	**Rapporteur**
Guy van der Sande Professeur d'université, Vrije Universiteit Brussel	**Rapporteur**
Lorena Anghel Professeur d'université, Grenoble INP – SPINTEC	**Présidente**
Ioana Vatajelu Chercheur CNRS, Laboratoire TIMA – Grenoble INP	**Examinatrice**
Fabio Pavanello Chargé de Recherche CNRS, CROMA – Grenoble INP	**Co-encadrant de thèse**
Andreas Boes Senior Lecturer, University of Adelaide, Australia	**Co-encadrant de thèse**
Arnan Mitchell Professor, RMIT University, Australia	**Directeur de thèse**
Ian O'Connor Professeur d'université, École Centrale de Lyon – INL	**Directeur de thèse**
Christelle Monat Professeur d'université, École Centrale de Lyon – INL	**Membre invitée**

Abstract

This work concerns time-delay reservoir computing (TDRC) in integrated photonic platforms, specifically the Lithium Niobate on Insulator (LNOI) platform. We propose a novel all-optical integrated architecture, which has only one tunable parameter in the form of a phase-shifter, and which can achieve good performance on several reservoir computing benchmark tasks. We also investigate the design space of this architecture and the asynchronous operation, which represents a departure from the more common framework of envisioning time-delay reservoir computers as networks in the stricter sense. Additionally, we propose to leverage the optical feedback to dispense with the input mask, which allows the bypassing of an O/E/O conversion, which is often necessary to apply the mask. In future work, this can allow the processing of real-time incoming signals, possibly for telecom/edge applications. The effects of the output electronic readout on the proposed architecture's performance are also investigated. Initial experimental work is also reported. The unifying theme of this work is to investigate the performance possibilities with minimum photonic hardware requirements, relying mainly on LNOI's low losses which enables the integration of the feedback waveguide, and using only interference and subsequent intensity conversion (through a photodetector) as the nonlinearity. This provides a base for future work to compare against in terms of performance gains when additional nonlinearities are considered (such as those available on the LNOI platform), and when overall system complexity is increased by means of introducing more tunable parameters.
Thus, the scope of this work is about the exploration of one particular unconventional computing approach (reservoir computing), using one particular technology (photonics), on one particular platform (lithium niobate on insulator). This work builds on the increasing interest of exploring unconventional computing, since it has been shown over the years that digital computers can no longer be a 'one-size-fits-all', especially for emerging applications like artificial intelligence (AI). The future landscape of computing will likely encompass a rich variety of computing paradigms, architectures, and hardware, to meet the needs of rising specialized applications, and all in coexistence with digital computers which remain — at least for now — better suited for general-purpose computing.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Context

The invention of the transistor in 1947 began the unprecedented development of computers, thrusting the world into what is now called the “Information Age". Prior to that, first-generation computers utilized vacuum tubes to carry out logical processing. These vacuum-tube computers were first used for solving linear systems of equations, then extended to general purpose computing (first in the ENIAC, 1945, then later in the binary EDVAC). The concepts behind digital computing were already developed in the Babbage engine as far back as the 1820s. Going further back, analog computers have existed for far longer in the form of devices to measure real-world physical variables; chronometer for time, slide rule for distance, among others. Even in ancient times, humans had the predecessors to these tools in sundials and measuring rods based on arm-spans. Today, computers play a vital role in our daily lives, both at the individual and societal level, and the modern world would not function as it does without them.

1.1.1 A brief history of computing

The rise of computing in the 20th century — from enabling materials to the scale and breadth of applications — was not simply a technological breakthrough. Mathematical progress in the 19th century had already transformed philosophical logic into a formal language which gave birth to symbolic logic, and subsequently Boolean algebra, which is the foundation of digital computing. In symbolic logic, an expression is evaluated for its truth value, yielding one of two possible outputs: True or False, which can in practice be represented as ’ON’ and ’OFF’ states of a switch. Pioneering work in computability theory in the 1930s, catalyzed by Gödel’s then-

contemporary work in formal logic, led to the Turing machine: a conceptual computer that manipulates symbols on an infinite length of tape according to a table of rules. This simple machine could implement, in principle, any computer algorithm. Boolean algebra was also heavily utilized by Claude Shannon to develop Information Theory, laying the groundwork for modern communication theory. On the more practical side, advancements in semiconductor physics and material science enabled the engineering of the p-n junction. Thus, modern computing could only emerge thanks to the successive breakthroughs spanning the theoretical and practical sciences in the first half of the 20th century.
Vacuum tube computers were being used well beyond the first half of the 20th century. However they were difficult to maintain due to constant failures (unreliable), and were also operated at hundreds of volts (power hungry). The invention of the transistor promised (and later delivered) rapid, energy-efficient, and reliable switching. This allowed for the scaling up of computing in terms of number of switches (through scaling down of device size and on-chip integration) and also the scope of applications, making them the ubiquitous machines we have today. The incredible success of digital electronics, driven by Moore's law [1], which predicted that the number of transistors on an integrated circuit roughly doubles every two years, sustained the improvement of computing performance for more than 50 years. Continuous progress in device architectures has allowed further scaling down to 2-nm process (including Fin-Fet, GAAFET, etc.. [2]). However, this came at the expense of rising R&D and fab costs [3, 4]. Furthermore, today's computers — which essentially use the architecture known as the "Princeton" architecture or more commonly as the "von Neumann" model — suffer from a well-known bottleneck.

1.1.2 The von Neumann Bottleneck

In the first draft of the EDVAC report [5], John von Neumann described a design for a digital electronic computer. The proposed architecture boils down to the schematic shown in Fig. 1.1, comprising of:

- A central processing unit that includes an instruction register and a program counter
- Shared memory for data and instructions
- Input and output peripherals
- External mass storage for long-term data storage

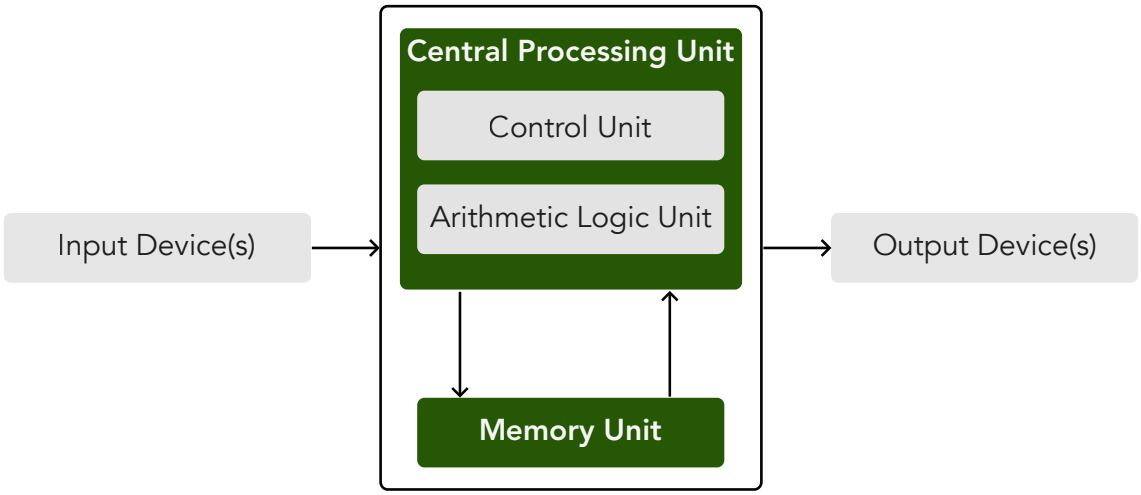

Fig. 1.1: Simplified diagram of the von Neumann architecture

In conventional (Harvard) architectures, data and instructions share the same bus and memory. In electronic implementations, one must resort to a serial method to transfer instructions and data one at a time. A certain time t' is consumed to fetch data/instructions at each clock cycle. Consider a simple program of adding two numbers x and y:

1. Fetch 'add' instruction
2. Fetch x
3. Fetch y
4. Execute 'add' instruction
5. Store sum(x, y)

With such a simple program suffering already from an excess run time (overhead) of $4t'$ (everything except the execution), it can be seen how inefficiently this scales with the size and complexity of the program. Modern architectures, which are still largely von Neumann based [6], mitigate this problem by providing, for example, multi-level memory (L1, L2 cache). This can allow for some of the data and instructions to be 'closer' to the processor, thereby reducing t'. On the other hand, pipelining strategies can also help, which is done in Microprocessor without Interlocked Pipeline Stages (MIPS) processors (in itself a Reduced Instruction Set Computer (RISC) architecture). Pipelining allows for several phases of instructions to execute simultaneously, thus reducing CPU idle time. On the other hand, multi-core processing enables parallelization and more efficient computing, especially if programs are optimized to run on such hardware. However, all these mitigation techniques have

become less effective over the years with the advances in CPU speeds while memory access speeds plateaued [7]. Furthermore, the rise of AI and the scale at which it is deployed nowadays means that application-specific hardware would be more suitable, especially one that can natively accomplish, e.g., multiply-accumulate (MAC) operations ($sum = sum + (a \times b)$), which are widely used in digital signal processing, image processing, and AI applications (especially neural networks). Recurrent neural networks (RNNs) are a widely-used type of neural networks, optimized for temporal processing such as speech recognition. In RNNs, the number of MACs required is roughly

$$\text{MACs} = T \times \left[(x \times h) + h^2 + hy\right] \tag{1.1}$$

where T refers to the number of time steps considered, h is the hidden dimension, x is the input size, and y is size of the output. For more advanced RNNs, like the widely used Long Short-Term Memory (LSTM), the number of MACs increases further by a factor of four. In contrast to the example program above, which sums two numbers, this involves additional overheads to retrieve a and b, the multiply instruction, and the previous sum as well as storing the new sum. Graphical processing units (GPUs), which were originally made for the purpose of 3D graphic rendering, can handle MACs more efficiently than CPUs thanks to their highly parallel architecture and matrix processing units. In addition to designing specialized hardware that can outperform CPUs in MAC operations, it is also possible to implement MAC operations implicitly in some paradigms of computing, where the information processing is done in a more unconventional way, such as exploiting the properties and dynamics of physical systems. Thus, there has been increasing interest in the exploration of unconventional computing paradigms, ranging from in-memory computing [8], stochastic computing [9], neuromorphic computing [10], and reservoir computing [11], among others. Optical computing paradigms have also been explored as a means to leverage particular advantages of working with light over RF signals, or simply as a basis for the implementation of some of the above-mentioned paradigms.

1.1.3 The Fall and Rise of Optical Computing

Optical computing refers to the use of optical devices and systems to perform signal processing. 'Information optics' — one of the earlier names of this field — could arguably be traced back to 1859 with the knife-edge test by Foucault [12], where light propagated through a lens or curved mirror and was intercepted by a knife-edge, such that the resulting observed image would give information about the mirror/lens aberrations and surface quality. The first initiatives for optical computing (in the

modern sense) emerged in the 1950s, with the first all-optical schemes proposed as early as the 1950s, coinciding with the development of holography [13]. The linking factor here is that both, proposed optical processors and holography, relied on a Fourier optics treatment. The first such system, known as the 4-f correlator [14], is composed of 3 planes: input, reference (processing), and output. Inputs were initially realized as static slides which were later substituted with spatial light modulators (SLM), which are optical devices that could be tuned by applying electrical signals. A lens is positioned between each pair of planes, hence the '4-f' designation implying 4 focal distances separating the input and output layers. The processing plane can be composed of lenses, holograms, or even nonlinear elements, depending on the desired computation. The output layer comprises a photodetector array or a camera. While the processing itself takes place at the speed of light, the whole system is limited by the rather slow SLMs (modern units operate at the ∼100 Hz order of magnitude). Thus the applications for real-time signal processing remained out of reach.

On the other hand, the rise of communication theory and information theory paved the way for their adaptation to optical systems [15]. There was great optimism for this then-emerging field, which continued onward until the 1990s. From a retrospective point of view, it is interesting to note that some of that optimism had been in a surprisingly realistic direction. In 1963, a symposium was held to discuss the role and future of optical information processing with about 425 participants from the fields of optics and information theory. In the published proceedings [16], the authors clearly show a preference for the use of optics within application-specific contexts such as pattern recognition and character recognition, where the parallelization offered by optics can be harnessed. In the same preface, the authors also expressed skepticism over "the possibility of developing a general-purpose optical computer", but which could be overcome "with the discovery and application of new optical effects and phenomena in laser research and fiber optics". It is interesting to note that, in retrospect, the scope of applications where optical computing could really make a difference was accurately predicted, i.e. within application-specific settings. International conferences on the theme of optical processing were held almost on a yearly basis from the 1980s to the 1990s by various organizations, the journal Applied Optics held a feature issue on a monthly basis, and the momentum was strong. Owing to the nature of light, most of the efforts were naturally directed towards analog computing paradigms.

The subsequent rise and rapid progress of electronic digital computers, backed by the self-fulfilling prophecy of Moore's law, dealt a heavy blow to the optical comput-

ing community. This was especially aggravated when optical digital computing was also considered, where it became quickly apparent that it could not compete against the rapidly developing digital electronics industry. This was primarily due to the limitations of cascadability, density of integration, non-volatile information storage, and difficulty of achieving nonlinear elements in optical platforms as opposed to electronic ones. In the early 2000s, there were no longer specialized conferences on the topic of "Optical Processing", and the Applied Optics monthly feature was dropped [17].

Far from being the end for optical computing, and in an ironic turn of events, it is the slowing down of progress in digital electronics that paved the way back for optical computing. In fact, this comeback was largely possible thanks to two major — though unrelated — strides: the first being the rise of silicon photonics (focused initially on interconnects) in the mid 2000s, overcoming the limitations of electrical I/O through replacement by on-chip optical communication, and the second being the rapid developments in AI and in particular neural networks in the late 2000s and early 2010s. Thus, the rise of silicon photonics, coinciding with the need to address the increasing demands of AI, has been key in the resurrection of the previously almost-dismissed forays into optical computing. In addition to the possibility of using optical interconnects with conventional computing platforms for AI, optical hardware is being harnessed as a vehicle to implement various emerging computing paradigms, such as in-memory computing [18], Ising machines [19], physical neural networks [20], and reservoir computing [21].

So where are the opportunities in optical computing? By now it is obvious that light-speed processing is in fact not the main catch, the reason being the unavoidable bottlenecks from hardware required for domain conversions (i.e. between optical to electronic and vice-versa), in addition to the simple fact that radio frequency (RF) electrical signals do in fact travel at the speed of light as well. A recent 2023 perspective article lists some features where photonics can make a difference in computing [22], most notably:

- **Bandwidth**: photonics operate on lightwaves (100s of THz), while digital electronics operate on RF waves (GHz). The difference in potential bandwidth is 5 orders of magnitude. This means massive parallelism can be achieved through frequency-multiplexing, where frequency combs can generate $\sim 10^7$ lines [23]. Data in each line can be processed in parallel.

- **Near-dissipationless dynamics and low-loss transmission**: photons can propagate with negligible losses in free-space setups and very low losses in some integrated platforms (e.g. losses demonstrated on lithium niobate on

insulator and silicon-nitride platforms ~ 0.2 dB cm^{-1} in the C-band [24, 25]). It is anticipated that optical interconnects will also play a role in electronic intra-chip communication.

Thus, there is great promise in utilizing photonic platforms for computing applications, especially those that would benefit from fundamentally different paradigms that promise boosts in speed and energy efficiency, e.g., application-specific signal processing, and which can exploit and benefit from the above-mentioned strengths of working with light.

1.2 Structure of the Dissertation

The rest of this work is structured as follows. In chapter 3, recent advances in photonic neural networks, and in particular, photonic RC, are summarized. This is then followed by the research questions addressed in this dissertation and a statement of the research contributions. Chapter 4 introduces the proposed minimum complexity architecture alongside numerical results on some standard benchmark tasks. Chapter 5 proposes the dispensing of the input masking protocol through the use of the same architecture presented albeit in the asynchronous regime. Chapter 6 shows the experimental work done on fabricated devices on the lithium-niobate-on-insulator platform. Finally, chapter 7 presents the summary and provides suggestions and perspectives for future work.

Bibliography

[1] G. E. Moore, "Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff." *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 3, pp. 33–35, 2006.

[2] S. BintiMdSallah, H. Mohamed, M. Mamun, and M. Amin, "Cmos downsizing: Present, past and future," *Journal of Applied Sciences Research*, vol. 8, pp. 4138–4146, 01 2012.

[3] "Navigating the costly economics of chip making," Apr 2023. [Online]. Available: https://www.bcg.com/publications/2023/navigating-the-semiconductor-manufacturing-costs

[4] "It is time to take intel seriously as a chip foundry," Apr 2024. [Online]. Available: https://www.forbes.com/sites/tiriasresearch/2024/04/05/it-is-time-to-take-intel-seriously-as-a-chip-foundry/

[5] J. Neumann, "First draft of a report on the EDVAC," *archived from the original (PDF) on March 14*, vol. 25, 1945.

[6] D. A. Patterson and J. L. Hennessy, *Computer Organization and Design: The hardware/software interface revised 4th edition*. Elsevier Science, 2011.

[7] C. Carvalho, "The gap between processor and memory speeds," 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:16553457

[8] S. Bavikadi, P. R. Sutradhar, K. N. Khasawneh, A. Ganguly, and S. M. Pudukotai Dinakarrao, "A review of in-memory computing architectures for machine learning applications," in *Proceedings of the 2020 on Great Lakes Symposium on VLSI*, ser. GLSVLSI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 89–94. [Online]. Available: https://doi.org/10.1145/3386263.3407649

[9] W. J. Gross and V. C. Gaudet, Eds., *Stochastic Computing: Techniques and Applications*. Springer International Publishing, 2019. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-03730-7

[10] A. Ben Abdallah and K. N. Dang, *Neuromorphic Computing Principles and Organization*. Springer International Publishing, 2022. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-92525-3

[11] K. Nakajima and I. Fischer, Eds., *Reservoir Computing*. Springer Singapore, 2021. [Online]. Available: http://dx.doi.org/10.1007/978-981-13-1687-6

[12] L. Foucault, "Memoire sur la construction des telescopes en verre argente," *Annales de l'Observatoire Imperial de Paris*, vol. 5, p. 197–237, 1859.

[13] D. GABOR, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, May 1948. [Online]. Available: https://doi.org/10.1038/161777a0

[14] A. Marechal and P. Croce, "Un filtre de frequences spatiales pour l'amelioration du contraste des images optiques, Comptes Rendus de l'academie des Sciences," vol. 237, no. 12, pp. 607–609, 1953.

[15] P. Elias, "Optics and communication theory," *J. Opt. Soc. Am.*, vol. 43, no. 4, pp. 229–232, Apr 1953. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-43-4-229

[16] D. K. Pollock, C. J. Koester, and J. T. Tippett, *Optical Processing of Information, Spartan Books, Baltimore, Md, USA*, 1963.

[17] P. Ambs, "Optical computing: A 60-year adventure," *Advances in Optical Technologies*, 2010.

[18] C. Ríos, N. Youngblood, Z. Cheng, M. L. Gallo, W. H. P. Pernice, C. D. Wright, A. Sebastian, and H. Bhaskaran, "In-memory computing on a photonic platform," *Science Advances*, vol. 5, no. 2, p. eaau5759, 2019. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.aau5759

[19] F. Böhm, G. Verschaffelt, and G. Van der Sande, "A poor man's coherent ising machine based on opto-electronic feedback systems for solving optimization problems," *Nature Communications*, vol. 10, no. 1, p. 3538, Aug 2019. [Online]. Available: https://doi.org/10.1038/s41467-019-11484-3

[20] K. Liao, T. Dai, Q. Yan, X. Hu, and Q. Gong, "Integrated photonic neural networks: Opportunities and challenges," *ACS Photonics*, vol. 10, no. 7, pp. 2001–2010, Jul 2023. [Online]. Available: https://doi.org/10.1021/acsphotonics.2c01516

[21] G. V. der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017. [Online]. Available: https://doi.org/10.1515/nanoph-2016-0132

[22] P. L. McMahon, "The physics of optical computing," *Nature Reviews Physics*, vol. 5, pp. 717–734, 12 2023.

[23] S. A. Diddams, K. Vahala, and T. Udem, "Optical frequency combs: Coherently uniting the electromagnetic spectrum," *Science*, vol. 369, no. 6501, p. eaay3676, 2020. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aay3676

[24] A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (lnoi) for photonic integrated circuits," *Laser and Photonics Reviews*, vol. 12, 4 2018.

[25] N. F. Tyndall, D. A. Kozak, M. W. Pruessner, P. G. Goetz, W. S. Rabinovich, T. H. Stievater, N. M. Fahrenkopf, and A. Antohe, "A low-loss, broadband, nitride-only photonic integrated circuit platform," in *Quantum 2.0 Conference and Exhibition*. Optica Publishing Group, 2022, p. QTu4B.5. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=QUANTUM-2022-QTu4B.5

Chapter 2

Background

2.1 Background on Photonic Integrated Circuits

The history of integrated optics is almost as old as integrated electronics. A seminal Bell-Labs 1969 paper discussed the benefits of photonic integration [1]. Just a year earlier, the first complementary metal-oxide semiconductor (CMOS) integrated circuits were made [2]. This section presents a brief overview of photonics, with a focus on photonic integrated circuits (PICs). Firstly, some material platforms are discussed, with their advantages and drawbacks. This is followed by a tutorial on photonic building blocks and their compact models, which aims to familiarize the reader with the simulation part of this work.

2.1.1 Material platforms

Much research has gone into finding suitable materials for integrated photonic devices and circuits. For most platforms, photolithography is used to pattern wafers for etching and material deposition. Unlike with electronic circuits, where electrons exist readily inside materials, photonic circuits require the means of light *generation*, usually achieved by generating electron-hole pairs within a gain medium through an external excitation process, and *detection*, through the combination of those electron-hole pairs. From a historical perspective, device-level design preceded circuit-level thinking for decades. This resulted in the specialization of platforms for these different devices, for example: indium phosphide (InP) for lasers, gallium arsenide (GaAs) for detectors, and lithium niobate ($LiNbO_3$) for high-speed modulators. Each of these devices were made under different fabrication processes which are fundamentally incompatible with the ones used for the other devices. If one could integrate all these functionalities on silicon, with a standard fabrication pro-

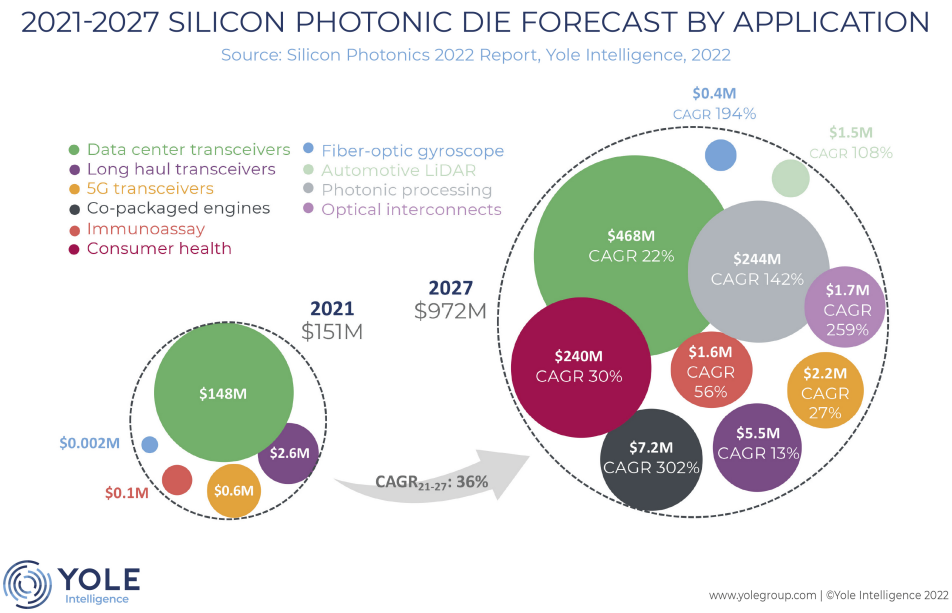

Fig. 2.1: Projection of silicon photonics compound annual growth rate (CAGR) and comprising industries [6].

cess, then photonics would become the next fabless semiconductor industry [3]. As of now, it remains an open engineering problem to achieve efficient, integrate-able light emission on silicon platforms (due to silicon having an indirect bandgap), with some recent demonstrations of such phenomena using epitaxially grown hexagonal silicon nanowires (with direct bandgap) [4]. Earlier efforts to achieve lasing in silicon used Raman nonlinearities in the indirect bandgap regime [5], but this scheme requires optical pumping which necessitates another laser. Thus, it is of interest to explore other material platforms and the possibility of integrating multiple ones. CMOS-compatible monolithic integration used to be attractive for the integrated photonics community. The desire was to leverage the 50+ year-old multi-billion dollar electronics industry for the currently booming photonics industry, and use existing foundries and mature fabrication facilities as a quick means to achieving large scale production of photonic chips. Additionally, many of the older facilities that were once required for larger gate widths could be repurposed for photonic chips. Going further in this direction, one could consider fabrication with only full CMOS compatibility in mind, i.e. using the existing fabs without introducing fabrication steps that may contaminate them or affect their use for traditional electronics [7, 8]. However, a fundamental limitation of this direction, observed through the years, is the very low yield of such a scheme. The first large-scale use of CMOS-compatible silicon photonics in an industrial setting was with optical transceivers, pioneered by Luxtera in 2006 on a then state of the art 130 nm SOI platform [9]. Optical

transceivers solved a fundamental issue in data centers which have large amounts of modules and racks that require interconnectivity, initially done with coaxial cables. The main problem with electrical cables is their impedance-induced heat generation, which required extensive cooling equipment and associated energy costs, in addition to bandwidth limitations. Integrated photonics solved this problem by introducing a module that is both a transmitter and receiver, combining optical modulation circuitry with photodetectors, and replacing electrical cables with optical fibers where light can travel without generating heat. Furthermore, wavelength-division multiplexing is leveraged to send multiple signals onto the same optical fiber, greatly increasing the bandwidth. While initially InP was the dominating integrated photonic platform, silicon photonics (SiPh) has been gaining traction over the years and is expected to contribute to larger portions of this industry over the years to come [6], in addition to other application areas, most notably photonic processing [10, 11], which is projected to play a major role in the industry, as shown in Fig. 2.1.

Another appeal of SiPh is the capability of integrating electronics. Active optical devices, such as lasers and semiconductor optical amplifiers (SOAs), along with tunable photonics (e.g. filters) require electronic control circuits. This opened the door to integrated photonics beyond research labs, and many kinds of applications, from sensing to information processing, have been envisioned and implemented. This also gave rise to the desire of integrating multiple materials on the SiPh platform. In other words, if the strengths of all materials can be leveraged and combined together, this would enable competitive products in a wide array of applications. Hybrid and heterogeneous integration schemes pave this way [12], where the former is done at the packaging stage, and the latter is done at the fabrication stage.

In recent years, silicon nitride (Si_3N_4) and lithium-niobate-on-insulator (LNOI), both CMOS-compatible platforms, gained interest owing to their very low losses and their nonlinearities. LNOI is a relatively new platform where a layer of Si_3N_4 is loaded onto thin-film-$LiNbO_3$ (TFLN), which is deposited on a buried oxide (BOX) layer, to allow for CMOS compatibility, since TFLN in itself is not CMOS-compatible owing to specialized etching and deposition techniques [13]. LNOI enables the accessing of the wide range of advantages of $LiNbO_3$, which has been dubbed as the 'silicon of photonics' [14]. For example, $LiNbO_3$ possesses — among other nonlinearities — the linear electro-optic 'Pockel's' effect, a wide transparency window spanning the ultraviolet to the mid-infrared (400 nm — 5 µm), and benefits from widespread commercial availability of optical-grade wafers [15]. Among other applications such as sensing, these properties can be exploited for the purposes of high-speed photonic circuits, including photonic computing applications.

2.1.2 Photonic Building Blocks

The physics of light can be described by different theories, which follow the evolution of our understanding of the nature of light [16]. In chronological order, these theories are ray optics, scalar wave optics, beam optics, Fourier optics, electromagnetic optics, and quantum electrodynamics. In photonic circuits, depending on whether device level or system-level simulation is desired, and also on the kinds of phenomena taken into account, different theories of light may be employed for the modeling process. For integrated photonic system-level implementations, the modeling process is mainly reliant on guided-wave optics (e.g. coupled-mode theory).

Light can be manipulated by exploiting the phenomena associated with it, such as reflection, interference, polarization, dispersion, in addition to nonlinearities and further light-matter interaction phenomena. Bulk components can be used in optical table setups to control light in free space or in fiber optics. With the advances in photonic integration, many of these components can now be integrated together on a single chip using the above-mentioned CMOS-compatible processes. In PICs, components are either *passive* (draw no power), or *active* (draw some power). On chip, light can be confined in passive waveguides which are built into the chosen material/platform through e.g. an etching process. Active components include light sources (laser/LED), photodiodes, optical amplifiers, among others. A laser is a particular solution of the wave equation [17], which describes a coherent, monochromatic light field. The electric field of this source of frequency $\omega = 2\pi f$, assuming a plane wave (i.e no spatial dependence), can be described by its amplitude A and initial phase ϕ as

$$E_{\text{src}}(t) = A \exp(i\omega t + \phi) \tag{2.1}$$

For ease of notation, the dependence on time will be implicit in subsequent equations. Consider the following example of a photonic chip, shown in Fig. 2.2. Laser light can be injected into a photonic chip by means of electric field coupling between an optical fiber and a device/structure on the chip. In practice the two most common ways are (i) butt-coupling, where the cleaved end of the fiber is aligned as close as possible to the on-chip waveguide, and by (ii) using grating couplers (GC), which are Bragg structures that, when a certain angle condition is satisfied, can efficiently couple the light into the waveguide mode for a range of wavelengths. Butt-coupling, when properly done and for permanence, can be highly efficient. However, a limitation is that there is a high degree of alignment sensitivity. On the other hand, GCs offer

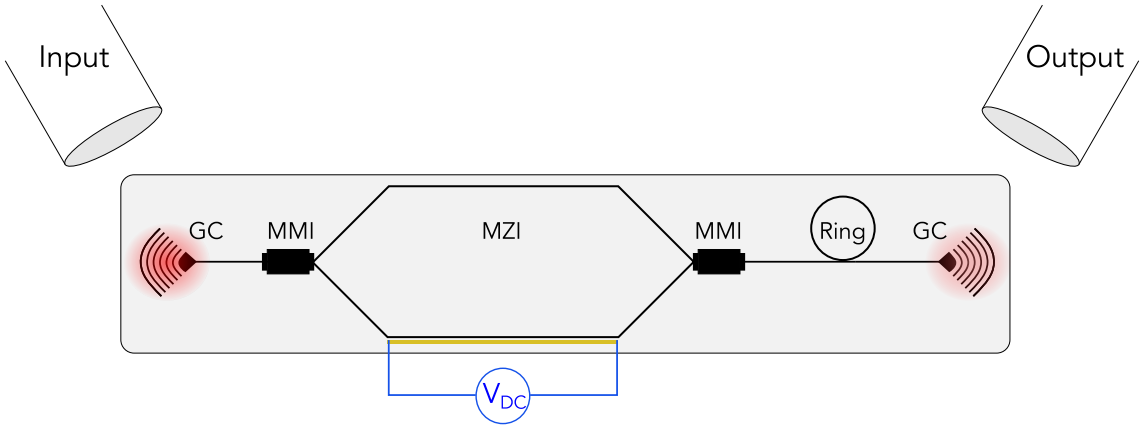

Fig. 2.2: Example of an integrated photonic chip with some common components.

a larger alignment flexibility for on-chip coupling, but they intrinsically suffer from more losses due to mode mismatch and energy dissipation into higher diffraction orders. After the light is coupled from the fiber into the chip via the GC, the field can be described as

$$E_{\mathrm{gc}} = \gamma_{\mathrm{gc}}(\lambda, \theta) E_{\mathrm{src}} \tag{2.2}$$

where γ_{gc} is the coupling coefficient of the GC which depends on the laser wavelength λ and the angle θ between the fiber and the normal. After being coupled into the on-chip waveguide and travelling some distance L_{wg}, the field evolves to:

$$E_{\mathrm{wg}} = E_{\mathrm{gc}} \exp\left(i\beta L_{\mathrm{wg}}\right) \exp\left(-\alpha_{\mathrm{wg}} L_{\mathrm{wg}}\right) \tag{2.3}$$

where L_{wg} is the length of the waveguide [m], α_{wg} is the waveguide loss [m^{-1}], and $\beta = 2\pi n_{\mathrm{eff}}/\lambda$ [m^{-1}] is the propagation constant of the guided mode with effective refractive index n_{eff}. The effective index describes the approximate index encountered by the propagating field due to the fact that not all the mode travels inside the waveguide core. This means that some of the mode will leak into the surrounding medium which in turn causes the field to encounter an 'effective' index that takes into account this effect. In practice, the effective index is usually found through mode solvers which apply coupled-mode theory to the boundary conditions defined by the waveguide dimensions and the properties of the materials involved. The details of this calculation are not discussed here. The field can then be split up using splitters, either in the form of Y-junctions, directional couplers, or multimode inteferometers (MMI). A 1×2 50-50 (3-dB) MMI, or the equivalent Y-splitter, will split the field such that one half of the associated power will go into each arm, yielding

$$E_1 = E_2 = \frac{\gamma_{\mathrm{mmi}}}{\sqrt{2}} E_{\mathrm{wg}} \tag{2.4}$$

where $\gamma_{\text{mmi}} \in (0, 1)$ is the MMI gain and depends largely on the inaccuracies encountered during fabrication processes. For a perfectly symmetric Y-splitter/1×2 MMI, there is no phase delay between the splitter arms. In general, MMIs are more tolerant to fabrication variations, as they rely on interference as the mechanism of power splitting. A 1×2 MMI and 2×1 MMI can be used together with two waveguides in between to construct an important device: the Mach-Zender interferometer (MZI), also shown in Fig. 2.2. An MZI exploits the interference property of light to generate a target output intensity from the interference of the two fields. This becomes especially useful when combined with a phase shifter, which is usually implemented by simply placing an electrical heater close to one of the waveguides. Thanks to the material's thermo-optic coefficient, the applied heat will alter the optical path length experienced by the field in that waveguide, thus controlling the MZI output intensity. This can be either slowly/statically controlled with a DC supply, or rapidly with a radio-frequency (RF) wave, as in the case of electro-optic modulators, which exploit a material's electro-optic coefficient for that purpose. For certain kinds of materials which have a non-centrosymmetric crystal structure, Pockel's effect is observed, whereby the change in refractive index is proportional to the applied electric field. The output field of the unbalanced MZI is expressed as

$$E_{\text{mzi}} = \frac{\gamma_{\text{mmi}}}{\sqrt{2}} E_1 \exp\left(-i\beta L_1\right) + \frac{\gamma_{\text{mmi}}}{\sqrt{2}} E_2 \exp\left(-i\beta L_2 - i\pi\left(\frac{V_{\text{m}}}{V_\pi^{\text{RF}}} + \frac{V_{\text{b}}}{V_\pi^{\text{DC}}}\right)\right) \tag{2.5}$$

where L_1 and L_2 are the lengths of the top and bottom MZI arms, respectively, V_{m} is the modulation voltage, V_{b} is the MZI bias voltage, V_π^{DC} is the DC voltage that induces a π phase shift, and similarly V_π^{RF} for the RF voltage. The electric field now encounters a resonant structure: a microring resonator. The microring filters the signal according to its resonant frequency with the optical path length condition given by

$$2\pi n_{\text{eff}} r = m\lambda_0 \tag{2.6}$$

where r is the radius of the ring resonator, m is the mode number, and λ_0 is the resonance wavelength [18]. Ring resonators are typically used in either an all-pass configuration (as in this example) or an add-drop configuration, which includes an additional 'drop' waveguide above the ring. Ignoring the waveguide between the combiner MMI and the ring resonator, and assuming symmetric point coupling (i.e. bidirectionally equivalent), the output of an all-pass or 'notch' ring resonator in the time domain can be expressed as

$$E_{\text{ring}}(t) = \gamma_{\text{c}} r E_{\text{mzi}}(t) - i\gamma_{\text{c}}\kappa \exp\left(-i\beta L_{\text{rt}}\right) E_{\text{mzi}}(t - \tau) \tag{2.7}$$

where $\gamma_c \in (0, 1)$ is the coupler gain, r is the through coupling coefficient, κ is the cross coupling coefficient, L_{rt} is the round-trip length, and τ is the round-trip time. In this equation the time-dependence and delay are explicitly written, as the resonator is a feedback structure described by both the incoming field and the stored field.

The presented compact models are useful for designing photonic circuits. The photonic devices are considered as blocks that affect the complex field either linearly or nonlinearly, in terms of phase and amplitude. More rigorous modeling techniques, such as solving the Maxwell equations in 3D, e.g. the finite-difference time-domain (FDTD) algorithm, which are suitable for device-level simulations, would not be practical for system-level modeling and design (which often take into account multiple components).

2.2 Background on Reservoir Computing

Machine learning (ML) can be defined as the building of a statistical model that solves a practical problem, based on a gathered dataset that is either present in the real-world or artificially created. ML can be supervised, unsupervised, or semi-supervised. In supervised ML, the dataset contains a collection of L labeled examples $(x_i, y_i)_{i=1}^{L}$, where x_i is a feature vector of dimension j that describes the example in some way. The goal is then to create a model that trains on these examples such that it can make sufficiently accurate predictions on unseen data. Reservoir computing (RC) is one type of supervised machine learning approaches. The underlying mathematical framework for reservoir computing has been developed in two ways independently of each other (also motivated by different reasons): the “echo-state network" (ESN) [19] and the “liquid-state machine" (LSM) [20]. The former deals with discrete-time systems, while the latter describes a continuous-time perspective useful for spiking systems [21]. Historically, the motivation for ESN was to circumvent the problems of training large-scale artificial neural networks (ANNs) implemented in software, which suffered from e.g. the vanishing gradient problem. Currently, the interest has shifted more towards physical RC implementations which leverage the dynamics of physical systems to do the computing, and which have been explored in various domains. One of the earliest RC hardware implementations involved mechanical actuators that acted as input to a ‘reservoir’ of water, which was able to solve the XOR and speech recognition benchmark tasks [22]. RC has since been implemented in electronics and photonics, among other platforms [23].

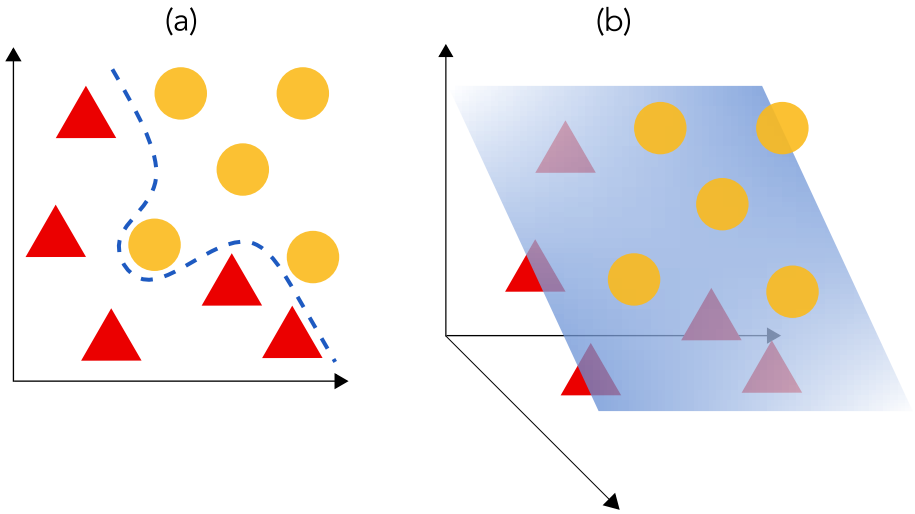

Fig. 2.3: Visualization of dimensionality expansion: (a) only nonlinear solutions can separate the different classes, while in (b) a plane can linearly separate them.

2.2.1 Formalism

We will concern ourselves with the ESN framework, as it is the one considered in the rest of this work. The ESN is a nonlinear dynamical system which exhibits particular characteristics, namely:

- the echo-state property (ESP); the system allows reverberations of past input that decay with time (fading memory),
- dimensionality expansion, usually achieved through a nonlinear activation function (Fig. 2.3).

The ESP is a condition that necessitates the forgetting of past inputs such that the reservoir does not depend on its initial conditions. This is an essential feature of RC as it would allow the reservoir to respond to the same input in the same manner. The consistent mapping of input data is crucial for ensuring that a model trained on the output of the reservoir can generalize effectively to different datasets. Dimensionality expansion enables the separation of multiple classes with simple linear classifiers, even for tasks that would normally require larger and deeper neural networks. The idea is, by projecting the data onto a higher dimensional space, there would exist more hyperplanes that can separate well the various classes of data, which enables high performance accuracy.

As shown in Fig. 2.4a, any RC scheme consists of the following 3 layers:

- Input layer: where data is preprocessed (optional) and then injected into the reservoir,

- Reservoir layer: which consists of nonlinear nodes connected to each other,
- Output layer: where the reservoir states are collected by means of a readout mechanism.

Each layer is connected to the next one via a series of weighted connections. What separates reservoir computing from similar temporal NNs, such as RNNs, is that the weights between the input and the reservoir layers can be set and fixed, as well as the interconnections between the nonlinear nodes, whereas in RNNs all the connections need to be optimized. Thus, in RC only the output layer is trained, which means that only least-squares methods are required to find the optimal weights. Furthermore, a reservoir may be constructed from an ensemble of physical nodes (Fig. 2.4b), or with one node in delayed feedback (Fig. 2.4c), known as time-delay reservoir computing (TDRC), which is discussed in section 2.2.2. The reservoir state matrix, which is a collection of the states of each of the N nodes, is described at each discrete timestep according to

$$x\left(n\right) = f\left[W^{\text{in}}u\left(n\right) + W^{\text{res}}x\left(n-1\right)\right] \tag{2.8}$$

where f is the nonlinear activation function, $u[n]$ is the current input, $W^{\text{in}} \in \mathbb{R}^{N\times M}$ is the weight connectivity matrix between the M (usually $M = 1$) inputs and each reservoir node, and $W^{\text{res}} \in \mathbb{R}^{N\times N}$ is the reservoir connectivity matrix describing how the nodes are connected to one another. The prediction step is then done at the output layer

$$y\left(n\right) = W^{\text{out}}x(n) \tag{2.9}$$

where $W^{\text{out}} \in \mathbb{R}^{S\times N}$ is the output weight matrix with S being the readout dimensionality (usually $S = M = 1$). After recording all the states, they are arranged together in a matrix which describes the N microstates of the reservoir for each training example L. W^{out} is then found by e.g. ridge regression:

$$W^{\text{out}} = \left[\mathbf{X}^{\top}\mathbf{X} + \lambda\mathbf{I}\right]^{-1}\mathbf{X}^{\top}\mathbf{y} \tag{2.10}$$

where $\mathbf{X} \in \mathbb{R}^{N\times L}$ is the matrix containing the reservoir states corresponding to each training example, $\mathbf{y} \in \mathbb{R}^{S\times L}$ is the labels matrix, $\mathbf{I}$ is the identity matrix, and λ is the Tikhonov regularization term which prevents overfitting by limiting the range of values available for the weights to take.

2.2.2 Time-delay Reservoir Computing

The emergence of TDRC came from the realization that networks of physical nodes can have their equivalents in the time-dimension of delayed-dynamical systems [24].

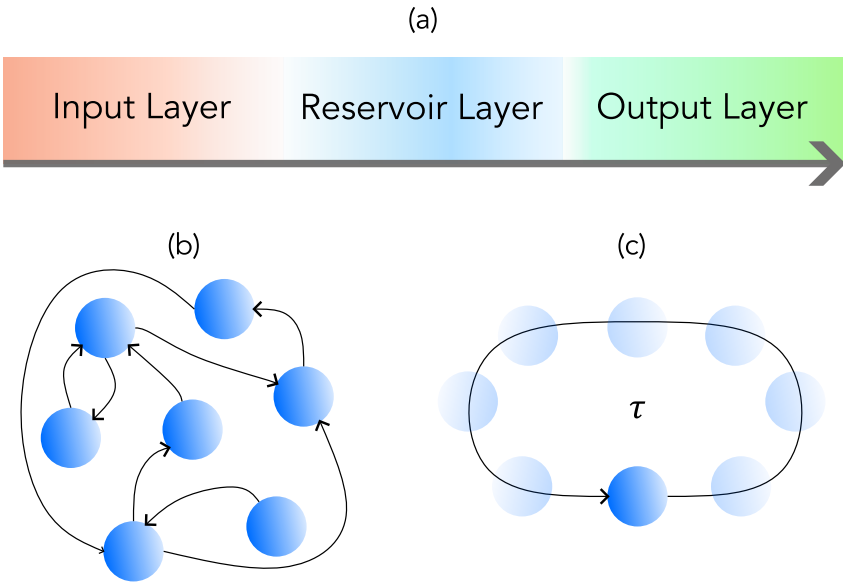

Fig. 2.4: (a) A reservoir computing scheme consists of 3 layers: input, reservoir, and output. The reservoir may be in the form of (b) physical nonlinear nodes connected to each other according to the reservoir connection matrix W^{res}, or as (c) one nonlinear node in a delayed feedback configuration.

Such systems are described by delay-differential equations (DDE) [25]. For example, the DDE of a system that considers a single constant delay has the form

$$\dot{X}(t) = F\left[t, X(t), X(t-\tau_{\text{d}})\right] \tag{2.11}$$

where τ_{d} is the delay time and is a positive constant. In other words, the system's rate of change is described by its present state and its state τ_{d} into the past. When stable, an input-driven system possesses some initial state $X(t_0) = X_0$, where F maps a subsequent trajectory until the system is brought back to its rest state after a certain time constant t_{s}, which depends on the system parameters. For RC, it is fundamental that F is nonlinear and that the system eventually goes to rest after being perturbed, which permits the forgetting of past inputs at t_{s} and thus removes dependence from the initial state.

Due to consideration of network equivalents, initial attempts at hardware TDRC were constrained by a quasi-synchronization of the input clock cycle T with the delay time, i.e. $T \approx \tau_{\text{d}}$. This yields a 'network' of equidistant virtual nodes N_{v} which are sampled at equal divisions of the clock cycle, i.e. $T = \theta N_{\text{v}}$, with θ being the inter-node temporal distance. In physical systems the network topology is defined by the interplay of the input datarate with the characteristic timescale of the system, which is dependent on τ_{d} and the timescale of the nonlinearity τ_{nl}. When $\tau_{nl} \ll \tau_{\text{d}}$, a perfect synchronization $\tau_{\text{d}} = T$ would couple each node to its past state

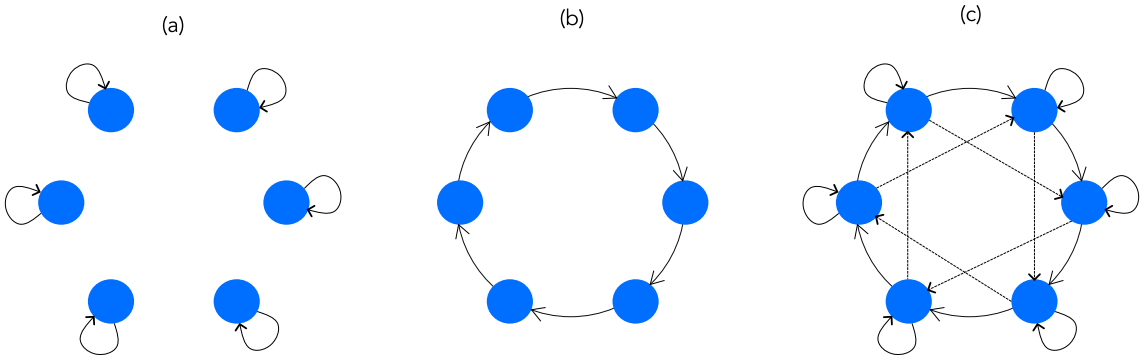

Fig. 2.5: Different network topologies in TDRC arising from the temporal dynamics, (a) $\tau_{\mathrm{nl}} \ll \tau_{\mathrm{d}}$ and $T = \tau_{\mathrm{d}}$, (b) $\tau_{\mathrm{nl}} \ll \tau_{\mathrm{d}}$ and $T = \tau_{\mathrm{d}} + \theta$, (c) $\tau_{\mathrm{nl}} \gg \tau_{\mathrm{d}}$

(Fig. 2.5a). Mismatching the clock cycle with the input delay by one node distance, i.e. $\tau_{\mathrm{d}} = T + \theta$, achieves the forward connectivity of the states of the reservoir, through what is usually described as 'coupling by inertia' (Fig. 2.5b). In the last case, $\tau_{\mathrm{nl}} \gg \theta$, which entails an even richer interconnectivity between the nodes thanks to the participation of the nonlinearity's timescale (Fig. 2.5c). However, this comes at the cost of slowing down the overall speed of the system. Interestingly, this forward connectivity with delay lines was explored theoretically and published in 2011 by Rodan and Tino [26], coinciding with the pioneering experimental demonstrations of TDRC that were published around the same time [24, 27]. Rodan's work was motivated by the construction of minimum complexity ESNs in software, where various time-multiplexed network topologies are investigated. Prior to that, ESNs were usually randomly initialized, but optimizing the internal connectivity to yield optimal dynamics for the task at hand was desirable. The optimization procedure proved difficult for interconnection weight matrices where most of the elements were nonzero, especially when this optimization was usually done through a parameter space exploration. This was addressed in [26] through the use of 'simple cycle reservoirs' where most of the internal weights were zeroed out. In this respect, physical TDRC based on ultrashort timescales is in fact the hardware realization of simple cycle reservoirs. One key feature of TDRC is the applying of an input mask onto the input signal in the pre-processing stage, as shown in Fig. 2.6. The mask repeats every clock cycle, and is usually done using either binary values, or using pseudo-random values drawn from some distribution. This ensures that the reservoir can have a rich temporal response, which results in a higher effective dimensionality of the reservoir. This is due to the nodes occupying different states from one another, which essentially translates to approaching linear independence from one another. A higher effective dimensionality ensures that the number of nodes obtained are useful for solving the task at hand.

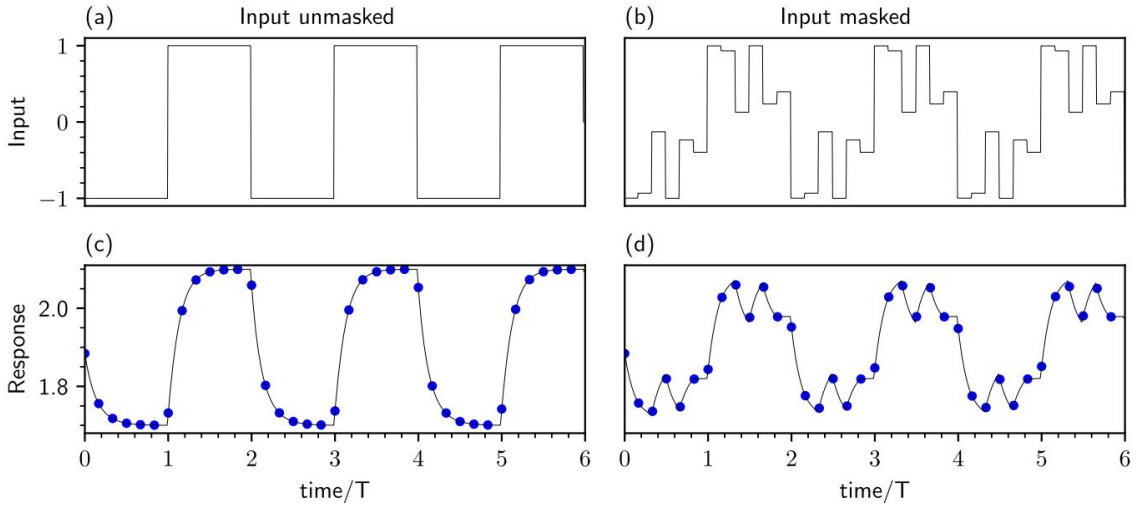

Fig. 2.6: The input masking protocol is commonly used in TDRC, adapted from [28].

Bibliography

[1] S. E. Miller, "Integrated optics: An introduction," *The Bell System Technical Journal*, vol. 48, no. 7, pp. 2059–2069, 1969.

[2] P. Reynaert and M. Steyaert, *RF POWER AMPLIFIERS FOR MOBILE COMMUNICATIONS*. Springer Netherlands, 2006. [Online]. Available: http://dx.doi.org/10.1007/1-4020-5117-4

[3] M. Hochberg, N. Harris, R. Ding, Y. Zhang, A. Novack, Z. Xuan, and T. Baehr-Jones, "Silicon photonics: The next fabless semiconductor industry," *IEEE Solid-State Circuits Magazine*, vol. 5, pp. 48–58, 2013.

[4] E. M. T. Fadaly, A. Dijkstra, J. R. Suckert, D. Ziss, M. A. J. van Tilburg, C. Mao, Y. Ren, V. T. van Lange, K. Korzun, S. Kölling, M. A. Verheijen, D. Busse, C. Rödl, J. Furthmüller, F. Bechstedt, J. Stangl, J. J. Finley, S. Botti, J. E. M. Haverkort, and E. P. A. M. Bakkers, "Direct-bandgap emission from hexagonal ge and sige alloys," *Nature*, vol. 580, no. 7802, pp. 205–209, Apr 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2150-y

[5] H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave raman silicon laser," *Nature*, vol. 433, no. 7027, pp. 725–728, Feb 2005. [Online]. Available: https://doi.org/10.1038/nature03346

[6] "Silicon photonics: to SOI and beyond! — yolegroup.com," https://www.yolegroup.com/press-release/silicon-photonics-to-soi-and-beyond/, [Accessed 30-08-2024].

[7] A. H. Atabaki, S. Moazeni, F. Pavanello, H. Gevorgyan, J. Notaros, L. Alloatti, M. T. Wade, C. Sun, S. A. Kruger, H. Meng, K. A. Qubaisi, I. Wang, B. Zhang, A. Khilo, C. V. Baiocco, M. A. Popović, V. M. Stojanović, and R. J. Ram, "Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip," *Nature*, vol. 556, pp. 349–353, 4 2018.

[8] V. Stojanović, R. J. Ram, M. Popović, S. Lin, S. Moazeni, M. Wade, C. Sun, L. Alloatti, A. Atabaki, F. Pavanello, N. Mehta, and P. Bhargava, "Monolithic silicon-photonic platforms in state-of-the-art cmos soi processes [invited]," *Optics Express*, vol. 26, p. 13106, 5 2018.

[9] C. Gunn, "Cmos photonics for high-speed interconnects," *IEEE Micro*, vol. 26, no. 2, pp. 58–66, 2006.

[10] European Commission, "Photonics," https://digital-strategy.ec.europa.eu/en/policies/photonics, 2024, accessed: 2024-04-28.

[11] https://www.yolegroup.com/press-release/could-optical-computing-solve-ais-power-demands/, [Accessed 19-10-2024].

[12] P. Kaur, A. Boes, G. Ren, T. G. Nguyen, G. Roelkens, and A. Mitchell, "Hybrid and heterogeneous photonic integration," 6 2021.

[13] V. Mere, F. Valdez, X. Wang, and S. Mookherjea, "A modular fabrication process for thin-film lithium niobate modulators with silicon photonics," *JPhys Photonics*, vol. 4, 4 2022.

[14] M. Manzo, F. Laurell, V. Pasiskevicius, and K. Gallo, *Lithium Niobate: The Silicon of Photonics!* Springer Netherlands, Sep. 2012, p. 421–422. [Online]. Available: http://dx.doi.org/10.1007/978-94-007-5313-6_42

[15] A. Boes, L. Chang, C. Langrock, M. Yu, M. Zhang, Q. Lin, M. Lončar, M. Fejer, J. Bowers, and A. Mitchell, "Lithium niobate photonics: Unlocking the electromagnetic spectrum," *Science*, vol. 379, no. 6627, p. eabj4396, 2023. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abj4396

[16] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. Wiley, Aug. 1991. [Online]. Available: http://dx.doi.org/10.1002/0471213748

[17] G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*. Springer US, 1993. [Online]. Available: http://dx.doi.org/10.1007/978-1-4613-0481-4

[18] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser & Photonics Reviews*, vol. 6, no. 1, pp. 47–73, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.201100017

[19] H. Jaeger, "The" echo state" approach to analysing and training recurrent neural networks-with an erratum note'," *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, vol. 148, 01 2001.

[20] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: A new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2002.

[21] Z. Konkoli, *Reservoir Computing*. Springer Berlin Heidelberg, 2017, pp. 1–12.

[22] C. Fernando and S. Sojakka, "Pattern recognition in a bucket," in *Advances in Artificial Life*, W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, and J. T. Kim, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 588–597.

[23] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 7 2019.

[24] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature Communications*, vol. 2, no. 1, p. 468, Sep 2011. [Online]. Available: https://doi.org/10.1038/ncomms1476

[25] M. Lakshmann and D. V. Senthilkumar, *Dynamics of nonlinear time delay systems*. Springer. [Online]. Available: http://www.springer.com/series/712

[26] A. Rodan and P. Tiňo, "Minimum complexity echo state network," *IEEE Transactions on Neural Networks*, vol. 22, pp. 131–144, 1 2011.

[27] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Scientific Reports*, vol. 2, no. 1, p. 287, Feb 2012. [Online]. Available: https://doi.org/10.1038/srep00287

[28] T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, "Role of delay-times in delay-based photonic reservoir computing," *Opt. Mater. Express*, vol. 12, no. 3, pp. 1214–1231, Mar 2022. [Online]. Available: https://opg.optica.org/ome/abstract.cfm?URI=ome-12-3-1214

Chapter 3

State of the Art

In this chapter, a brief overview of photonic neural networks is given, followed by a detailed report on the state of the art in photonic reservoir computing, shedding light on the most important advancements in the field. The research aims will then be presented within the context of current research trends and gaps in the literature.

3.1 Photonic Neural Networks

The rapid developments in AI and deep neural networks (DNNs) in the last decade signaled the need for new technologies that can allow faster, energy-efficient processing of MAC operations through hardware accelerators. On the one hand, photonic-based AI accelerators are being explored, for example in a systolic-array topology [1], or with optoelectronic time-frequency interleaving schemes [2]. Hardware accelerators are usually not full neural network implementations, it suffices that they serve to execute the resource-intensive parts of a neural network (such as MAC operations/convolution), with the rest being done on conventional hardware (e.g. computer/FPGA). On the other hand, there has been considerable research into realizing physical NNs in photonic hardware.

Photonic neural networks (PNNs) have recently gained considerable interest [3], owing largely to their promise of higher energy-efficiency per MAC and fast processing speeds. Like software implementations of NNs, PNNs are suitable for solving image recognition tasks, video processing (such as object detection), and natural language processing, among others. PNNs do not need to be neuromorphic, i.e. they do not need to be spike-driven. However, they usually realize the artificial neuron model,

whose output y is given by

$$y = \phi\left(\sum_i w_i x_i + b\right) \tag{3.1}$$

where w_i is the weight of the input x_i (e.g. an output of a preceding neuron with index i), b is a bias term, and ϕ being an activation function. The McCulloch-Pitts neuron model is usually associated with the term 'Perceptron' [4], where a hard thresholding (Heaviside) function allows a neuron to fire only when it has accumulated a specific value from its weighted inputs. However, it is more desirable that photonic implementations approximate some well-known nonlinear functions which are bounded and monotonic such as Re-Lu and sigmoid, which yield a richer variety of output. There are several possibilities to realize weight multiplication, summation, and nonlinearity in photonic hardware [5].

Weight multiplication

Multiplication can be done in photonics in multiple ways. In on-chip settings, interference paths can be tuned to apply the desired weight. For example, if two beams with the same wavelength interfere and combine, the strength of their combination can be tuned by, e.g., a heater element. The heater changes the optical path length experienced by one beam with respect to the other, yielding an output dependent on the phase difference between the two beams. Another way to realize weight multiplication is via resonance tuning with a ring resonator. The idea here is to modify the resonance of a ring resonator via a tuning element, consequently determining how much of the incoming signal, at a fixed wavelength, couples into the resonator. The resulting weighted signal can be collected in a drop waveguide. Phase change materials (PCMs), such as GeSbTe (GST), can be used to control the transmission of the optical signal in a non-volatile manner[6, 7]. They require an optical or electrical pulse to change their state between amorphous and crystalline, which then changes their optical transmission, thus the exiting signal carries the corresponding weight. Weight multiplication can also be implemented off-chip, e.g., with bulk MZIs or SLMs, which are usually used for diffraction-based implementations. Recent implementations mostly use MEMS-based digital micromirror devices (DMDs) such as SLMs, which provide the weight multiplication through tuning the positions of the individual mirrors, often referred to as pixels. Usually, they are either set to 'ON' (facing the light source) or 'OFF' state (facing away from the light source). Thus, the ensemble of pixels yield a rich diffraction pattern that cor-

responds to a weighted multiplication, which can be collected on a plane/screen. It is worth mentioning that, while the process of tuning consumes energy, the actual multiplication is done at no energy cost, i.e. it is an entirely passive process which only relies on the interference of the beams.

Summation

Summation can be done by several means, depending mainly on whether the implementation is coherence-based or utilizes wavelength-multiplexing. One way to do the summation in a coherent implementation is by using PCMs which can aggregate until a certain saturation point [8]. For spike-driven systems, micropillar lasers based on saturable absorbers can integrate multiple spike trains [9], which can in principle be done also using VCSELs and DFB lasers [5, 10, 11]. Furthermore, microrings can implement the summation function through light confinement up to the cavity lifetime. For wavelength-multiplexed systems, and depending on the channel spacing, a photodetector can be a means of achieving summation for signals on different wavelengths, which can be tuned individually through spectral filtering approaches [12].

Nonlinear activation function

Nonlinearities can be achieved in a number of ways in photonics, whether in integrated or bulk settings. In integrated approaches, one way is to use SOAs which, through gain saturation, exhibit a nonlinear, truncated 'S-curve' response to the input signal [13]. Optical nonlinearities can also be exploited, depending on the platform. For example, in SOI one can leverage the free carrier dispersion (FCD) or two photon absorption (TPA) to yield a nonlinear response in a microring resonator [14]. A reconfigurable all-optical nonlinearity was proposed, which is based on a microring and couplers which control the strength of the nonlinearity, and can achieve good approximations of many commonly used nonlinear functions in deep learning [15]. For optoelectronic implementations, a Mach-Zender Modulator (MZM) yields the $\sin(.)$ nonlinearity, the strength of which can be controlled by the biasing electrical signal on the modulator.

If the designed network is one layer deep, and is subsequently converted to the electrical domain (which is the common case), a nonlinearity can be obtained by what is known as the square law of the photodetector. The complex optical field E_{opt} is translated to a photodetector current I_{pd} which is proportional to the associated optical power $I_{\text{pd}} \propto P_{\text{opt}} \propto |E_{\text{opt}}|^2$, which means this scheme has the activation

function $\phi = |.|^2$. This type of nonlinearity is in fact very common in photonic implementations (especially photonic RC), sometimes dubbed as the 'photodetector nonlinearity', even though the photodetector does not actually provide the nonlinearity. A more accurate term could be the 'intensity-conversion nonlinearity' since it is associated with interfering fields in coherent implementations. For the intensity-conversion nonlinearity to take place, there must exist nonzero-phase-difference interference paths in the photonic circuit. From an optics point of view, interference itself is not a nonlinear phenomena, since it is the linear sum of the associated interfering fields. Consider two interfering complex fields E_1 and E_2 which have the same wavelength, the sum of their interference is $E_{\text{sum}} = E_1 + E_2$. The associated intensity (from $|.|^2$) becomes

$$I_{\text{sum}} = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos(\theta) \tag{3.2}$$

where θ is the phase difference between the two interfering fields. If we consider a black box implementation of this function, it is clear that, for $\theta \neq 0$, it is nonlinear as neither the additivity nor homogeneity conditions are satisfied. It is also clear that it is a sinusoidal nonlinearity. Thus, it is clear that the photodetector, as a stand-alone component, does not provide the nonlinearity.

Learning

One other feature that is essential to NNs is the means of implementing learning algorithms. This is usually done off-chip (offline learning), where the learned weights are then transformed to the photonic hardware. However, the goal is to eventually implement these learning algorithms on-chip (online learning). For real-time learning, on-chip learning is desirable as it avoids bottlenecks caused by electronic bandwidth limitations and increases the system's overall energy efficiency by reducing domain conversion overheads. Additionally, integrating electronic control with photonic components can limit scalability and increase overall system complexity. For DNNs, backpropagation is often used to train the weights for the different layers. In optical implementations, these weights are typically complex-valued (as they are associated with the complex-valued fields), where one phase shifting element, e.g., a heater, controls the intensity and another controls the phase. This can be utilized using MZIs or meshes of MZIs. The output of the network is probed and typically a backpropagation algorithm tunes the voltage on each heater. Reinforcement learning has also been explored in photonics [16] and implemented on RC-like physical networks [17]. Further strategies for online training have been explored and reported

in the literature [18, 19].

3.2 Examples of PNNs

In this section, we examine some examples of PNNs found in the literature.

3.2.1 Coherent PNNs

A recent on-chip scheme, shown in Fig. 3.1 (a), implements the perceptron using spatio-temporal multiplexing and passive delay lines [20]. The scheme relies on the intensity-conversion nonlinearity, and the complex weights are applied on-chip through heaters which change the phases of the signal, before being combined at the output photodiode. Similarly, feed-forward coherent PNNs rely on a similar setup, where they have multiple nodes arranged in a mesh, usually of a triangular or rectangular topology. The MZI mesh can implement any unitary transformation, as first demonstrated in [23]. Each node is represented by a MZI with two heaters that change the incoming light's phase and intensity, making it very popular for on-chip implementations [21, 24]. For free-space implementations, a popular scheme known as 'Deep diffractive neural networks' (D^2NN), similar to the 4-f correlator discussed in section 1.1.3, utilizes layers of diffractive planes situated between an input plane and an output plane [22, 25]. In this scheme, each layer is designed through simulating the network and training it with backpropagation to find the suitable complex transmission coefficients for each segment of each layer. The layers are then fabricated using 3-D printing (lithography is another possibility). However, one of the limitations is the need for a photodetector array that scans the spots corresponding to the nodes, requiring large energy overheads. Another limitation is the difficulty of scaling down this scheme to a chip level. Moreover, most coherent implementations can only allow the processing of one signal at a time.

3.2.2 Wavelength-multiplexed PNNs

On the other hand, PNNs can leverage wavelength-division-multiplexing (WDM) to process several signals at once, each on a carrier at a different wavelength. In a WDM system, multiple channels can use the same physical medium. Respecting the minimum distance between channels will result in good isolation and prevent unwanted crosstalk and photodetector interference. A spatially-multiplexed WDM-based neuromorphic photonic implementation can realize SNNs, such as in [12]. In this case, each neuron block comprises: an array of tunable add-drop

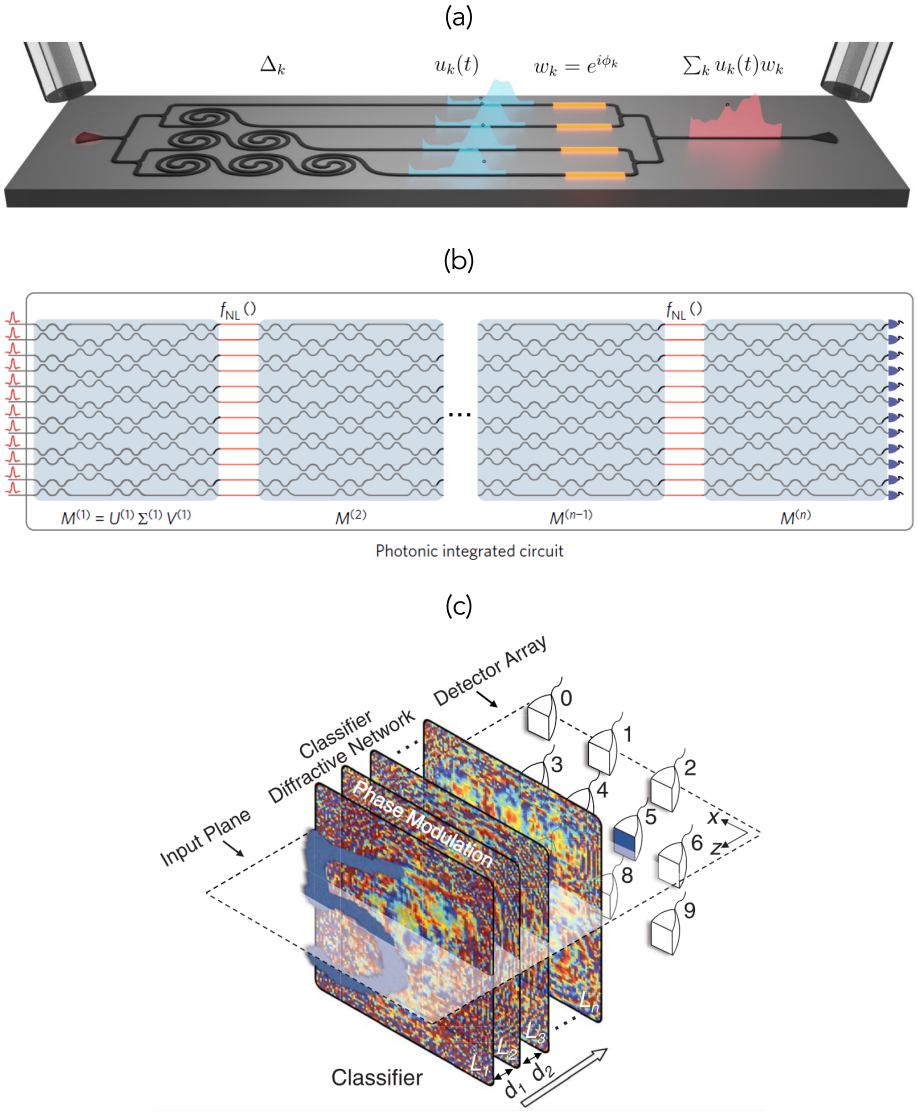

Fig. 3.1: Some coherent PNNs in the literature, (a) all-optical perceptron (adapted from [20]), (b) MZI-meshes implementing the unitary transformation (adapted from [21]), (c) free-space D^2NN implementation (adapted from [22]).

microrings to implement the weights via spectral filtering, a balanced photodetector to implement the summation (with +/- sign for the weights), and a laser or MZI to realize the E/O conversion. In this scheme, the signal is sent back to the broadcast interconnect which contains the outputs of the other neurons. This approach was shown to solve the well-known differential system emulation task: the Lorenz system (also a popular RC benchmark task). As can be seen in Fig. 3.2 (a), this approach is footprint-demanding, and can impose some restrictions due to the sensitivity of the microring elements, which would be used in abundance if a large number of neurons are needed. Furthermore, there are large energy overheads involved for the tunable elements, detection, and E/O conversion. In Fig. 3.2 (b), a photonic

WDM-based perceptron (which can also serve as a convolutional accelerator) is shown [26], which exploits 49 wavelengths produced from a kerr-microcomb source. The scheme was able to achieve high accuracy on the handwritten digit recognition (93%) and cancer-cell detection (86%) benchmark tasks.

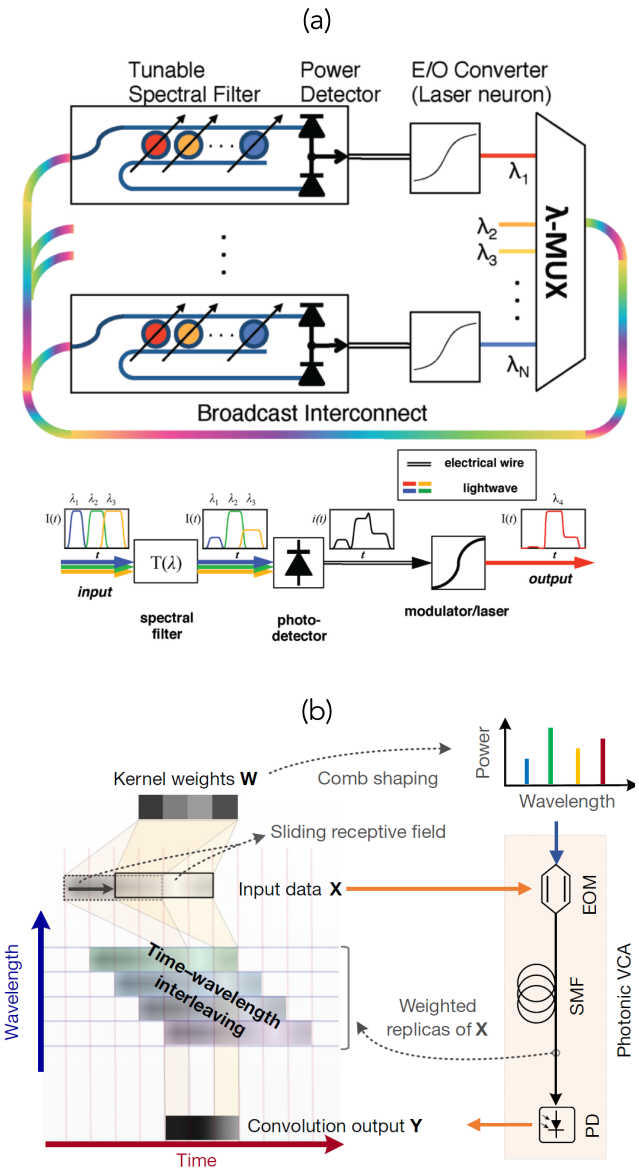

Fig. 3.2: Some WDM-based PNNs in the literature, (a) Broadcast and weight protocol (adapted from [12]), (b) Time-wavelength multiplexed vector convolution accelerator (adapted from [2]).

The difficulties associated with realizing, scaling, and training fully connected NNs in photonics has made RC an attractive choice, since in RC only the output weights need to be trained. RC can even be realized with entirely passive photonic circuits, as will be discussed in the following section.

3.3 Photonic Reservoir Computing

Over the years, photonics has proven to be especially well-suited to RC. RC can be classified, at least in principle, as a type of RNN, where the input weights and internal (connectivity) weights do not participate in the training process. The hardware approach to RC differs fundamentally from PNNs, one of the key differences being that the weights are usually applied post-readout in RC as a post-processing step. From an applications standpoint, RC works best for time-series prediction tasks, where the memorization of inputs in the recent past is required. However, it has also been shown to be capable of solving image recognition tasks such as the Iris dataset [27] and the MNIST hand-digit recognition [28]. A concise list of RC applications and their application domains can be found in this review article [29].

A standard photonic RC scheme comprises:

- Input generation stage; the RF input, which is typically high-speed, is usually defined in software and generated with an arbitrary waveform generator. Input pre-processing, if required, is also done at this stage.
- Modulation stage; the RF signal is used to modulate a laser source. Indirect (e.g. electro-optic) or direct (e.g. current) modulation approaches can be used, depending on the reservoir scheme and system bandwidth.
- Reservoir stage; essentially a high-dimensional substrate. This can be implemented using a laser diode, electro-optic modulator, passive optical waveguides (provided interference is involved), or devices leveraging optical nonlinearities.
- Output stage; typically in the form of an electronic readout consisting of photodiode(s), associated amplification circuitry, and analog-to-digital (ADC) conversion for sampling and recording data.
- Data processing stage; part of the recorded data is then used to train the reservoir output weight matrix. The learned weights can then be applied on the remaining data. For continuous operation, the same weights can be applied only if the reservoir interconnectivity weight matrix does not change. This is not a trivial matter.

Different RC implementations can be compared regardless of the substrate or implementation through some well-known benchmark tasks, which are explained in detail in a recent review article [30]. In the following, we will discuss some of the various implementations of photonic RC using off-the-shelf bulk components and PICs.

3.3.1 Bulk implementations

Similar to the synchronicity in the development of the LSM and ESN mathematical frameworks, the first RC demonstrations using bulk optics were published around the same time by two different research groups [31, 32]. Both used an optoelectronic implementation which comprises a MZM as the nonlinear node, an optical fiber as the delay line, and optoelectronic feedack, as shown in Fig. 3.3. The feedback signal is then combined with the electronic input before both are sent back to the MZM. This time-multiplexing approach to yield 'virtual' nodes is referred to in the literature as 'Time-delay Reservoir Computing' (TDRC) and as 'Delay-based Reservoir Computing'. In [32], a 400-node reservoir of this configuration was driven at a relatively slow speed of 50 kHz (τ_d = 20.9 µs) to solve the Santa Fe one-step prediction (NMSE = 0.124) and the spoken digit recognition, yielding a word error rate WER < 0.005. In [31], a 50-node optoelectronic configuration was driven at around 118 kHz (τ_d = 8.5 µs) to solve the NARMA-10 task (NMSE = 0.168), signal classification task (NMSE = 1.5×10^{-3}), nonlinear channel equalization task (symbol error rate SER = 1.3×10^{-4}), and isolated spoken digit recognition task (WER = 0.004 with 200 nodes).

All-optical configurations were proposed shortly after the first optoelectronic demonstrations, which used different sources of nonlinearity. The following examples, which were instrumental in advancing photonic RC, were all carried out within the same research group led by S. Massar. One implementation used the SOA saturation as the nonlinearity [33]. The following one used a semiconductor saturable absorption mirror (SESAM) [34] as a passive nonlinearity. A third implementation leveraged a coherently-driven passive cavity to realize the intensity-conversion nonlinearity [35]. In the first all-optical implementation [33], the nonlinear channel equalization task was solved with 50 nodes (SER = 4×10^{-4}), slightly lower than the optoelectronic performance. The same number of nodes was used to solve the radar time-series prediction task (NMSE = 3.0×10^{-3} for 1-step prediction). Furthermore, the isolated spoken digit recognition task was solved with 200 nodes, yielding an order of magnitude worse performance (WER = 0.04) than the optoelectronic RC. In an improved implementation [34], the same research group investigated the low-power nonlinearity from the SESAM implementation which, unlike the SOA implementation [33], does not need to be powered. In this light, this work can be considered as the first step towards leveraging passive nonlinearities. For certain values of input power, the performance on the nonlinear channel equalization task is comparable to the SOA and optoelectronic reservoirs (SER $\sim 10^{-4}$). For the radar task, the results also showed comparable performance to the SOA and op-

toelectronic implementations. For the isolated spoken digit recognition, the result (WER = 0.046) is close to the SOA implementation. In [35], the passive cavity was able to outperform the previous all-optical and optoelectronic implementations while relying only on the intensity-conversion nonlinearity. For the NARMA-10 task, it achieved a minimum NMSE = 0.107 using 50 nodes. For the same number of nodes, the nonlinear channel equalization task was solved perfectly with SER = 0.0 at sufficiently high SNR. For the isolated spoken digit task, a performance of WER = 0.0 was obtained using 200 internal variables. The excellent results in this study speak for the merits of the careful experimental setup, which reduced noise to a minimum and allowed good performance with lower SNR than previous implementations. This was also made possible by removing active components from the reservoir itself. More importantly, it showed the potential of a passive cavity, which relies only on the intensity-conversion nonlinearity, to achieve top performance on several RC benchmark tasks. More recently, VCSEL polarization dynamics have been explored within a fiber cavity [36], showing good results on the Santa-Fe (NMSE = 1.6×10^{-2}) and nonlinear channel equalization (SER = 1.5×10^{-2}) tasks. VCSELs have also been used in diffractive RC networks [37] to solve the temporal bitwise XOR task (BER < 0.01, 2-bits) and the optical header recognition task.

One can summarize that the main theme of research that links the aforementioned studies is the exploration of what can be achieved by off-the-shelf photonic components in RC. Such an approach was necessary, as the formal framework of RC does not readily predict the computational power and performance of certain substrates/implementations. This is further augmented by the fact that, while ESNs are considered universal approximators under certain conditions, there has been no straightforward link between a dynamical system's characteristics and predicting its performance on specific RC tasks. Thus, these concerns have been in themselves a topic of recent exploration in theoretical RC, whether through attempting to provide a substrate-independent framework for RC [38], or linking dynamical system analysis and task-independent metrics to task-specific performance [39, 40].

Alternatively, some studies were concerned with the theme of reservoir connectivity. Deep RC was proposed theoretically to leverage the interplay between various timescales in a dynamical system [41], where layers of RC are stacked in a forward, series connectivity. This was followed later by the first proposals (in bulk) of deep photonic reservoirs in 2023 and 2024 [42, 43, 44, 45]. In [44], it was experimentally shown that, for high enough SNR values, two orders of magnitude improvement was obtained with deep RC over shallow RC for the Santa Fe and nonlinear channel equalization benchmark tasks.

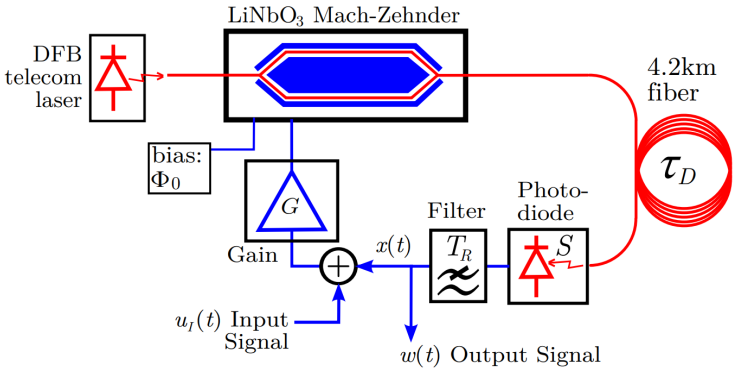

Fig. 3.3: First demonstration: An optoelectronic RC, (adapted from[32]). The MZM acts as the nonlinear node, where an optoelectronic feedback closes the loop back to the MZM after being combined with the electronic input.

Other themes relating to solving multiple tasks and techniques for RC performance enhancement were also explored. Shortly after the first photonic RC demonstrations, the ability to solve two tasks at the same time with the same RC was demonstrated in a pioneering study in 2013 [46]. The impact of delay times was explored in [40], where the authors highlighted the historical convention of the synchronous regime and showed that good performance can be obtained outside of it. Furthermore, they have showed that resonance between the input clock cycle and the delay time can significantly degrade RC performance, especially in terms of the memory capacity of the system. In [47], the operation in the asynchronous regime was experimentally investigated with a reservoir comprising a semiconductor laser in optical feedback, where the authors reported significantly better performance on the Santa-Fe task while using a considerably longer delay time than the input clock cycle ($\tau_\mathrm{d} = 3.5T_\mathrm{c}$). In another study of delay-based RC with semiconductor lasers, the authors investigate the impact of different input masks on the Santa Fe benchmark tasks [48]. The masks considered were binary, chaos, and colored noise. The key finding was that chaos and colored noise masking provided much better performance than binary masking. In another study, the effect of post-filtering was investigated for an optoelectronic reservoir [49], showing improved performance with lower bandwidth post-filtering.

More recently, different modulation schemes have been investigated as well. In an RC implementation using semiconductor lasers with external cavity [50], the authors explored the difference in performance when considering direct intensity modulation and phase modulation of the semiconductor laser. The key finding was that performance significantly improved when using phase modulation. Another approach that used phase-modulation of a pulsed laser was investigated, where the readout was implemented using homodyne detection (i.e. access to information regarding

both phase and amplitude) [51], showing good results on the NARMA-k task for various k. Furthermore, another study investigated parameter-drift on a reservoir based on a semiconductor with optical feedback [52]. In this study, the authors present transfer learning as a method to compensate for the ambient temperature fluctuations. In [53], a method of countering phase noise – due to changes in the optical path length of a Fabry-Perot resonator – was proposed, where the output weights themselves vary with the phase noise. The result is that only a small performance degradation was observed over a range of fluctuations, without the need for employing setup stabilization techniques.
It is worth mentioning that for all the above-mentioned photonic RC schemes (except [53]), the time-multiplexing approach was used. In other words, they are all delay-based/TDRC implementations. This shows the considerable popularity of time-multiplexed RC schemes over their spatially-multiplexed counterparts. For parallel and deep RC implementations, one could consider these schemes to be of a spatiotemporal nature, since multiple physical nodes would be typically required.

3.3.2 PIC implementations

The first PIC RC implementations were proposed a few years before the first bulk demonstrations. The first proposed approach dates back to 2008, which involved SOI-based SOAs as nonlinear nodes in a spatially multiplexed approach comprising 9 nodes [54] (Fig. 3.4 (a)). In this study, the authors compared the performance of the SOA model's nonlinearity and the widely used – in machine learning – $\tanh(.)$ nonlinearity for a feed-forward topology of SOAs and one with feedback. Using an input clock rate of 500 MHz, they obtained good results on the signal pattern recognition task for the SOA feedback topology. The topology was later optimized with what was coined the 'swirl' topology, which realized 16 nodes of SOAs [55] (Fig. 3.4 (b)). In this scheme, the input was fed to each node, and the impact of coherent vs incoherent source was investigated. Three years later, in a pioneering study reporting the first experimental on-chip (SOI) RC demonstration, P. Bienstman's research group used the same topology of connectivity, but dropped the SOAs, thus realizing the first all-passive integrated reservoir which relied on the intensity-conversion nonlinearity at the readout [56] (Fig. 3.4 (c)). The scheme showed good performance on the bitwise XOR task between various bits into the past. Spirals were introduced between the nodes to slow down the system to speeds which are detectable by high-speed electronics, which significantly impacted the chip footprint (16 mm^2). Furthermore, 5 of the 16 nodes were unusable, due to the low power values obtained at these nodes. A further improvement in this scheme over its SOA

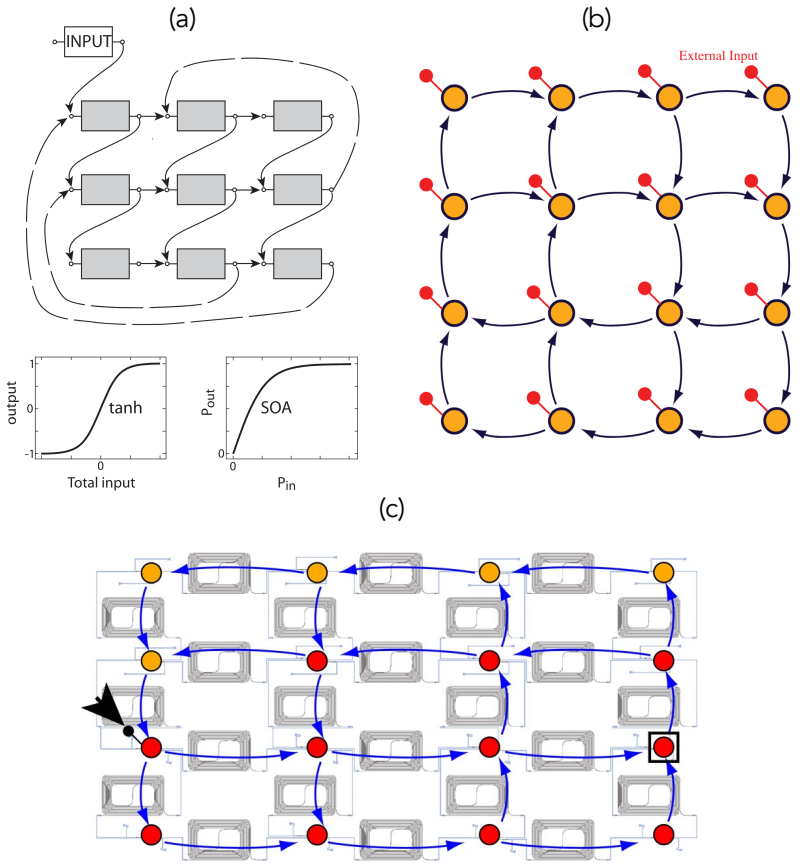

Fig. 3.4: Evolution of the 'swirl' spatially-multiplexed topology: (a) 9 SOA nodes, (b) 16 SOA nodes, (c) 16 passive nodes. Adapted from [54, 55, 56].

predecessor is the use of only one node for the input. Moreover, since this scheme is not reconfigurable, the good performance on the various tasks was accomplished through tuning the input bitrate with respect to the fixed interconnection delay, which can be considered another limitation of this study. Additionally, it required one photodiode per node to implement the readout.

The same research group did further investigations on the swirl topology. In [57], it was shown that using a binary mask (masking is usually done with TDRC) with SOAs allows the timescale of the system to be reduced such that it can be sampled at slower speeds. The motivation for this study was the naturally high on-chip bandwidth that requires costly pre-processing and detection electronics – in the form of signal generators and oscilloscopes which may not be readily available – to operate the system. Moreover, real-time signal processing applications such as speech recognition would naturally require lower bandwidths. Significantly better results were obtained for the speech recognition task through using the mask (WER ~ 0.1), which allowed $40\times$ reduction in signal processing speed, than the unmasked version (WER ~ 0.3). In [58, 59], an all-optical readout was considered. The weights are

pre-trained in simulation, which are then applied to the reservoir through optical modulators, and then fine-tuned on-chip to account for fabrication variations. This entails that the on-chip weights are complex, which is further explored in their work. This scheme improves on the previous experimentally demonstrated work [56] by requiring only one photodiode at the readout which sums all the states of the nodes. However, it requires one MZM per node, which can be a limitation when scaling up the reservoir size. It has been recently demonstrated experimentally in [19], where a detailed algorithm was proposed to perform the on-chip tuning of the complex weights. Another avenue of research explored by the Bienstman group was the use of an ensemble of passive reservoirs, which proposes for the first time parallel/deep on-chip RC [60], which also predates the first bulk implementations of such networks. In this numerical study, the authors consider different network topologies between the reservoirs, whether through ensembling, boosting, stacking, or chaining. They also consider a photodiode model to account for its noise. The key finding was that a considerable (10×) improvement was obtained on the 5-bit header recognition task over the baseline, which constituted a single reservoir comprising the same total number of nodes distributed among the different reservoirs. For the 3-bit XOR and Santa-Fe tasks, the ensemble topology achieved little improvement (BER = 0.001, NMSE = 0.057) over the baseline (BER = 0.006, NMSE = 0.073). The authors concluded that different types of connection topologies are more suited to some tasks over others.

Another interesting direction pursued by the same group was the training for robustness on the passive swirl architecture (similar to the aforementioned studies for bulk RC). Since PICs are susceptible to environmental fluctuations (e.g. temperature), it is of importance to consider ways to mitigate its effects on the RC performance. In their study [61], the authors train the output of a reservoir under input of several wavelengths in the vicinity of the principle one. The key finding is that a weight matrix can be found which achieves good performance on the XOR task for a range of wavelengths corresponding to what would be encountered in an experimental setting through parameter drift (mainly temperature).

In [62], the same group considered a different approach using on-chip DFB lasers and a Fabry-Perot cavity and experimentally verified it. Similarly, another scheme utilizing a DFB laser and a tunable external cavity was investigated by another group [63], which can be considered the first on-chip proposal and experimental demonstration of photonic TDRC. In another study by the Bienstman group [64], the potential of SOI-based microring resonators (MRRs) and their nonlinearities for RC is discussed. In that numerical study, the authors essentially substitute the

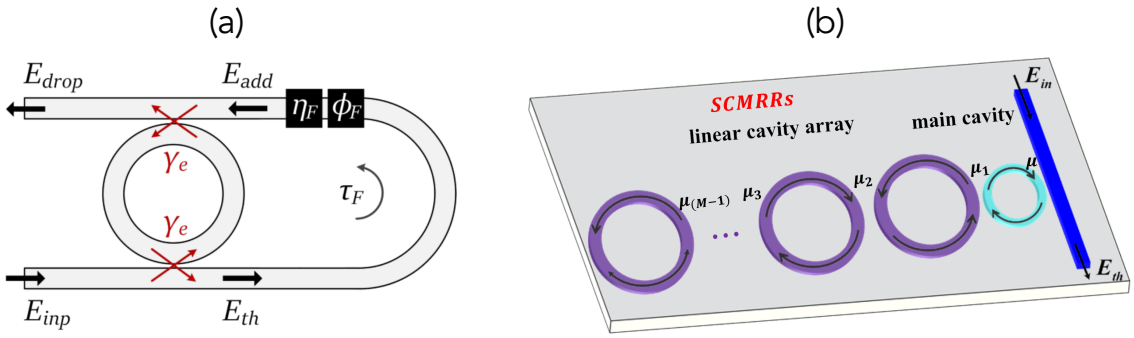

Fig. 3.5: MRR-based TDRC: (a) a MRR in add-drop configuration can solve certain tasks by itself and larger-memory tasks with an external feedback loop, (b) A cascaded MRR scheme significantly enhances the memory capacity without requiring an external feedback. Adapted from [65, 66].

SOAs for microring resonators as the nonlinear nodes, using the same swirl topology, and investigate its RC performance. The use of MRRs was later investigated in more depth by another research group led by L. Pavesi, albeit for TDRC rather than spatially-multiplexed RC.

In the first work on RC by Pavesi's group, MRR-based TDRC was experimentally validated for the first time [67, 27], where the 1-bit XOR (BER = 0.0, 3 nodes with input datarate < 25 Mbps and no signal masking) and Iris image classification (50 nodes at an input datarate of 380 MHz) were successfully demonstrated. In a follow-up numerical study [65], the add-drop configuration was augmented with an off-chip delay line – in the form of a single mode fiber – that was shown to enhance the memory capacity of the system, thereby yielding good results on large-memory tasks such as NARMA-10, which are not solvable by the MRR's intrinsic memory. This was followed by a recently published study which reports the experimental results of this approach [68]. In [69] the authors investigate the effect of noise, which was later expanded to include the effects of the setup and input non-idealities on the RC performance, isolated from the device under test (DUT) [70]. A recent work by another group addresses the limited intrinsic memory capacity of MRRs by adding cascaded ring resonators of larger radii (operating in the linear regime) in a CROW configuration to the principal nonlinear MRR [66]. In that numerical study, the authors took into account the effects of fabrication variations on the performance of the system. For the NARMA-10 task, the study reports NMSE = 0.156 with an array of 10 linear MRRs as opposed to NMSE = 0.534 obtained with a single MRR. Lastly, another research group recently proposed exploiting cross-gain modulation (XGM) using a SOA as a nonlinear node for TDRC [71]. In this scheme, the input is fed through two ports to the saturated SOA where they modulate each other. Two SOAs work in parallel by means of a silicon-based MZI which divides the input

power evenly across them. The delay was implemented with an external (off-chip) delay line consisting of a fiber loop. The authors report good performance on the Santa-Fe task (NMSE = 0.103).

3.4 Beyond State of the Art

By now, I hope it is clear to the reader the extent and breadth of the literature on photonic RC. What is most impressive is the fact that, while photonic RC could still be considered a nascent field, there have been already many pioneering studies in different directions that explored much of the important questions that needed addressing. Nevertheless, we did notice a few gaps that may have required the attention given in this dissertation.

3.4.1 Original Contributions

First of all, we have seen many implementations in the literature utilizing different kinds of nonlinearities (whether in bulk or integrated). The passive, spatially multiplexed architecture [56], and the time-multiplexed bulk passive architecture [35], have shown great promise when utilizing only the intensity-conversion nonlinearity. However, in [56], the only way to tune the proposed scheme is by means of changing the input data rate, which may not be a practical degree of freedom within an applications setting. Another merit of using only the intensity-conversion nonlinearity – aside from the simplified design considerations – is the fact that it is not limited by the timescales of some nonlinear processes (e.g. thermal nonlinearities in SOI). Thus the RC system is not speed-limited except by the readout electronics and post-readout processing. Furthermore, schemes that utilize MRR-based nonlinearities [27, 65, 66] require extensive hardware to probe the MRR detuning (which in itself requires precision). Moreover, many of the photonic nonlinearities previously/subsequently investigated did not provide significant improvement across the RC spectrum in terms of memory capacity and benchmark tasks (in fact, in some cases they show inferior performance to the passive schemes). This prompted us to consider the design of a reconfigurable TDRC architecture that minimizes hardware complexity by reducing the number of physical parameters that need to be controlled, while leveraging the low losses of the LNOI platform to fully integrate a delay line on-chip to provide a high enough memory capacity on a PIC implementation.

Additionally, while the impact of using different input masks on photonic TDRC has been investigated [48], there has been no mention of a mask-less TDRC scheme

in the literature. In fact, it has been only briefly brushed upon in [27] where the authors reported solving the 1-bit XOR task in the absence of an input mask. This gave rise to our second direction of investigation, where we address the absence of input masking in TDRC using our designed architecture. This is of interest as it further reduces the complexity of the pre-processing stage, thereby allowing an incoming optical signal to directly interact with the reservoir and bypassing possible domain conversions at the input layer. Furthermore, it is of interest to also consider the effects of the electronic readout on this scheme, as it imprints noise and filtering which affect the final acquired signal, as has been done in the optoelectronic implementation (albeit for the electronic feedback) [49].
Finally, we consider the asynchronous operation of the proposed reservoir, in line with recent trends in the literature [40, 47], and in contrast to the historical convention of the quasi-synchronization of the input clock cycle with the feedback loop's delay time. This would allow the optimization of the reservoir design as well as the possible reduction of the feedback length when the considered task does not benefit from longer delay lines, thereby reducing excess on-chip footprint, in addition to providing interesting system dynamics which can be exploited.

3.4.2 Research Questions

The research questions addressed in this work can be summarized as follows:

1. **RQ1**: Can we design an integrated photonic reservoir to minimize operational complexity while still being reconfigurable and achieving good performance on benchmark tasks?

2. **RQ2**: What is the impact of the absence of input masking on this scheme's performance, in terms of task-independent metrics and benchmark tasks?

3. **RQ3**: How is the performance of the proposed reservoir affected when driven in the asynchronous regime?

4. **RQ4**: What is the impact of the electronic readout on the reservoir performance? In particular, how does the photodiode bandwidth and noise affect the RC performance?

5. **RQ5**: What is the experimental performance of this architecture when fabricated on LNOI and operated in the asynchronous, mask-less regime?

RQ1 is addressed in chapter 4, **RQ2**, **RQ3**, and **RQ4** are addressed in chapter 5, and **RQ5** is addressed in chapter 6.

Bibliography

[1] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, no. 7840, pp. 52–58, Jan 2021. [Online]. Available: https://doi.org/10.1038/s41586-020-03070-1

[2] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 tops photonic convolutional accelerator for optical neural networks," *Nature*, vol. 589, no. 7840, pp. 44–51, Jan 2021. [Online]. Available: https://doi.org/10.1038/s41586-020-03063-0

[3] A. Tsakyridis, M. Moralis-Pegios, G. Giamougiannis, M. Kirtas, N. Passalis, A. Tefas, and N. Pleros, "Photonic neural networks and optics-informed deep learning fundamentals," *APL Photonics*, vol. 9, no. 1, p. 011102, 01 2024. [Online]. Available: https://doi.org/10.1063/5.0169810

[4] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec 1943. [Online]. Available: https://doi.org/10.1007/BF02478259

[5] F. P. Sunny, E. Taheri, M. Nikdast, and S. Pasricha, "A survey on silicon photonics for deep learning," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 17, 10 2021.

[6] Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, "On-chip photonic synapse," *Science Advances*, vol. 3, no. 9, p. e1700160, 2017. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.1700160

[7] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, May 2019. [Online]. Available: https://doi.org/10.1038/s41586-019-1157-8

[8] C. D. Wright, Y. Liu, K. I. Kohary, M. M. Aziz, and R. J. Hicken, "Arithmetic and biologically-inspired computing using phase-change materials," *Advanced Materials*, vol. 23, no. 30, pp. 3408–3413, 2011. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201101060

[9] F. Selmi, R. Braive, G. Beaudoin, I. Sagnes, R. Kuszelewicz, and S. Barbay, "Temporal summation in a neuromimetic micropillar laser," *Opt. Lett.*, vol. 40, no. 23, pp. 5690–5693, Dec 2015. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-40-23-5690

[10] J. Robertson, E. Wade, and A. Hurtado, "Electrically controlled neuron-like spiking regimes in vertical-cavity surface-emitting lasers at ultrafast rates," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 6, pp. 1–7, 2019.

[11] I. Aldaya, C. Gosset, C. Wang, G. Campuzano, F. Grillot, and G. Castañón, "Periodic and aperiodic pulse generation using optically injected dfb laser," *Electronics Letters*, vol. 51, no. 3, pp. 280–282, 2015. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/el.2014.3927

[12] A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Scientific Reports*, vol. 7, no. 1, p. 7430, Aug 2017. [Online]. Available: https://doi.org/10.1038/s41598-017-07754-z

[13] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, P. Bienstman, R. Baets, and J. Van Campenhout, "Photonic reservoir computing with coupled semiconductor optical amplifiers," in *Optical SuperComputing*, S. Dolev, T. Haist, and M. Oltean, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 46–55.

[14] G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Optics Express*, vol. 30, p. 522, 1 2022.

[15] A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," *Opt. Lett.*, vol. 45, no. 17, pp. 4819–4822, Sep 2020. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-45-17-4819

[16] M. Naruse, Y. Terashima, A. Uchida, and S.-J. Kim, "Ultrafast photonic reinforcement learning based on laser chaos," *Scientific Reports*, vol. 7, no. 1, p. 8772, Aug 2017. [Online]. Available: https://doi.org/10.1038/s41598-017-08585-8

[17] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural

network," *Optica*, vol. 5, no. 6, pp. 756–760, Jun 2018. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-5-6-756

[18] P. Antonik, M. Haelterman, and S. Massar, "Online training for high-performance analogue readout layers in photonic reservoir computers," *Cognitive Computation*, vol. 9, no. 3, pp. 297–306, Jun 2017. [Online]. Available: https://doi.org/10.1007/s12559-017-9459-3

[19] C. Ma, J. V. Kerrebrouck, H. Deng, S. Sackesyn, E. Gooskens, B. Bai, J. Dambre, and P. Bienstman, "Integrated photonic reservoir computing with an all-optical readout," *Opt. Express*, vol. 31, no. 21, pp. 34 843–34 854, Oct 2023. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-21-34843

[20] M. Mancinelli, D. Bazzanella, P. Bettotti, and L. Pavesi, "A photonic complex perceptron for ultrafast data processing," *Scientific Reports*, vol. 12, no. 1, p. 4216, Mar 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-08087-2

[21] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nature Photonics*, vol. 11, no. 7, pp. 441–446, Jul 2017. [Online]. Available: https://doi.org/10.1038/nphoton.2017.93

[22] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aat8084

[23] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, pp. 58–61, Jul 1994. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.73.58

[24] Y. Tian, Y. Zhao, S. Liu, Q. Li, W. Wang, J. Feng, and J. Guo, "Scalable and compact photonic neural chip with low learning-capability-loss," *Nanophotonics*, vol. 11, no. 2, pp. 329–344, 2022. [Online]. Available: https://doi.org/10.1515/nanoph-2021-0521

[25] Y. Qu, H. Lian, C. Ding, H. Liu, L. Liu, and J. Yang, "High-frame-rate reconfigurable diffractive neural network based on superpixels," *Opt. Lett.*, vol. 48, no. 19, pp. 5025–5028, Oct 2023. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-48-19-5025

[26] X. Xu, M. Tan, B. Corcoran, J. Wu, T. G. Nguyen, A. Boes, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, D. G. Hicks, and D. J. Moss, "Photonic perceptron based on a kerr microcomb for high-speed, scalable, optical neural networks," *Laser & Photonics Reviews*, vol. 14, no. 10, p. 2000070, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.202000070

[27] M. Borghi, S. Biasi, and L. Pavesi, "Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations," *Scientific Reports*, vol. 11, no. 1, p. 15642, Aug 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-94952-5

[28] A. Jalalvand, G. Van Wallendael, and R. Van De Walle, "Real-time reservoir computing network-based systems for detection tasks on visual contents," in *2015 7th International Conference on Computational Intelligence, Communication Systems and Networks*, 2015, pp. 146–151.

[29] G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, "Recent advances in physical reservoir computing: A review," *Neural Networks*, vol. 115, pp. 100–123, 7 2019.

[30] C. Wringe, M. Trefzer, and S. Stepney, "Reservoir computing benchmarks: a review, a taxonomy, some best practices," 2024. [Online]. Available: https://arxiv.org/abs/2405.06561

[31] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Scientific Reports*, vol. 2, no. 1, p. 287, Feb 2012. [Online]. Available: https://doi.org/10.1038/srep00287

[32] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing," *Opt. Express*, vol. 20, no. 3, pp. 3241–3249, Jan 2012. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-20-3-3241

[33] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express*, vol. 20, no. 20, pp. 22 783–22 795, Sep 2012. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-20-20-22783

[34] A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of

absorption," *Opt. Express*, vol. 22, no. 9, pp. 10 868–10 881, May 2014. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-22-9-10868

[35] Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, no. 5, pp. 438–446, May 2015. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-2-5-438

[36] J. Vatin, D. Rontani, and M. Sciamanna, "Experimental reservoir computing using vcsel polarization dynamics," *Opt. Express*, vol. 27, no. 13, pp. 18 579–18 584, Jun 2019. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-13-18579

[37] M. Pflüger, D. Brunner, T. Heuser, J. A. Lott, S. Reitzenstein, and I. Fischer, "Experimental reservoir computing with diffractively coupled vcsels," *Opt. Lett.*, vol. 49, no. 9, pp. 2285–2288, May 2024. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-49-9-2285

[38] M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer, "A substrate-independent framework to characterize reservoir computers," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 475, no. 2226, p. 20180723, 2019. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2018.0723

[39] F. Köster, S. Yanchuk, and K. Lüdge, "Insight into delay based reservoir computing via eigenvalue analysis," *JPhys Photonics*, vol. 3, 4 2021.

[40] T. Hülser, F. Köster, K. Lüdge, and L. Jaurigue, "Deriving task specific performance from the information processing capacity of a reservoir computer," *Nanophotonics*, 3 2022.

[41] C. Gallicchio, A. Micheli, and L. Pedrelli, "Deep reservoir computing: A critical experimental analysis," *Neurocomputing*, vol. 268, pp. 87–99, 2017, advances in artificial neural networks, machine learning and computational intelligence. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217307567

[42] H. Hasegawa, K. Kanno, and A. Uchida, "Parallel and deep reservoir computing using semiconductor lasers with optical feedback," *Nanophotonics*, vol. 12, no. 5, pp. 869–881, 2023. [Online]. Available: https://doi.org/10.1515/nanoph-2022-0440

[43] Y.-W. Shen, R.-Q. Li, G.-T. Liu, J. Yu, X. He, L. Yi, and C. Wang, "Deep photonic reservoir computing recurrent network," *Optica*, vol. 10, no. 12, pp. 1745–1751, Dec 2023. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-10-12-1745

[44] A. Lupo, E. Picco, M. Zajnulina, and S. Massar, "Deep photonic reservoir computer based on frequency multiplexing with fully analog connection between layers," *Optica*, vol. 10, no. 11, pp. 1478–1485, Nov 2023. [Online]. Available: https://opg.optica.org/optica/abstract.cfm?URI=optica-10-11-1478

[45] E. Picco, A. Lupo, and S. Massar, "Deep photonic reservoir computer for speech recognition," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–9, 2024.

[46] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature Communications*, vol. 4, no. 1, p. 1364, Jan 2013. [Online]. Available: https://doi.org/10.1038/ncomms2368

[47] J.-Y. Tang, B.-D. Lin, Y.-W. Shen, R.-Q. Li, J. Yu, X. He, and C. Wang, "Asynchronous photonic time-delay reservoir computing," *Opt. Express*, vol. 31, no. 2, pp. 2456–2466, Jan 2023. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-2-2456

[48] Y. Kuriki, J. Nakayama, K. Takano, and A. Uchida, "Impact of input mask signals on delay-based photonic reservoir computing with semiconductor lasers," *Opt. Express*, vol. 26, no. 5, pp. 5777–5788, Mar 2018. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-26-5-5777

[49] G. O. Danilenko, A. V. Kovalev, E. A. Viktorov, A. Locquet, D. S. Citrin, and D. Rontani, "Impact of filtering on photonic time-delay reservoir computing," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 1, p. 013116, 01 2023. [Online]. Available: https://doi.org/10.1063/5.0127661

[50] K. Kanno, A. A. Haya, and A. Uchida, "Reservoir computing based on an external-cavity semiconductor laser with optical feedback modulation," *Opt. Express*, vol. 30, no. 19, pp. 34 218–34 238, Sep 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-19-34218

[51] J. Henaff, M. Ansquer, M. C. Soriano, R. Zambrini, N. Treps, and V. Parigi, "Optical phase encoding in a pulsed approach to reservoir computing,"

Opt. Lett., vol. 49, no. 8, pp. 2097–2100, Apr 2024. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-49-8-2097

[52] I. Bauwens, K. Harkhoe, P. Bienstman, G. Verschaffelt, and G. V. der Sande, "Transfer learning for photonic delay-based reservoir computing to compensate parameter drift," *Nanophotonics*, vol. 12, no. 5, pp. 949–961, 2023. [Online]. Available: https://doi.org/10.1515/nanoph-2022-0399

[53] R. Alata, J. Pauwels, M. Haelterman, and S. Massar, "Phase noise robustness of a coherent spatially parallel optical reservoir," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–10, 2020.

[54] K. Vandoorne, W. Dierckx, B. Schrauwen, D. Verstraeten, R. Baets, P. Bienstman, and J. V. Campenhout, "Toward optical signal processing using photonic reservoir computing," *Opt. Express*, vol. 16, no. 15, pp. 11 182–11 192, Jul 2008. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-16-15-11182

[55] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1469–1481, 2011.

[56] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Communications*, vol. 5, no. 1, p. 3541, Mar 2014. [Online]. Available: https://doi.org/10.1038/ncomms4541

[57] B. Schneider, J. Dambre, and P. Bienstman, "Using digital masks to enhance the bandwidth tolerance and improve the performance of on-chip reservoir computing systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 12, pp. 2748–2753, 2016.

[58] M. Freiberger, A. Katumba, P. Bienstman, and J. Dambre, "On-chip passive photonic reservoir computing with integrated optical readout," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, 2017, pp. 1–4.

[59] ——, "Training passive photonic reservoirs with integrated optical readout," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 7, pp. 1943–1953, 2019.

[60] M. Freiberger, S. Sackesyn, C. Ma, A. Katumba, P. Bienstman, and J. Dambre, "Improving time series recognition and prediction with networks and ensembles of passive photonic reservoirs," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–11, 2020.

[61] E. Gooskens, F. Laporte, C. Ma, S. Sackesyn, J. Dambre, and P. Bienstman, "Wavelength dimension in waveguide-based photonic reservoir computing," *Opt. Express*, vol. 30, no. 9, pp. 15 634–15 647, Apr 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-9-15634

[62] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Opt. Express*, vol. 28, no. 3, pp. 3086–3096, Feb 2020. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-28-3-3086

[63] K. Takano, C. Sugano, M. Inubushi, K. Yoshimura, S. Sunada, K. Kanno, and A. Uchida, "Compact reservoir computing with a photonic integrated circuit," *Opt. Express*, vol. 26, no. 22, pp. 29 424–29 439, Oct 2018. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-26-22-29424

[64] Rontani, Damien and Denis-Le Coarer, Florian and Katumba, Andrew and Freiberger, Matthias and Dambre, Joni and Bienstman, Peter and Sciamanna, Marc, "Reservoir computing with nonlinear micro-resonators on a silicon photonics chip," in *2017 International Symposium on Nonlinear Theory and Its Applications, NOLTA2017, Proceedings*, 2017, pp. 253–255. [Online]. Available: https://www.ieice.org/nolta/symposium/archive/2017/nolta17toc.pdf

[65] G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Opt. Express*, vol. 30, no. 1, pp. 522–537, Jan 2022. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-30-1-522

[66] H. Ren, Y. Li, M. Li, M. Gao, J. Lu, C.-L. Zou, C.-H. Dong, P. Yu, X. Yang, and Q. Xuan, "Photonic time-delayed reservoir computing based on series-coupled microring resonators with high memory capacity," *Opt. Express*, vol. 32, no. 7, pp. 11 202–11 220, Mar 2024. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-32-7-11202

[67] M. Borghi, S. Biasi, and L. Pavesi, "Experimental demonstration of reservoir computing with a silicon resonator and time multiplexing," in *2021 IEEE 17th International Conference on Group IV Photonics (GFP)*, 2021, pp. 1–2.

[68] G. Donati, A. Argyris, M. Mancinelli, C. R. Mirasso, and L. Pavesi, "Time delay reservoir computing with a silicon microring resonator and a fiber-based optical feedback loop," *Opt. Express*, vol. 32, no. 8, pp. 13 419–13 437, Apr 2024. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-32-8-13419

[69] G. Donati, A. Argyris, C. R. Mirasso, M. Mancinelli, and L. Pavesi, "Noise effects on time delay reservoir computing using silicon microring resonators," in *Integrated Optics: Devices, Materials, and Technologies XXVI*, S. M. García-Blanco and P. Cheben, Eds., vol. 12004, International Society for Optics and Photonics. SPIE, 2022, p. 120040U. [Online]. Available: https://doi.org/10.1117/12.2609056

[70] D. Bazzanella, S. Biasi, M. Mancinelli, and L. Pavesi, "A microring as a reservoir computing node: Memory/nonlinear tasks and effect of input non-ideality," *Journal of Lightwave Technology*, vol. 40, no. 17, pp. 5917–5926, 2022.

[71] T. Tsurugaya, T. Hiraki, T. Aihara, M. Nakajima, N.-P. Diamantopoulos, T. Segawa, and S. Matsuo, "Reservoir computing using on-chip xgm-based nonlinear processing by membrane soas on si-mzi," *Journal of Lightwave Technology*, vol. 42, no. 8, pp. 2859–2867, 2024.

Chapter 4

The Minimum Complexity Approach to Photonic Time-Delay Reservoir Computing

In this chapter, we address **RQ1** by proposing a novel integrated photonic architecture which leverages the low-loss LNOI platform, which is developed at RMIT's Integrated Photonics and Applications Center (InPAC), led by Arnan Mitchell. The proposed scheme is designed with the goal of minimizing complexity; i.e. reducing hardware requirements while also simplifying the overall operation of the RC scheme, especially in terms of reconfigurability. We numerically investigate its performance on various reservoir computing benchmark tasks. Some of the presented in this chapter are published in a journal article [1].

4.1 Motivation

One notable advantage of using hardware-based RC is that the input and internal weights do not require tuning. This means that the RC framework is robust to process variations typically encountered during PIC manufacturing steps. With this in mind, however, most coherent (single wavelength) nanophotonic systems suffer from sensitivity to environmental factors such as temperature fluctuations, limiting the RC operation time and making it difficult to have one set of weights that are reusable. In general, the parameters are optimized every time the photonic RC is utilized due to ambient fluctuations, especially in the cases of all-optical feedback. To the authors' knowledge, this remains an area to be explored with only a few recent examples in the literature proposing techniques to solve this issue, such as training for a given range of wavelengths corresponding to the range of

thermal fluctuations in a controlled setting [2], or using other machine learning techniques such as transfer learning [3]. Thus, while the cost of optimizing the weights themselves is minimal, it is paid for by the need for optimizing the system parameters, often requiring the scanning of a multi-dimensional parameter space. For example, the proposed VCSEL scheme in [4] and the microring-based scheme in [5] both require 4 (hyper)parameters to be optimized. Furthermore, while it is useful for some parameters such as the bitrate or input power to be tuned and optimized for the purpose of finding the global optimum for each specific task, they may not be readily available degrees of freedom for general-purpose RC within an application setting. This may be of interest for designing reservoirs for a specific target application, where those parameters would be more or less fixed. On the other hand, for more general-purpose RC applications, the search for local optima within a more constrained parameter space gives a better idea of the usability of the design, while giving a fairly accurate picture of the information processing capabilities of the system.

Considering the above, we propose a novel architecture based on an asymmetric Mach-Zehnder interferometer (MZI) for TDRC with only one tunable parameter: a phase shifting element, and which leverages the intensity-conversion nonlinearity. We show that such a minimum complexity approach, i.e. using a minimal number of simple hardware components and control parameters, is sufficient for obtaining good performance on the various tasks investigated. Our approach enables GSa/s processing speeds which are only limited by the photodetector electronics, and we consider the Lithium-Niobate-on-Insulator (LNOI) platform [6, 7, 8] to leverage the low waveguide losses (< 0.2dB/cm) that enable an on-chip feedback loop, in addition to high-speed on-chip modulation (> 100GHz).

4.2 Operation principle

The integrated photonic reservoir is based on an asymmetric MZI used in a feedback configuration by means of a delay line. The asymmetric MZI is based on two 3-dB directional couplers and different arm lengths (3.0 mm and 1.5 mm), as shown in Fig. 4.1. The top ports of the MZI are connected to each other by a spiral waveguide of length 4.55 cm, which introduces a delayed feedback and thus short-term memory to the system. A phase shifting element on the bottom MZI arm controls both the feedback phase and feedback strength in this configuration, thereby essentially tuning the memory (without coupling optical power out of the system in the process as with using optical attenuators), thanks to the coupling modulation scheme [9].

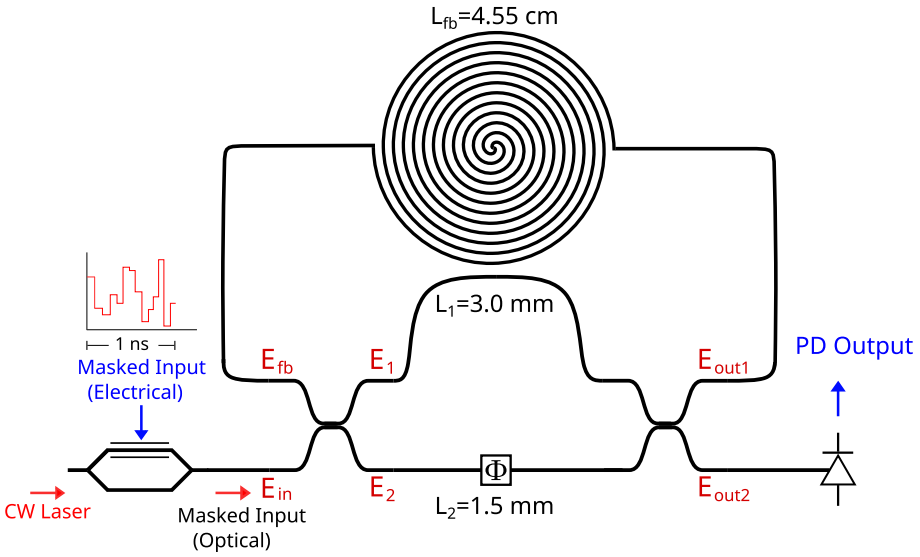

Fig. 4.1: Proposed architecture: a CW laser is modulated by the electrical input using a Mach-Zender modulator, the reservoir layer consists of the asymmetric MZI connected onto itself with a spiral waveguide and a photodetector, which also performs the readout.

Considering an input optical field $E_{in}(t) = A_{in} \exp(i\omega t)$ with amplitude A_{in} and $\omega = 2\pi c/\lambda_0$, where λ_0 is the source wavelength, it enters the first coupler at $t = 0$, and considering the 3-dB couplers as point couplers, we can describe the fields' evolution in time everywhere in the system using the scattering matrix approach (see Fig. 4.1):

$$\begin{pmatrix} E_1(t) \\ E_2(t) \end{pmatrix} = \sqrt{\alpha_c} \begin{pmatrix} -i\kappa & r \\ r & -i\kappa \end{pmatrix} \begin{pmatrix} E_{\text{in}}(t) \\ E_{\text{fb}}(t) \end{pmatrix} \tag{4.1}$$

$$\begin{pmatrix} E_{\text{out1}}(t) \\ E_{\text{out2}}(t) \end{pmatrix} = \sqrt{\alpha_c} \begin{pmatrix} \sqrt{\alpha_1} r e^{-i\beta L_1} & -\sqrt{\alpha_2} i\kappa e^{-i(\beta L_2 + \Phi)} \\ -\sqrt{\alpha_1} i\kappa e^{-i\beta L_1} & \sqrt{\alpha_2} r e^{-i(\beta L_2 + \Phi)} \end{pmatrix} \begin{pmatrix} E_1(t - \tau_1) \\ E_2(t - \tau_2) \end{pmatrix} \tag{4.2}$$

$$E_{\text{fb}}(t) = \sqrt{\alpha_{\text{fb}}} E_{\text{out1}}(t - \tau_{\text{fb}}) \exp(-i\beta L_{\text{fb}}) \tag{4.3}$$

where α_c is the fraction of optical power exiting from the coupler (considered equal for both ports), $\alpha_{1,2,fb} = 10^{-AL/10}$ are the overall fractions of power after waveguide propagation for a loss factor A [dB/m] and the respective waveguide lengths $L_{1,2,fb}$ [m], which are the lengths of the upper MZI arm, bottom MZI arm, and the feedback loop, respectively, κ and r are the cross and through field coupling coefficients, respectively, $\beta = 2\pi n_{eff}/\lambda_0$ [m^{-1}] is the propagation constant of the guided mode with effective refractive index n_{eff}, $\tau_{1,2,fb}$ [s] are the delay times of the upper MZI arm, bottom MZI arm, and delay line, respectively, and Φ [rad] is the applied phase shift on the bottom arm. The $-i$ in front of κ results from the $\pi/2$ phase shift

encountered when the field is crossing in the coupler.

The choice of the spiral waveguide length L_{fb} is important for enabling the desired maximum memory of the system. Normally, the feedback length would be constrained by the desired operation speed through one of two techniques: (i) Matching the sample hold duration with the delay time of the feedback, (ii) using a slightly longer sample hold duration than the delay time of the feedback. The first case is useful when the temporal distance between the 'virtual' nodes is smaller than the timescale of the nonlinearity such as the electronic implementation in [10]. Indeed, this creates a forward coupling of these nodes, in addition to remembering their previous states, by equating the delay time to the bit period. On the other hand, when the timescale of the nonlinearity is too small such that it can be considered instantaneous in the system [11, 12], using (i) will result in the nodes remembering only their own previous states (provided that the system does not reach steady state during one bit period) and so they become completely disconnected from each other. This can be alleviated by mismatching the sample hold duration with respect to the delay line (ii), which allows the nodes to remember the previous state of their neighboring node instead of their own (for a de-sychronization time of one node distance). However, as discussed in [13], it is not necessary for the delay time to be constrained by these two regimes for a variety of applications, especially those that do not require a large short-term memory. These constraints are due to considering the network equivalents of the TDRC scheme.

In fact, the memory capacity, discussed later in section 4.3.1, is significantly affected by the ratio of delay time to the sample hold duration. In [14] it has been shown that resonance between the delay time and the sample hold duration can especially be detrimental to the memory of the system for some tasks, specifically when they are integer multiples of each other. However, the specific components of the total metric are affected differently and thus the detriment in performance is task-dependent. Considering the above, we leverage these insights to design the delay line of our system for good performance and to reduce overall footprint. At an input sample rate $B = 1$ Gbit/s (i.e. sample hold time $\tau_B = 1$ ns), the equivalent length is $L_T = c\tau_B/n_g \approx 13.24$ cm. While the study in [13] uses the opto-electronic model as their basis, their findings show the impact of the ratio of the delay time to the input clock cycle on the memory capacity, which are also applicable in our case, the feedback phase is further considered in our case of optical feedback. Therefore we have done a few sweeps around $L_T/L_{fb} = 3$ while avoiding the resonant condition of having the exact integer value. Our design choice of $L_T/L_{fb} = 2.91$ thus reflects a region where indeed the memory capacity has been reduced below its maximum,

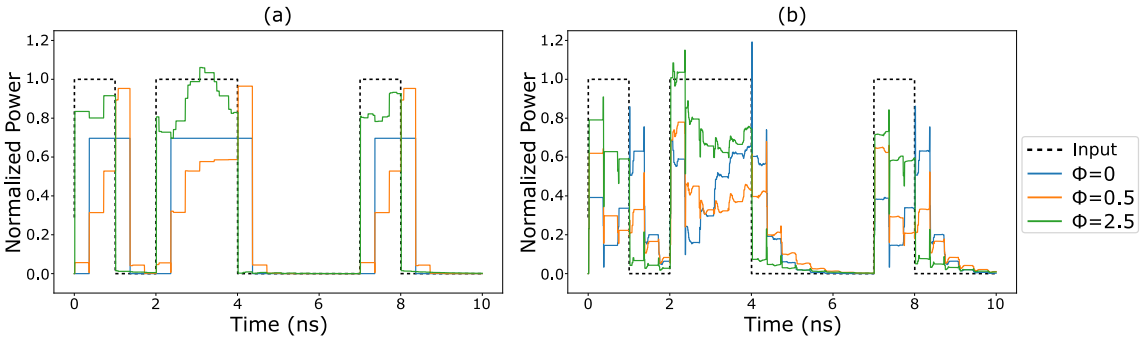

Fig. 4.2: Simulated dynamical response of the system subjected to different applied phase shifts Φ[rad] to an input bitstream with (a) symmetric MZI, (b) asymmetric MZI [1].

but is still enough for performing the nonlinear tasks presented here, while saving around 3× on footprint.

The second point to consider in this architecture is the choice of an asymmetric MZI as opposed to a symmetric one, where $\tau_1 = \tau_2$. As shown in Fig. 4.2 (b), the dynamics of the system are more interesting than that of the symmetric case in Fig. 4.2 (a) due to the different number of delays introduced in the system. Using an asymmetric MZI, there is one additional delay which enriches the dynamics further and this temporal mismatch enables the system to provide a more complex spectro-temporal response. These rich dynamics correspond to a more interesting mapping of input to output, thereby allowing the reservoir to solve highly nonlinear tasks more effectively.

4.3 Metrics and Benchmarks

The inputs of the various tasks, described in sections 4.3.1 – 4.3.5, are fed one at a time to the simulated system and its response is recorded and then trained on the various tasks using linear regression. All the simulations were carried out with the model presented in 4.2 and with an open source S-matrix based photonic circuit solver [15] to validate our analytical model. In this study, we consider the operation of the phase shifting element up to V_π and divide the interval into 101 points, constituting our applied phase values, to obtain an accurate view of the trend between the reservoir's predicted results and the applied phase. Masks were applied to the inputs for all the benchmark tasks presented here, with values drawn pseudo-randomly from a uniform distribution on the interval (0,1] corresponding to the number of 'virtual' nodes in the reservoir N_v. In this way, each mask value can be viewed as the input weight connecting the input layer to its corresponding N^{th} node. For all the tasks, we passed the photonic circuit's response (E_{out2} in Eq. 4.2)

Parameter	Designation	Value
Input sample rate	B	1 GSa/s
Simulation timestep	Δt	10 ps
Laser wavelength	λ_0	1550 nm
Waveguide effective index	n_{eff}	2.2111
Waveguide group index	n_{g}	2.2637
Directional coupler gain	α_{c}	0.966
Waveguide propagation loss	A	20 dB/m
Input optical power	P_{in}	10 mW
Photodiode responsivity	r	0.7
Load resistance	R_{l}	100 Ω
Dark current	I_{d}	5 nA

Table 4.1: Simulated photonic circuit and photodetector parameters.

through photodetectors of 5 GHz, 10 GHz, 20 GHz, and 25 GHz bandwidth (corresponding to $N_{\text{v}} = 5$, $N_{\text{v}} = 10$, $N_{\text{v}} = 20$, and $N_{\text{v}} = 25$, respectively) to determine the required size of the reservoir for the various tasks presented here. Furthermore, the trained output layer was tested on different inputs wtih the results presented in 4.4. To ensure accurate circuit-level simulations, the simulation timestep Δt was chosen to be 100 times smaller than the span of one input clock cycle, which is also small enough to take into account the short delays of the MZI arms. These short delays constitute the fastest timescale in the dynamics of this system. All the simulated photodetectors were bandwidth-dependent (incorporating a 4th order butterworth filter) and exhibited Gaussian noise with variance corresponding to the different contributions to noise. All the simulation parameters are listed in Table 4.1.

The use of standard ordinary least squares regression proved sufficient for the model to generalize and predict accurately the unseen test data, because we considered almost ideal inputs and also injected noise into the data through simulating the photodetector's response. The added noise is also a form of regularization in itself, as it prevents the model from overfitting (and from depending too strongly on very few features) and increases the robustness of the model to irregular variations or outliers. For experimental verification, however, Tikhonov regularization or Bayesian regression may need to be employed, where nonidealities and outliers in the gathered data may result in an ill-posed problem when attempting to calculate the matrix pseudoinverse [16]. The following tasks, listed in no particular order of importance, are standard RC benchmark tasks. Their widespread use in the literature allows for a direct comparison of this architecture with other photonic RC schemes.

4.3.1 Linear memory capacity

The linear memory capacity is one of the fundamental tasks for RC, which aims to test the echo state property by training the reservoir to reconstruct a given input stream of values $\in [0, 1)$ drawn from an independent and identical distribution (i.i.d.) k inputs later. It was first introduced in [17] and is given by:

$$\mathrm{MC}_k = \frac{\mathrm{cov}(u[n-k], y_k(n))^2}{\mathrm{var}(u(n))var(y_k(n)))} = r_c^2(u[n-k], y_k(n)) \tag{4.4}$$

$$\mathrm{NMSE} = \frac{\langle \|y_k(n) - y_{exp}(n)\|^2 \rangle}{\langle \|(y_{exp}(n)) - \langle y_{exp}(n) \rangle\|^2 \rangle} \tag{4.5}$$

where $u(n)$ is the input at discrete timestep n, $y_{exp}(n)$ is the expected value, $y_k(n)$ is the predicted value (after training), NMSE is the normalized mean square error, and $MC_k \in$ [0,1] is the memory capacity for a k bits shift. $MC_k = 1$ corresponds to a perfect recall of the input sequence after k input samples/bits, while $MC_k = 0$ corresponds to the complete absence of any information regarding the input sample/bit k steps into the past.

A sequence of 4000 samples was constructed from an i.i.d. stream. The target sequence is a k-bits delayed copy of the input, testing the reservoir's ability to faithfully reconstruct the input sequence after k input samples. The model was trained on the first 1000 samples and then tested on the remaining 3000 samples. The performance of specific components of the memory capacity (Eq. 4.4) are investigated as it gives a better indication of the usability of the stored information, in contrast to just the amount of information stored given by the summation of all the components, given by $\mathrm{MC}_{\mathrm{total}} = \sum_{k=1}^{N} MC_k$, where N is the number of nodes. The relevance of this evaluation for tasks requiring specific memory has also been mentioned and taken into account in other works [5, 13].

4.3.2 Temporal bitwise XOR

The temporal bitwise XOR task is a nontrivial, nonlinear memory-specific task commonly used for evaluating RC performance which was first introduced in [18]. For this task, a quasi-ideal bit stream of 4000 bits was generated, where the first 1000 bits were fed to the circuit for training and the rest were used for testing. The target bit streams were constructed by applying the XOR operation on the bit stream and a k time steps shifted version of it, yielding $x(n) \oplus x[n-k]$, where $x(n)$ is the current input bit (similar to the treatment of this task in [19]). The performance (up to

$k = 4$) is evaluated with the bit error rate (BER) metric, given by

$$\text{BER} = \frac{\#\text{incorrectly classified bits}}{\text{length of test set}} \tag{4.6}$$

4.3.3 Mackey-Glass

The Mackey-Glass sequence was first used as a RC benchmark in [20], and is generated from solving the following differential equation numerically using the 4th order Runge-Kutta method:

$$\frac{dy(t)}{dt} = \frac{ay(t-\tau)}{1+y(t-\tau)^{10}} - by(t) \tag{4.7}$$

with the commonly used parameters $a = 0.2$, $b = 0.1$, $\tau = 17$, and an integration step of 0.1. The behavior resulting from these chosen parameters is fairly periodic and only slightly irregular in the sense of causing minor fluctuations for each repeated cycle. The training set was 5000 samples long and the test set consisted of another 3000 samples. The task is a one-step ahead prediction. The performance is then evaluated according to the NMSE between the target values and the predicted output.

4.3.4 Santa Fe

The Santa Fe dataset [21] comprises data points collected experimentally from a far infrared laser operating in a chaotic regime. This dataset is fairly chaotic in the vicinity of a few data points and fairly cyclic in terms of long-term dynamical behavior. The stream of 4000 data points was divided into 2000 points used for training and the other 2000 for testing. This task is also a one-step ahead prediction. The performance on the test set is then evaluated by NMSE.

4.3.5 NARMA-3

The nonlinear autoregressive moving average (NARMA) is a commonly used benchmark task for RC which mimics a randomly varying signal around a certain average value, similar to noise. It is often used in its 10th order form to test a reservoir with very large memory. Due to the smaller memory in our system, we test the performance on a 3rd order variant of this task, which would show how the system is solving a sufficiently nonlinear task, without imposing further memory requirements than the system is capable of. The discrete difference equation that produces the

NARMA-3 sequence is given by:

$$y[n] = 0.3y[n-1] + 0.05y[n-1]\sum_{i=1}^{3} y[n-i] + 1.5u[n]u[n-3] + 0.1 \tag{4.8}$$

where the input sequence u is drawn from a uniform distribution [0,0.5]. The task is to predict y[n] given u[n]. The performance on the test set is evaluated with NMSE.

4.3.6 Baseline: Asymmetric MZI

To better understand the role of the delay line and its impact on the various tasks presented here, and consequently their memory requirements, we proceed to compare the architecture presented in Fig. 4.1 with just the asymmetric MZI without the feedback spiral waveguide. To that end, we consider the same tasks mentioned above to evaluate the performance of the asymmetric MZI alone on solving them.

4.4 Results and Discussion

4.4.1 Linear Memory Capacity

The results for different k time steps shifts in Fig. 4.3 show the variation of MC_k with respect to the applied phase shift. For lower number of nodes, the effect of the phase shift is more pronounced on the memory as can be seen in Fig. 4.3(a),(b), with $k = 3$ and $k = 4$ improving significantly as the reservoir size increases (Fig. 4.3(c),(d)). We also show MC_{total} for each reservoir size in Fig. 4.4(a) and the optimal MC_k obtained for each k up to $k = 10$, as shown in Fig. 4.4. Beyond $N_v = 5$, the peak MC_k values for many k's appear to be close to each other. This suggests that further exceeding the studied number of nodes (i.e $N_v = 25$) would not enhance the memory further as it is fundamentally limited by the length of the spiral waveguide with the given input bitrate. Our results for MC_{total}, shown in Fig. 4.4 are consistent with those presented in [13], as we achieve $MC_{total} \approx 5.5$ which is close to the result obtained in the same work for our chosen ratio of delay time and input clock cycle. Thus, our MC_{total} exceeds that of the PIC implementation in [22], which requires copies of the delayed input at the modulation stage to exceed its intrinsic $MC_{total} = 1$. However, it is lower than the one presented in [19], where additional post-processing is performed to merge the responses to previous inputs with the current input. This increases the number of virtual nodes and subsequently the linear memory capacity from an intrinsic $MC_{total} \approx 6$ to $MC_{total} \approx 8$ with post-processing. Furthermore, our MC_k results for 20 nodes are almost equivalent to the simulated result in [23]

using 20 on-chip lasers from $k = 1$ to $k = 5$. However, it is lower for further k as our system was not designed for large memory for the purpose of the currently investigated tasks. This can be easily alleviated - thanks to the low losses of the LNOI platform - by utilizing a longer spiral length and using the desynchronized regime explained in section 4.2, and/or possibly using similar pre/post-processing techniques such as those described in [22, 19].

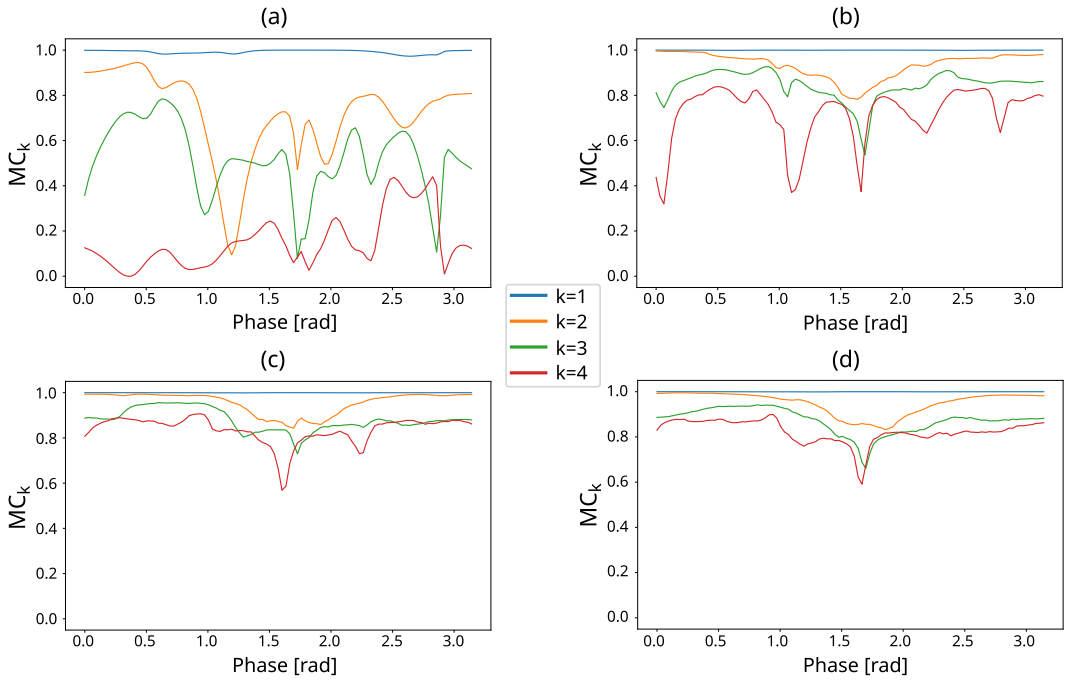

Fig. 4.3: Performance of the reservoir on solving the linear memory capacity task for different values of applied phase shift Φ for different reservoir sizes: (a) N_v=5 (b) N_v=10, (c) N_v=20, (d) N_v=25 [1].

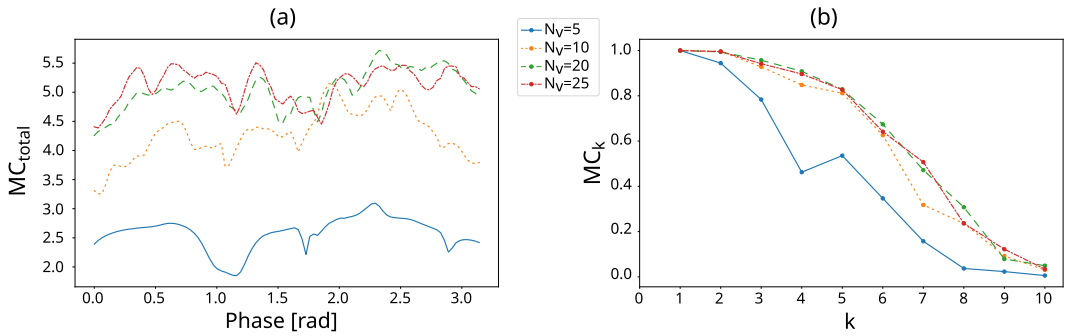

Fig. 4.4: MC_{total} (a) and peak obtained values of MC_k (b) for different reservoir sizes: N_v=5, N_v=10, N_v=20, and N_v=25 [1].

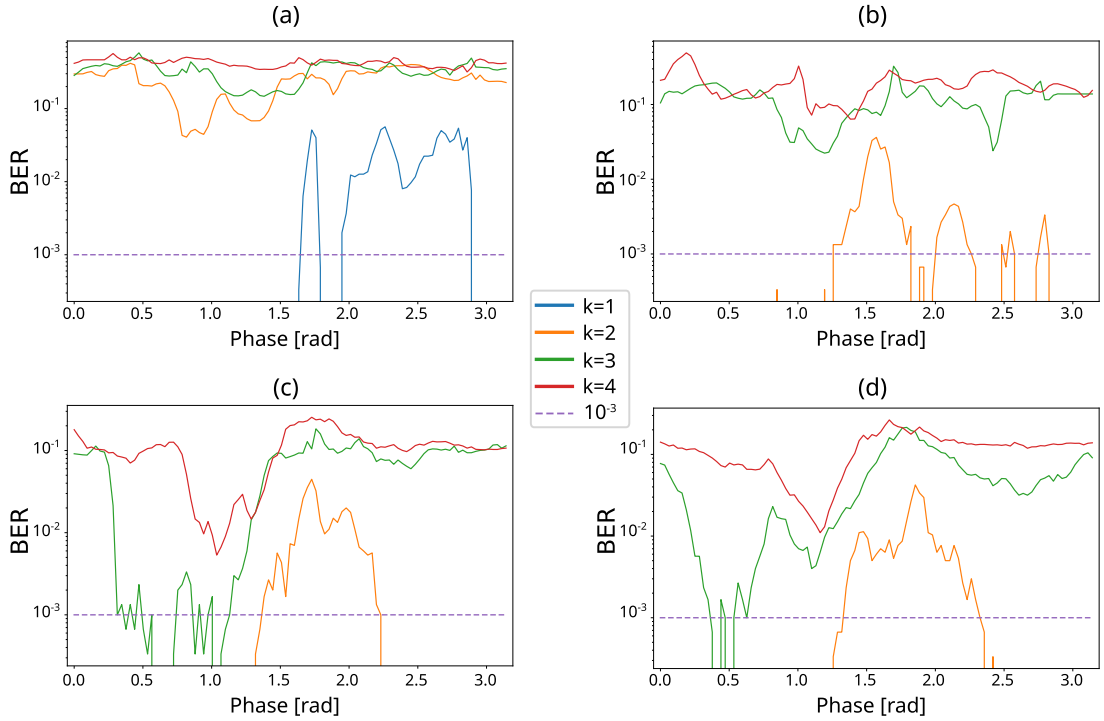

Fig. 4.5: Performance of the reservoir on solving the temporal bitwise XOR task for different values of applied phase shift Φ [rad] on the MZI arm for different reservoir sizes: (a) 5 Nodes, (b) 10 Nodes, (c) 20 Nodes, (d) 25 Nodes. When the blue line (k=1) is not visible, it is due to BER = 0 everywhere on the plot [1].

4.4.2 Temporal bitwise XOR

For the XOR task, the test sequence used consisted of 3000 bits, limiting the resolution of the BER to 3.33×10^{-4}. Thus, a BER below 10^{-3} is considered acceptable, as shown in dotted purple lines in Fig. 4.5, where results of $k = 1$ to $k = 4$ are shown for various reservoir sizes. Similar to the memory capacity results, the performance on the XOR task mostly improves as the reservoir size scales. It is shown in Fig. 4.5(a) that it is possible to perform the one bit XOR with only 5 nodes. It can be seen from Fig. 4.5(c) that the architecture can be used successfully for XOR-ing the current input bit with 3 bits into the past, for $N_v = 20$.

4.4.3 Mackey-Glass

The results in Fig. 4.6(a) are for differently sized reservoirs under applied phase shift. One of the interesting features in the curve is that the performance is only minimally affected by the number of nodes N_v presented here, which suggests that only a small memory is required for this task. We obtain NMSE = 0.0056 which is close to the optimum value obtained in [5], even when compared to a bulk setup implementation [24].

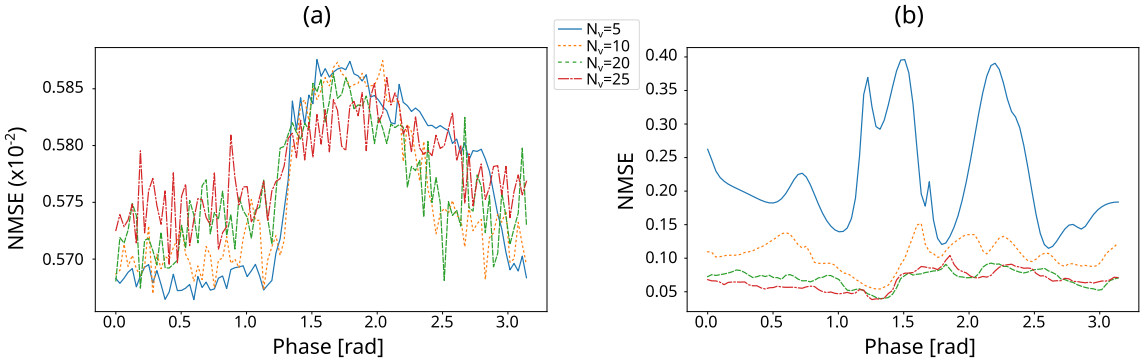

Fig. 4.6: Performance of the reservoir (NMSE) under applied phase shift on one-step ahead prediction time series tasks: (a) Mackey-Glass, and (b) Santa Fe [1].

Another interesting point this result shows is the minimal dependence on the varying phase shift, which prompts further investigation into which part of the architecture is responsible for the obtained NMSE performance. It was found that equivalent performance is obtainable by training the input data with linear regression, without going through the photonic circuit. This is discussed later in section 4.4.6.

4.4.4 Santa Fe

For the Santa Fe time series prediction, the results in Fig. 4.6(b) show a minimum NMSE of 0.038 using 25 nodes, which is close to the simulated result in the nonlinear microring approach in [5] (NMSE=0.038) and better than the experimental result in the multiple cavities approach based on a feed-forward photonic neural network in feedback reported in [25] (NMSE=0.06). It is also slightly better than the experimental result mentioned in [19] (NMSE=0.049), which was achieved by increasing the laser pump current and the semiconductor optical amplifier (SOA) current, using 23 virtual nodes, albeit with additional postprocessing techniques. Furthermore, we also obtain better prediction results than the approach in [23] where they reported a minimum NMSE$\approx$ 0.01 using 40 on-chip lasers with small external cavities of 10 mm.

4.4.5 NARMA-3

For the NARMA-3 task, beyond $N_v = 5$, the results show only a slight dependence on the number of nodes for all values of phase shift. This is especially the case around $\Phi = 0.5$ rad. However, the results are much more strongly influenced by the phase shift range, especially from $\Phi = 1$ rad to $\Phi = 2.5$ rad. This indicates that the memory and nonlinearity of the system are being strongly altered by the different interference conditions at the MZI, which produces sufficient changes in

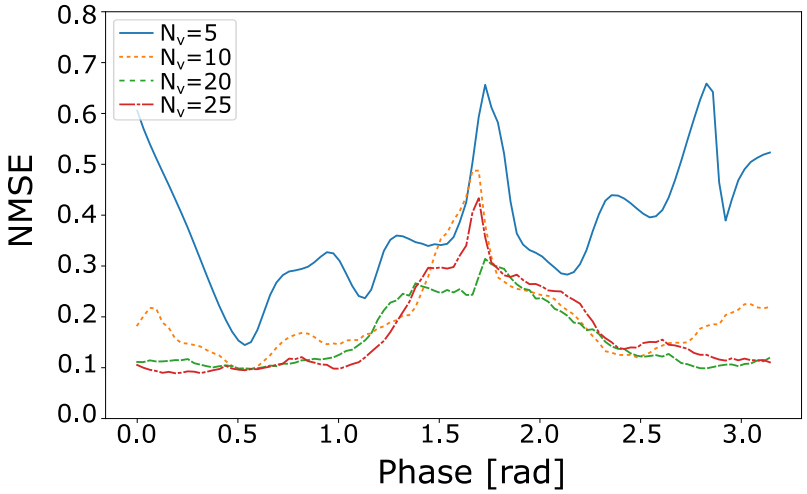

Fig. 4.7: Performance on the NARMA-3 task for various N_{v} [1].

the output dynamics to yield excellent (e.g. $\Phi \approx 0.5$ rad) or poor (e.g. $\Phi \approx 1.7$ rad) performance. A low NMSE= 0.096 is obtained using as few as 10 nodes.

4.4.6 Baseline: Asymmetric MZI

Performing the same numerical investigations on the MZI alone without the feedback loop helps in understanding the delay's role further. For our operation speed of 1 GSa/s, it can be seen that tasks requiring a memory of one sample/bit into the past are achievable, which is not surprising since at some point the current sample interacts with the previous input sample due to the differences between the arm lengths, and consequently the asymmetric MZI alone is sufficient. Such tasks are the memory capacity and XOR tasks for $k = 1$, where $MC_1 \approx 1.0$ everywhere for all phase shifts and for all reservoir sizes N_v. The XOR operation for $k = 1$ is successful beyond a certain value of phase shift, due to destructive interference at this value of applied phase (Fig. 4.8 (a), a similar behavior is seen in (b) as well). For tasks requiring deeper memories, the MZI alone fails completely: $MC_{k>1} \approx 0$ for all N and phase shift, BER$\approx$ 0.5 XOR for $k > 1$. For the Santa Fe task, the performance degrades considerably as shown in Fig 4.8 (b) with NMSE$\approx$ 0.34 being the best value achieved. For the NARMA-3 task (Fig. 4.8 (c)), it fails completely with NMSE $\approx$ 0.7 everywhere on the plot.

For the Mackey-Glass one-step ahead prediction task, we get equivalent performance (NMSE = 0.00587) with the MZI alone as shown in Fig. 4.8 (d), and in fact it is also similar to the performance obtained when training the input data itself (masked and unmasked) using linear regression, where we also found no degradation in NMSE for all N_v considered (NMSE = 0.00583). We find this result particularly interesting,

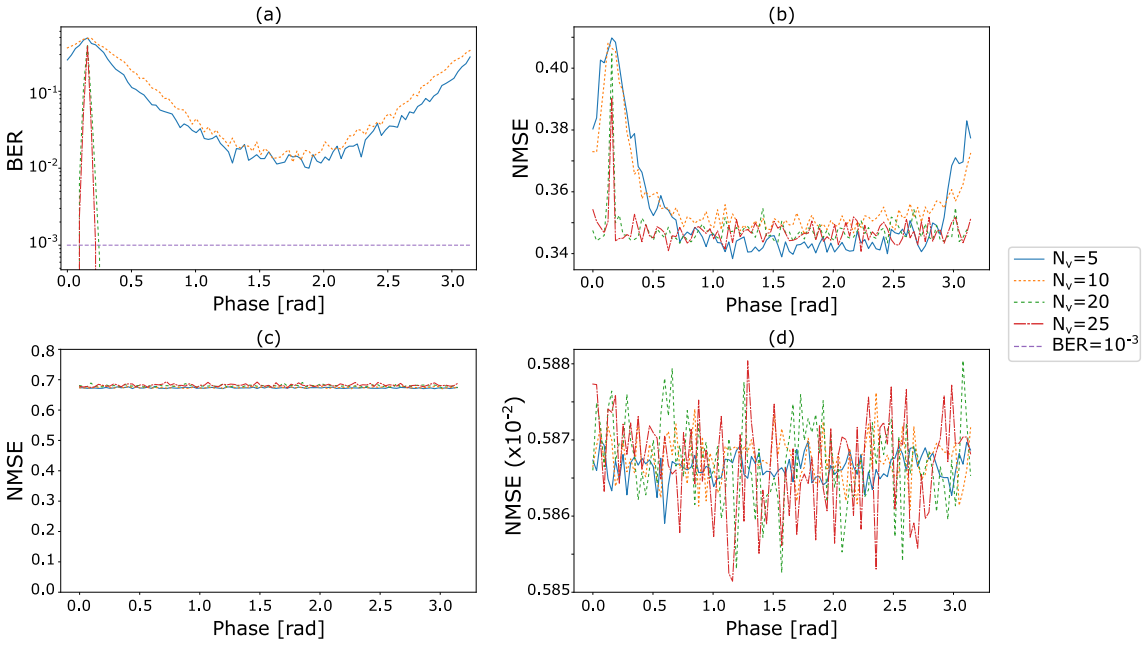

Fig. 4.8: Performance of the different tasks using only the asymmetric MZI under varying phase shift: (a) XOR for k=1, (b) Santa Fe, (c) NARMA-3, (d) Mackey-Glass [1].

since it shows that solving the one-step ahead task can be done with 5 trainable features and using a regression on the input data itself. Considering two and three steps ahead predictions on the same task, the full reservoir architecture only slightly outperforms ($\text{NMSE}_{k=2} = 0.0114$, $\text{NMSE}_{k=3} = 0.0170$) the almost equivalent result of training on both the input dataset directly, and passing it through just the MZI ($\text{NMSE}_{k=2} = 0.0119$, $\text{NMSE}_{k=3} = 0.0182$). According to these results, the one-step ahead Mackey-Glass task can be linearly separable, at least in the considered portion of the dataset used for training and validation.

4.4.7 Further discussion

For all the tasks presented here, a reservoir size of $N_\text{v} = 20$, with a 20 GHz photodetector, is sufficient for obtaining the best performance out of this architecture. Furthermore, the variation in prediction accuracy (under applied phase shift) for the several tasks presented are strongly related to the phases of the signals travelling into the spiral from the two MZI arms. To further explain this notion, we refer back to the memory capacity results in section 4.4.1. According to the value of Φ, interference occurs at the input and output couplers. Due to the low losses, multiple round trips can take place within both paths, yielding either constructive or destructive interference over one or multiple round trips. This directly influences the virtual nodes' connectivity matrix, and has a much stronger effect for lower number

of nodes such as $N_v = 5$, as can be seen from the larger variability in Fig. 4.3(a). Increasing the number of nodes allows more information to be captured each round trip, which especially improves the performance on tasks requiring deeper memories. The memory eventually saturates when there is no longer any representation of the information inside the system for further past inputs (Fig. 4.3(c) and (d)). Naturally, this behavior is also pronounced in other tasks (Fig. 4.6(b) and Fig. 4.7). In addition, there are multiple advantages for using the proposed RC scheme:

- First, the **fully integrated low-loss delay line**, which is made possible by considering low-loss platforms such as LNOI, which entails less power loss coupling into and out of the chip, similar to [19], and in contrast to [5] when using an external feedback.

- Second, the **high-speed operation**, limited only by the photodetector bandwidth, whereas other architectures employing relatively slow nonlinearities (especially thermal nonlinearities in the case of silicon-on-insulator) can significantly lower computation speeds [22].

- Third, the **exploiting of multiple timescales** in this RC scheme for response richness, which was also previously leveraged in [25]. However, our architecture reduces complexity by minimizing the number of phase shifters, while also obtaining better results on the Santa Fe task (compared to [25]).

Taking into consideration some PIC RC implementations from the literature, we see that compared to [23], which uses up to 40 on-chip lasers, the memory MC_k of our system is close to the one obtained by them using 20 on-chip lasers in the range of $k = 1$ to $k = 5$ (Fig. 4.4 (b)). In [26], the similarly passive architecture – which also relies only on the photodetector nonlinearity – is limited in scalability by the use of physical nodes instead of virtual ones, and the need to change the ratio of interconnection delay and bit period for solving different tasks. Since the former is fixed, this entails changing the input bitrate for tasks requiring different memories, such as the bitwise XOR with multiple bits in the past. This is often not practical or even possible in certain scenarios. However, in our case, as shown in section 4.4.2, only a single phase shifting element is required. In fact, it is even possible to do the XOR for $k = 1$ to $k = 3$ at the same value of phase shift, thus requiring only a change in the applied output weights to perform the three different tasks. The architecture in [19], consisting of a distributed Bragg reflector laser and amplifiers, as well as integrated delay lines, achieved similar performance on the Santa Fe task after additional post-processing.

On the other hand, it is also important to consider that simulation setups and learning algorithms are different between different investigations; for example, most often ridge regression is employed instead of linear regression, and the ridge parameter is usually scanned and optimized for the unique requirements of each scheme and setup. It is therefore not straightforward to compare these different examples from the literature, which is why a rigorously fair comparison is beyond the scope of this work. Instead, we aim to shed light on the fact that RC with matching performance to the above examples can be done on-chip with mostly passive components, without the need for nonlinearities beyond that of the photodetector, with minimum active components (no optical amplifiers or multiple laser sources), and with only one phase shifter controlling two tunable parameters (feedback strength and phase). Using a single phase shifter that is relatively easy to control – as opposed to scanning and optimizing for a multidimensional parameter space – can enable on-chip stabilization using optical feedback techniques [27], which could potentially enable photonic RC that is robust to ambient fluctuations. Further adding to the system complexity may indeed boost the system performance beyond simpler architectures such as that proposed in this work. We believe this work could therefore serve as a baseline in terms of performance for the given system and hardware requirements, and that future works could enable performance improvements that warrant the use of higher complexity PIC RC schemes.

4.5 Conclusion

We have proposed an integrated photonic architecture for RC which leverages the low losses of the LNOI platform to enable a fully integrated delay line and with only one tunable parameter to tune the feedback phase and the feedback strength simultaneously. The delay line was designed to be compact enough while still delivering equivalent performance to other PIC implementations for a variety of tasks. Further enhancement of the memory is possible by increasing the length of the spiral waveguide, at the cost of footprint. This approach also provides more efficient utilization of the optical power and of the information stored inside the reservoir layer. This is especially the case when compared to other photonic implementations requiring an optical attenuator block in the feedback loop to tune the feedback strength, where light is simply coupled out of the system. We conclude that minimum complexity RC can also open the doors towards robust RC in ambient conditions by only requiring the stabilization of one parameter, thereby increasing the longevity of each training cycle and possibly allowing the deployment of photonic RC in real-world settings and applications. Lastly, we believe this work can also serve as a baseline

to be compared against more complex photonic RC systems, since there could indeed be room for performance improvement through using more interesting (and complex) hardware schemes or physical phenomena.
In the next chapter, a more detailed investigation involving a design space exploration will be presented, while proposing novel approaches to bypassing the input mask.

Bibliography

[1] M. Abdalla, C. Zrounba, R. Cardoso, P. Jimenez, G. Ren, A. Boes, A. Mitchell, A. Bosio, I. O'Connor, and F. Pavanello, "Minimum complexity integrated photonic architecture for delay-based reservoir computing," *Opt. Express*, vol. 31, no. 7, pp. 11 610–11 623, Mar 2023. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-7-11610

[2] E. Gooskens, F. Laporte, C. Ma, S. Sackesyn, J. Dambre, and P. Bienstman, "Wavelength dimension in waveguide-based photonic reservoir computing," *Opt. Express*, vol. 30, no. 9, pp. 15 634–15 647, Apr 2022.

[3] I. Bauwens, K. Harkhoe, P. Bienstman, G. Verschaffelt, and G. V. der Sande, "Transfer learning for photonic delay-based reservoir computing to compensate parameter drift," *Nanophotonics*, 2022. [Online]. Available: https://doi.org/10.1515/nanoph-2022-0399

[4] X. X. Guo, S. Y. Xiang, Y. H. Zhang, L. Lin, A. J. Wen, and Y. Hao, "Four-channels reservoir computing based on polarization dynamics in mutually coupled vcsels system," *Opt. Express*, vol. 27, no. 16, pp. 23 293–23 306, Aug 2019. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-16-23293

[5] G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Optics Express*, vol. 30, p. 522, 1 2022.

[6] A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (lnoi) for photonic integrated circuits," *Laser and Photonics Reviews*, vol. 12, 4 2018.

[7] X. Han, L. Chen, Y. Jiang, A. Frigg, H. Xiao, T. G. Nguyen, A. Boes, J. Yang, G. Ren, Y. Su, A. Mitchell, and Y. Tian, "Integrated subwavelength gratings

on a lithium niobate on insulator platform for mode and polarization manipulation," *Laser & Photonics Reviews*, p. 2200130, 2022.

[8] X. Han, Y. Jiang, A. Frigg, H. Xiao, P. Zhang, T. G. Nguyen, A. Boes, J. Yang, G. Ren, Y. Su, A. Mitchell, and Y. Tian, "Mode and polarization-division multiplexing based on silicon nitride loaded lithium niobate on insulator platform," *Laser & Photonics Reviews*, vol. 16, no. 1, p. 2270001, 2022.

[9] W. D. Sacher, W. M. J. Green, S. Assefa, T. Barwicz, H. Pan, S. M. Shank, Y. A. Vlasov, and J. K. S. Poon, "Coupling modulation of microrings at rates beyond the linewidth limit," *Opt. Express*, vol. 21, no. 8, pp. 9722–9733, Apr 2013.

[10] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature Communications*, vol. 2, no. 1, p. 468, Sep 2011.

[11] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Scientific Reports*, vol. 2, 2012.

[12] Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, no. 5, pp. 438–446, May 2015.

[13] T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, "Role of delay-times in delay-based photonic reservoir computing," *Opt. Mater. Express*, vol. 12, no. 3, pp. 1214–1231, Mar 2022.

[14] F. Köster, D. Ehlert, and K. Lüdge, "Limitations of the recall capabilities in delay-based reservoir computing systems," *Cognitive Computation*, Oct 2020.

[15] F. Laporte, J. Dambre, and P. Bienstman, "Highly parallel simulation and optimization of photonic circuits in time and frequency domain based on the deep-learning framework pytorch," *Scientific Reports*, vol. 9, 12 2019.

[16] D. Li, M. Han, and J. Wang, "Chaotic time series prediction based on a novel robust echo state network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 5, pp. 787–799, 2012.

[17] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach," 2002.

[18] N. Bertschinger and T. Natschläger, "Real-Time Computation at the Edge of Chaos in Recurrent Neural Networks," *Neural Computation*, vol. 16, no. 7, pp. 1413–1436, 07 2004.

[19] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Optics Express*, vol. 28, p. 3086, 2 2020.

[20] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1091277

[21] A. Weigend and N. Gershenfeld, "Results of the time series prediction competition at the santa fe institute," in *IEEE International Conference on Neural Networks*, 1993, pp. 1786–1793 vol.3.

[22] M. Borghi, S. Biasi, and L. Pavesi, "Reservoir computing based on a silicon microring and time multiplexing for binary and analog operations," *Scientific Reports*, vol. 11, no. 1, p. 15642, Aug 2021. [Online]. Available: https://doi.org/10.1038/s41598-021-94952-5

[23] C. Sugano, K. Kanno, and A. Uchida, "Reservoir computing using multiple lasers with feedback on a photonic integrated circuit," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. PP, pp. 1–1, 07 2019.

[24] J. Bueno, D. Brunner, M. C. Soriano, and I. Fischer, "Conditions for reservoir computing performance using semiconductor lasers with delayed optical feedback," *Optics Express*, vol. 25, p. 2401, 2 2017.

[25] M. Nakajima, K. Tanaka, and T. Hashimoto, "Scalable reservoir computing on coherent linear photonic processor," *Communications Physics*, vol. 4, no. 1, p. 20, Feb 2021.

[26] K. Vandoorne, P. Mechet, T. V. Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Communications*, vol. 5, 3 2014.

[27] C. Sun, M. Wade, M. Georgas, S. Lin, L. Alloatti, B. Moss, R. Kumar, A. H. Atabaki, F. Pavanello, J. M. Shainline, J. S. Orcutt, R. J. Ram, M. Popović, and V. Stojanović, "A 45 nm cmos-soi monolithic photonics platform with bit-statistics-based resonant microring thermal tuning," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 893–907, 2016.

Chapter 5

Mask-less Photonic Time-Delay Reservoir Computing

In this chapter, we are going to focus on how we can optimize the design of the minimum complexity reservoir, discussed in Chapter 4, to achieve optimal performance on desired tasks. Furthermore, we look into opportunities to bypass the input masking protocol, which can constrain the input layer to costly and energy-consuming auxiliary electronics, in addition to possible domain conversions. Additionally, the correlation between the performance on multiple tasks is investigated, which can be leveraged to design a reservoir that can handle multiple tasks with the same input signal. The reported results in this chapter address **RQ2, RQ3, and RQ4**, and have been published in a journal article [1].

5.1 Motivation

Recently, some interest has been directed towards the exploration of the asynchronous regime in RC [2, 3, 4], where the delay time of the feedback loop is unconventionally chosen to be neither equivalent nor close to that of the input rate. The synchronicity convention stemmed from the initial approach of viewing TDRC systems as the network equivalent of spatially multiplexed networks, where the nodes are connected in a cyclic fashion [5]. Another conventional preference has been the use of an input masking protocol to prolong a reservoir's activity. This approach generally improves performance, but introduces complications in practical cases as well as associated power costs, in addition to a bottleneck for ultra-high speed signal processing [6]. In this study, we tackle this problem by dispensing with the input mask altogether, which allows the bypassing of an optical/electronic/optical conversion at the input, in case the signal to be fed to the reservoir is in the optical domain

(for example in telecom applications for in-network computing [7]). We use task-independent metrics alongside standard benchmark tasks to judge the performance of an all-optical integrated reservoir over its design space, with a specific focus on the feedback time and applied phase shift. The considered photonic reservoir has an ultrashort timescale, which is only limited by the RF bandwidth of the readout electronics [8]. We also take into account the associated effects of the electronic readout in our study, mainly the photodiode noise and filtering with subsequent 12-bit quantization, to understand its implications on the performance. Lastly, we show the correlation between the performance on benchmark tasks and the different task-independent metrics within the explored design space.

5.2 Methods

The performance of an RC scheme can be quantified in a number of ways; the most obvious is by using standard benchmark tasks. However, this constrains the evaluation to these specific tasks and does not provide information about the reservoir's overall memory/nonlinearity capabilities. To give a more comprehensive account of the RC performance, task-independent metrics must be employed as well. The information processing capacity (IPC) [9] is a task-independent metric that has been used to quantify the performance of reservoirs in a number of dedicated studies [10, 11, 12, 13]. However, when considering a large design space (as in our case), IPC calculation would require large computational time and resources. Instead, we opted for a combination of other metrics which can reflect the memory/nonlinearity abilities of a reservoir, while requiring less computing resources. The combined use of task-independent metrics and benchmark tasks aim to quantify the reservoir's performance such that the different instances can be compared to one another, and also compared to some of the proposed schemes in the literature.

5.2.1 Task-independent tests

The first metric considered is the linear memory capacity [14], which showcases the ability to reconstruct past inputs using a weighted linear combination of the reservoir's presently probed states. An input sequence u of length 2000, drawn from a uniform distribution $\in [0, 0.5)$, is used to train the model with linear regression to reconstruct k inputs into the past. The performance is then evaluated with the

square of the Pearson correlation coefficient r_c^2:

$$\mathrm{MC}_k = \frac{\mathrm{cov}(u[n-k], y_k[n])^2}{\mathrm{var}(u[n])var(y_k[n]))} = r_c^2(u[n-k], y_k[n]) \tag{5.1}$$

where cov(.) is the covariance, var(.) is the variance, $u[n]$ is the input at the current step, $u[n-k]$ is the input delayed by k discrete steps, and $y_k[n]$ is the predicted stream (after training). For this study, we calculate the total linear memory capacity $\mathrm{MC}_{\mathrm{tot}}$ by summing up the first $m = 100$ terms. Theoretically speaking, the sum includes the infinite past, i.e. all the $m \in [1, \infty)$ to be considered. A similar truncation strategy to [15] is used, where we do not consider the terms where $\mathrm{MC}_k < 0.01$.

$$\mathrm{MC}_{\mathrm{tot}} = \sum_{k=1}^{m} \mathrm{MC}_k \tag{5.2}$$

The second metric is the computational ability (CA), which is a measure of the RC nonlinearity strength or computation power. It is calculated by performing two tests on the reservoir: the kernel quality and the generalization tests [16]. The kernel quality test evaluates how the reservoir can map different inputs to sufficiently different states. For this test, we construct 100 input streams, with each being a sequence of length 100, drawn from an independent and identical distribution with values $\in [0, 0.5)$. All 20 nodes corresponding to the 100th input are then collected to construct a 100×20 matrix (#input streams$\times N$). The generalization test reflects how well the reservoir can map the same input samples under different initial conditions. A useful schematic which explains the procedures of this test can be found in [17]. Using the previously generated 100 different input streams, we concatenate to each an identical sequence of 10 inputs, and we collect the node responses corresponding to the 110th output sample to construct a 100×20 matrix. Using singular value decomposition with the appropriate threshold [18], the ranks of their corresponding matrices are then found. The normalized CA is then calculated:

$$\mathrm{CA} = (\mathrm{KQR} - \mathrm{GR})/N_{\mathrm{v}} \tag{5.3}$$

where KQR is the kernel quality matrix rank (from the kernel quality test), GR is the generalization matrix rank (from the generalization test), and N_{v} is the number of nodes in the reservoir. Thus, high KQR and low GR are desirable.

5.2.2 Benchmark tasks

Each realization over the entire design space was trained on 3 benchmark tasks: NARMA-k, XOR-k, and the Santa Fe dataset [19]. The NARMA-k task is given by:

$$y[n] = 0.3y[n-1] + 0.05y[n-1]\sum_{i=1}^{k} y[n-i] + 1.5u[n]u[n-k] + 0.1 \tag{5.4}$$

where the input sequence u is drawn from a uniform distribution $\in [0, 0.5)$. The task is to predict $y[n]$ given $u[n]$. The performance on these tasks is evaluated by the normalized mean square error (NMSE), which is given by:

$$\text{NMSE} = \frac{1}{n\sigma_{y_\text{t}}^2}\sum_{i=1}^{n}(y_\text{t}^{(i)} - y_\text{p}^{(i)})^2 \tag{5.5}$$

where y_t is the target (true) value and y_p is the predicted value, n is the number of data points, and $\sigma_{y_\text{t}}^2$ is the variance of the sequence of true values. For the XOR task, the commonly used bit error rate (BER) was also evaluated, which is given by:

$$\text{BER} = \frac{\#\text{incorrectly classified bits}}{\text{length of test set}} \tag{5.6}$$

NMSE and BER were already defined in chapter 4. They are reproduced here for convenience, using the same notation as the rest of this chapter. Training for all the benchmark tasks was done using ridge regression, with the regularization term $\alpha = 10^{-4}$ to prevent overfitting. For the linear memory capacity, linear regression was employed since this task does not require a test set. For the NARMA-k task, an input sequence of length 3500 was used for training and 500 for testing. For the XOR=k task, 1000 bits were used for training and 3000 for testing. For the Santa Fe task, the training sequence was 3000 samples, while 1000 were used for testing.

5.2.3 Simulation setup

The considered photonic circuit, shown in Fig. 5.1, consists of passive waveguides, 3-dB directional couplers, and a heater to tune the system's output dynamics. This circuit can be fabricated on a low-loss integrated photonic platform, such as Si_3N_4 [20] or LNOI [21]. The circuit is modeled using the transfer matrix approach and numerically solved with the following delayed coupled equations, which describe the

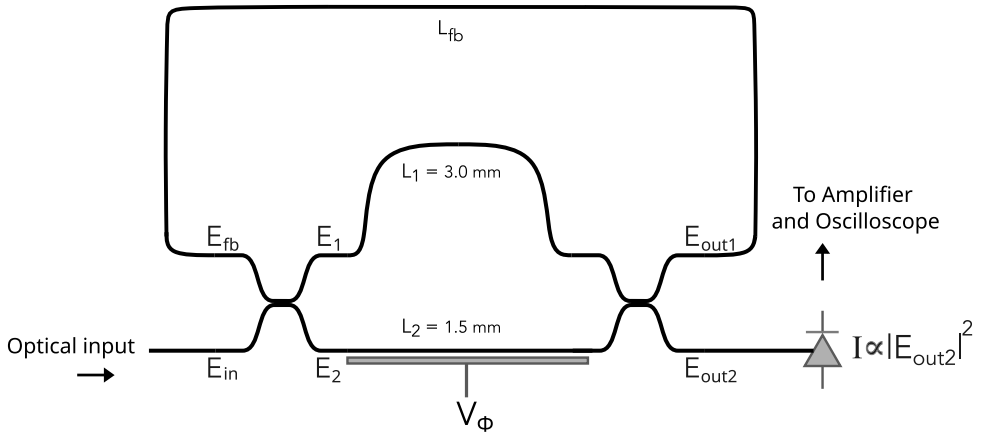

Fig. 5.1: Passive photonic reservoir. Depending on the desired L_{fb}, the feedback waveguide is either straight, with bends, or in a spiral configuration [1].

temporal evolution of the electric field everywhere in the system (Fig. 5.1).

$$\begin{bmatrix} E_1(t) \\ E_2(t) \end{bmatrix} = \sqrt{\gamma_c} \begin{bmatrix} -i\kappa & r \\ r & -i\kappa \end{bmatrix} \begin{bmatrix} E_{\text{in}}(t) \\ E_{\text{fb}}(t) \end{bmatrix} \tag{5.7}$$

$$\begin{bmatrix} E_{\text{out1}}(t) \\ E_{\text{out2}}(t) \end{bmatrix} = \sqrt{\gamma_c} \begin{bmatrix} \sqrt{\gamma_1} r e^{-i\beta L_1} & -\sqrt{\gamma_2} i\kappa e^{-i(\beta L_2+\Phi)} \\ -\sqrt{\gamma_1} i\kappa e^{-i\beta L_1} & \sqrt{\gamma_2} r e^{-i(\beta L_2+\Phi)} \end{bmatrix} \begin{bmatrix} E_1(t-\tau_1) \\ E_2(t-\tau_2) \end{bmatrix} \tag{5.8}$$

$$E_{\text{fb}}(t) = \sqrt{\gamma_{\text{fb}}} E_{\text{out1}}(t-\tau_{\text{fb}}) \exp\left(-i\beta L_{\text{fb}}\right) \tag{5.9}$$

Eqs. 5.7-5.9 show the circuit model, where γ_c is the percentage of optical power exiting from the coupler, $\gamma_{1,2,\text{fb}} = 10^{-AL/10}$ are the fractions of power after waveguide propagation for a loss factor A [dB/m] and waveguide lengths $L_{1,2,\text{fb}}$ [m], corresponding to the lengths of the upper MZI arm, bottom MZI arm, and the feedback loop, respectively. Since we are considering 50/50 couplers, the cross and through field coupling coefficients are equivalent ($\kappa = r$). For waveguide parameters, $\beta = 2\pi n_{\text{eff}}/\lambda_0$ [m^{-1}] is the propagation constant of the guided mode with effective refractive index n_{eff}, $\tau_{1,2,fb}$ [s] are the delay times of the upper MZI arm, bottom MZI arm, and feedback waveguide, respectively, and Φ [rad] is the applied phase shift on the bottom arm. The application of voltage on the heater allows the reservoir's dynamics to be changed through controlling both the feedback strength and phase by adjusting the interference at the directional couplers. For the electronic readout, the photodetector is modeled by adding the associated noise components of shot noise and thermal noise, and then low-pass-filtering the signal with the cut-off frequency as the chosen photodetector bandwidth. The signal is then converted

Table 5.1: parameters of the optical and electronic simulations

Parameter	Value	Description
τ_{fb}	scanned over [20 ps;500 ps]	Feedback time
τ	100 ps	Synchronized delay time
L_{fb}	$\tau_{\mathrm{fb}} \times c/n_{\mathrm{g}}$	Feedback length
Φ	scanned over [0;2π]	Applied phase shift
B_{m}	10 GBit/s	Input rate
f_{s}	200 GSa/s	Sampling rate (electronic)
N_{v}	$f_{\mathrm{s}}/B_{\mathrm{m}} = 20$	Number of nodes
Δt	1.6 ps	Simulation timestep (optical)
λ_0	1550 nm	Laser wavelength
n_g	1.996	Group index of Si_3N_4
γ_{c}	0.966	$P_{\mathrm{out}}/P_{\mathrm{in}}$ of directional coupler
κ	$\sqrt{0.5}$	Cross coefficient of coupler
r	$\sqrt{0.5}$	Through coefficient of coupler
$A_{(\mathrm{fb})}$	20 dB/m	(Feedback) waveguide loss
P_{in}	10 mW	Input laser power (amplitude)
r_{p}	0.8	Photodetector responsivity
I_{d}	2 nA	Photodetector dark current
R_{l}	100 Ω	Load resistance
f_{c}	scanned over [2 GHz;40 GHz]	Photodetector bandwidth
Q	12 bits	ADC bits of resolution
V_{RefHi}	1.2 V	ADC upper bound
V_{RefLo}	0.0 V	ADC lower bound

to a voltage through a transimpedence amplifier, where the only TIA-related noise considered is the one contributed by the load resistance to the thermal noise. Finally, the signal is binned to yield the final output as if it were passed through a 12-bit analog-to-digital converter (ADC), as would be found on most high-end digital oscilloscopes. The photodetector noise is modeled by:

$$\sigma_{\mathrm{th}}^2 = 4k_{\mathrm{B}}Tf_{\mathrm{c}}/R_{\mathrm{l}} \tag{5.10}$$

$$\sigma_{\mathrm{sh}}^2 = 2q(I_{\mathrm{p}} + I_{\mathrm{d}})f_{\mathrm{c}} \tag{5.11}$$

$$\sigma_{\mathrm{tot}}^2 = \sigma_{\mathrm{th}}^2 + \sigma_{\mathrm{sh}}^2 \tag{5.12}$$

$$I_{\mathrm{p}} = r_{\mathrm{p}}P_{\mathrm{opt}} \tag{5.13}$$

$$V_{\mathrm{p}} = R_{\mathrm{l}}I_{\mathrm{p}} \tag{5.14}$$

where σ_{th}^2 is the thermal noise variance, σ_{sh}^2 is the shot noise variance, σ_{tot}^2 is the total noise variance, k_{B} is the Boltzmann constant, T is the temperature, f_{c} is

the photodetector 3-dB bandwidth (cutoff-frequency), R_l is the load resistance, q is the electronic charge, I_p is the current proportional to the input optical power P_opt through responsivity r_p, I_d is the dark current, and V_p is the voltage converted through load resistance R_l. The voltage after passing through the ADC is then quantized by

$$V_\mathrm{Q} = \mathrm{floor}\left(\frac{V_\mathrm{p}}{\Delta_\mathrm{q}}\right) \times \Delta_\mathrm{q} \tag{5.15}$$

where Δ_q is the ADC quantization step, given by $\Delta_\mathrm{q} = \left(V_\mathrm{RefHi} - V_\mathrm{RefLo}\right)/(2^Q - 1)$, which depends on the considered voltage range and number of bits of resolution Q of the ADC. Further information about the simulation parameters can be found in Table 5.1.

5.3 Results and Discussion

We consider the passive all-optical integrated reservoir scheme shown in Fig. 5.1 operating with 10 mW laser power ($\lambda = 1550$ nm) at a fixed datarate of 10 GBit/s, which corresponds to a synchronized delay time $\tau = 100$ ps. The reservoir output is then sampled at 200 GSa/s to yield 20 time-multiplexed nodes per input clock cycle. We sweep the feedback time (50 values), in the form of the feedback waveguide's length, from 0.2τ (corresponding to approximately 3.0 mm on Si_3N_4 and 2.65 mm on LNOI) to 5τ (approximately 7.51 cm on Si_3N_4 and 6.62 cm on LNOI), and the applied phase shift (49 values) on the bottom MZI arm from 0 to 2π. The performance is then recorded with the relevant metric for each test/task.

5.3.1 Dispensing with the input mask

An input masking protocol is generally used in TDRC schemes to keep the reservoir continuously perturbed to delay it from falling to a steady state, and thereby aiding in improving the reservoir's memory [22, 23]. However, there are limitations imposed by the use of a mask in practice (especially in photonic implementations) which call for alternative approaches [6]. Masks normally have to be prepared electronically, which may necessitate a domain conversion at the input, instead of allowing an incoming optical signal to be directly processed by the reservoir. Moreover, for high speed implementations in the GHz range, the input mask preparation not only introduces an associated energy cost, but can also put additional constraints on modulation speeds. This would exhaust the arbitrary waveform generator's output sample rate on the mask, thus limiting the system's information processing speed. For example, a 10 GSa/s input signal, when masked for a reservoir of 20 nodes,

would be a 200 GSa/s signal, which is impractical.
One solution to this problem is by shortening the feedback delay time τ_{fb} to fractions of the input clock cycle τ. If the reservoir's timescale is short enough - as in the case of passive optical reservoirs - this would allow the same input sample to interact with past version(s) of itself, such that the system is kept continuously perturbed during the interval of one input sample. The effect produced can be explained in two complementary ways:

1. The input interacts with itself and thus increases the complexity, or 'richness' of the output signal;

2. The shorter feedback lengths allow the reservoir to more quickly 'forget' inputs further in the past, resulting in a lower GR, which in turn increases CA.

For this reason, it is of interest in this study to dispense with the masking procedure and observe how well the reservoir can solve some tasks when fed the input signal directly. As will be seen, a signal interacting with itself in the sub-τ regime can provide the necessary computational power to carry out some tasks, especially those not requiring a large memory. It is especially important that the feedback losses A_{fb} are kept to a minimum, as this prolongs the reservoir's activity and delays the time taken to reach steady state (Fig. 5.2(a,c)). As shown in Fig. 5.2(b,d,f), higher A_{fb} results in many nodes sharing the same state, which means that the nodes are not linearly independent and thus reduces the effective dimensionality of the reservoir. Furthermore, the delay time also affects the effective reservoir dimensionality, as longer delays (with respect to the input clock cycle) would increase the amount of nodes that share the same state (Fig. 5.2(e)). One must also ensure that the node states are sufficiently distinguishable above the system's noise floor, which primarily results from the electronic readout and photodetector sensitivity. Furthermore, signal filtering also affects the effective dimensionality of the reservoir, especially if the timescale of the filtering is close to the input clock cycle or slower. For this reason, it is of interest in this study to dispense with the masking procedure and observe how well the reservoir can solve some tasks when fed the input signal directly, in the presence of the electronic readout. As will be seen, a signal interacting with itself in the sub-τ regime can provide the necessary computational power to carry out some tasks with moderate memory requirements.

5.3.2 Task-independent performance

In Fig. 5.3, we show the performance of the proposed reservoir on MC_{tot}, CA, and its constituent metrics GR and KQR, while considering different feedback lengths, and

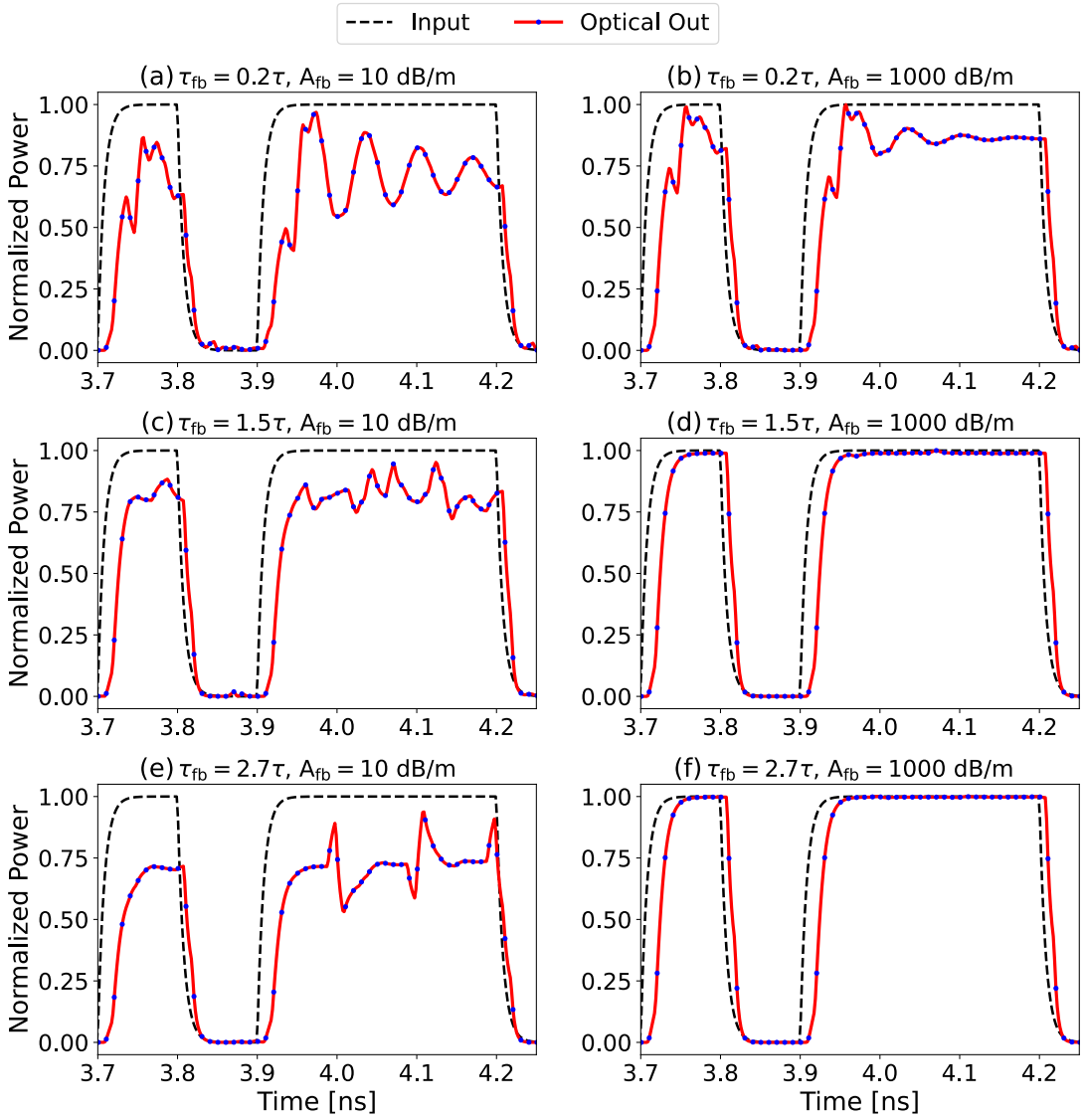

Fig. 5.2: Effect of τ_{fb} and A_{fb} on the dynamics and effective dimensionality of the reservoir. Blue dots represent the sampled nodes. τ is the delay time synchronized to the input clock cycle (100 ps) [1].

considering an applied phase shift between 0 rad to 2π rad. Here, we consider all the outputs passing through a photodetector with 40 GHz bandwidth and a 12-bit ADC. For feedback times less than τ, it is observed that MC_{tot} (CA) is generally low (high). A closer look at the constituent metrics of CA (Fig. 5.3(c,d)) shows that GR lowers with decreasing feedback time, especially in the sub-τ region, while also showing some dependence on the applied phase shift. However, for KQR, the variation is smaller over the entire design space, with an especially lower rank in the sub-τ region. Since GR becomes considerably lower in this region, the overall CA is highest. Less variation is observed in KQR when compared to GR. These variations are phase-dependent as the nonlinearity is effectively performed by the

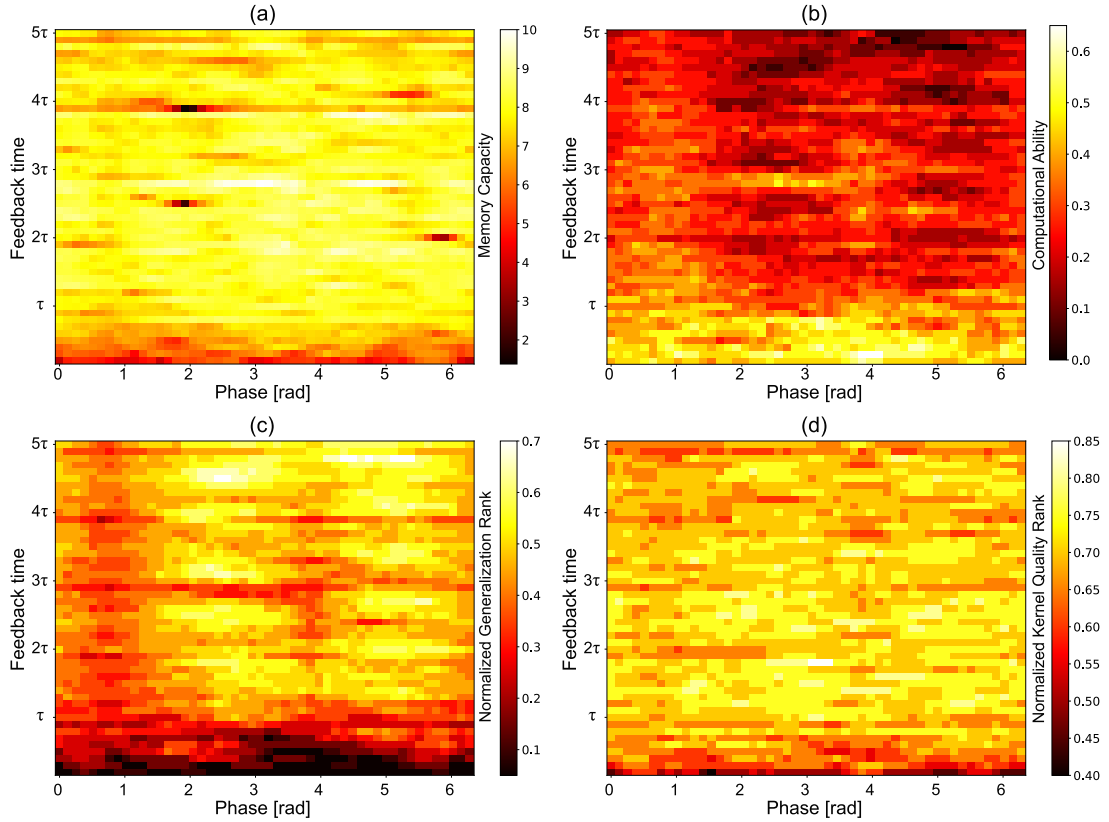

Fig. 5.3: Results on task-independent tests for different feedback lengths: (a) MC_{tot}, (b) CA, (c) GR, and (d) KQR [1].

strength of the interference between the fields at the couplers, which is controlled by the phase shifter. This also changes the signal-to-noise ratio (SNR) of the output signal and the overall Q-factor of the implicit cavity, which impacts the performance. Overall, these results show that there is opportunity in scaling down the feedback length while having enough MC_{tot} and CA to do some tasks in the absence of input masking, as will be explored in the following sections.

5.3.3 Benchmark task: NARMA-k

In this section, we consider the performance on the nonlinear autoregressive moving average (NARMA-k) benchmark task. This task requires both memory and nonlinearity. By varying the maximum number of memory steps k, we challenge the reservoirs in the design space with different memory/nonlinearity requirements. Furthermore, we also consider the temporal XOR-k task ($u[n] \oplus u[n-k]$) and the Santa Fe chaotic laser prediction, which are not displayed here for the sake of brevity, but which are taken into account in sections 5.3.4 and 5.3.7. For all NARMA-k, a thresholding trend can be observed for a feedback time longer than the most distant memory. This is due to the absence of interaction between the first input and

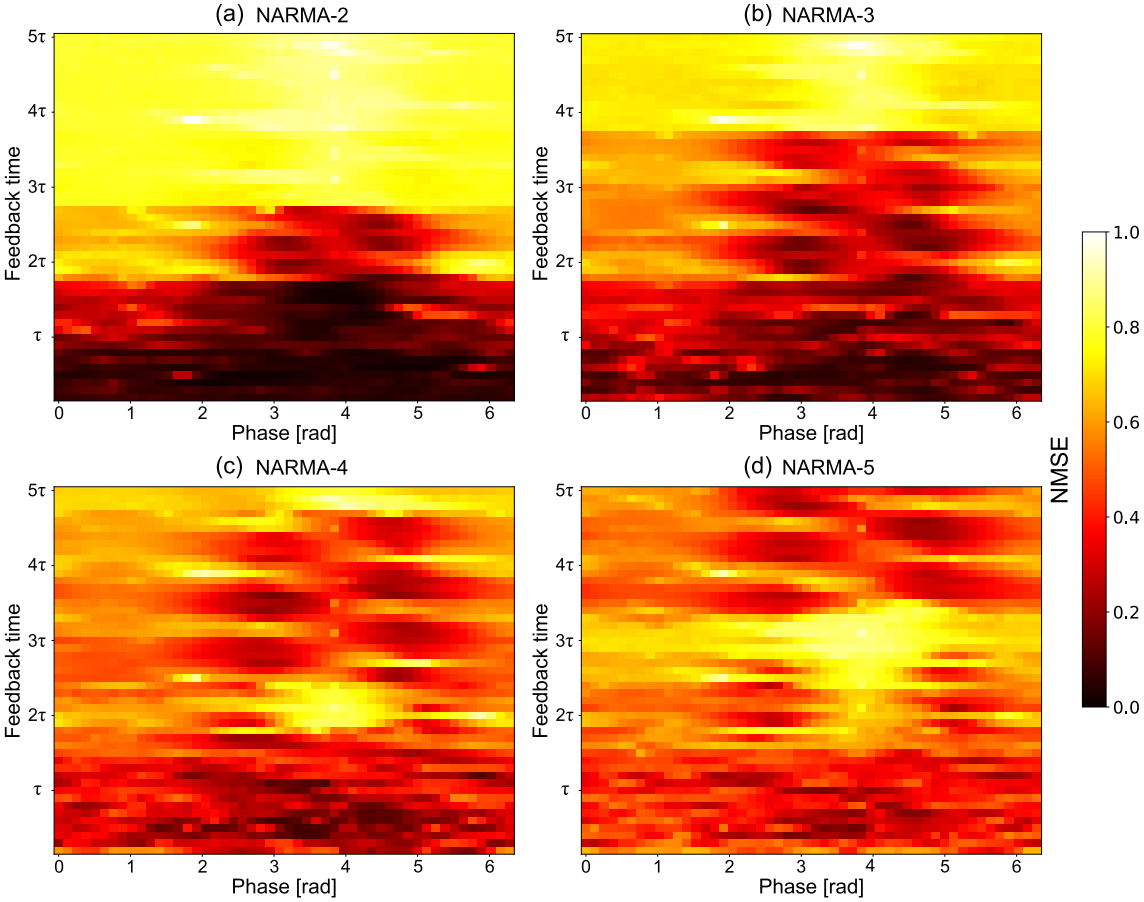

Fig. 5.4: Performance over the design space on the NARMA-k benchmark task. Shorter feedback delay times achieve excellent performance especially for lower k [1].

earlier inputs into the past, which are essentially skipped. Since the NARMA-k task considers the aggregate of previous inputs (Eq. 5.4), it depends on the total memory until the chosen k. This is observed in Fig. 5.4 where, beyond a certain feedback time, some earlier components of the memory are missing and thus lead to significantly worse performance. This sensitivity to earlier k reduces as k increases (as more terms are being aggregated). Additionally, for k's which are further back in time, their attenuated power in the circuit becomes more sensitive to readout noise, which is the reason that this task does not achieve a low enough NMSE for k higher than 5. In the sub-τ region, it can be seen that a low NMSE can be achieved for all reported NARMA-k. However, differing trends are observed due to the different memory/nonlinearity requirements from the RC system. Furthermore, upon closer examination, there seems to be a striking resemblance in pattern between both NARMA-4 and NARMA-5, and CA and GR results in Fig. 5.3(b,c). This is explored further in the section on correlation (section 5.3.7).

5.3.4 Performance comparison

In this section, we compare the obtained results on the mask-less protocol with results simulating the same architecture considering a pseudo-randomly generated mask from an i.i.d. $\in (0, 1]$, which is synchronized to the 20 nodes of the reservoir. Here, the same mask is considered for all the reported tasks. While, in practice, it would be difficult to generate such a mask due to limitations on the RF signal generation equipment (as discussed in section 5.3.1), it is nonetheless interesting – from the point of view of a numerical study – to assess how the mask-less performance compares to that of a masked version. Additionally, we compare to results reported in the literature for some RC schemes, including integrated TDRC [24, 25, 26, 27], integrated spatial-multiplexed RC [28], and bulk [29]. Similar to section 5.3.3, we scanned the parameter space in the length-phase space and considered a fixed high-speed 40 GHz photodetector with a subsequent 12-bit ADC. The best performance on the considered tasks is reported in Table 5.2. For the NARMA$-k$ task, it can be seen that the mask-less performance is close to the masked performance, and is also better than the simulated performance of the RC scheme in [29] except for NARMA-5. Since the test set of the XOR task is 3000 bits, the BER resolution is 3.33×10^{-4}. Therefore, we have also included the NMSE between brackets to observe the performance variations between XOR-1 and XOR-2 which both yield BER $= 0.0$. It can be seen that the unmasked performance is good up to XOR-3, beyond which the masked scheme outperforms it. However, the mask-less BER is still close to the state of the art performance [28, 27]. For the Santa Fe task, we can observe no significant difference between the optimal NMSE for the masked and mask-less scheme. Compared to the state of the art, the mask-less scheme achieves better performance than what is reported in [25], and performs worse than the simulated scheme in [26]. Furthermore, it can be seen in Table 5.2 that our scheme can provide a large enough $\mathrm{MC}_{\mathrm{tot}}$ compared to state of the art integrated photonic TDRC schemes, and the optimal $\mathrm{MC}_{\mathrm{tot}}$ improves by 1 when using the masked scheme. For the majority of tasks considered here, the performance obtained for the unmasked reservoir is comparable to that of the masked reservoir, and even sometimes slightly outperforms it. However, for tasks which require a larger memory, the mask indeed provides a considerable improvement, e.g., in NARMA-5 and XOR-4,5. When comparing to examples from the literature, it is important to consider the different factors involved. For example, the number of nodes, training and test dataset lengths, regularization parameter(s), and post-processing are almost different in every case, which makes it difficult to have a completely fair comparison. Nevertheless, it is shown that the mask-less scheme achieves results that are within

Table 5.2: Performance Comparison of Proposed vs. Existing Schemes

Task	Masked	Unmasked	State of the Art
NARMA-2	9.37×10^{-3}	6.69×10^{-3}	5×10^{-2} [29]*
NARMA-3	1.73×10^{-2}	3.75×10^{-3}	6×10^{-2} [29]*
NARMA-4	6.42×10^{-2}	7.95×10^{-2}	0.12 [29]*
NARMA-5	9.93×10^{-2}	0.149	0.14 [29]*
XOR-1	0.0 (7.49×10^{-5})	0.0 (1.72×10^{-5})	10^{-3}[28] $10^{-0.7}$[27]
XOR-2	0.0 (7.51×10^{-3})	0.0 (1.26×10^{-2})	2×10^{-3}[28] $10^{-0.9}$[27]
XOR-3	0.0 (4.68×10^{-3})	6.67×10^{-4} (0.10)	2×10^{-2}[28] $10^{-2.6}$[27]†
XOR-4	0.0 (5.39×10^{-3})	3.20×10^{-2} (0.22)	7×10^{-2}[28] — [27]
XOR-5	0.0 (5.19×10^{-3})	2.30×10^{-2} (0.19)	— [28] — [27]
Santa Fe	0.102	0.108	0.135 [25], 4×10^{-2} [26]
MC_{tot}	10.06	9.06	∼1.0 [26]‡, 1.5 [24], 6.0 [25]
Nodes	20	20	124 [24], 16 [28], 35 [29] 23 [26], 23 [25], 5 [27]

For the XOR task BER is reported, NMSE is between brackets. For Santa Fe and NARMA-k, NMSE is reported.
— means unreported
Simulation results [26, 28, 29], experimental results [25, 27, 24].
* NMSE calculated from the reported R2 scores as NMSE= 1−R2
† Post-processing involved
‡ Ref.[26] reports $MC_{tot} = 2$ without external feedback, but the authors consider also the term MC_0.

the vicinity of state of the art performance, and comparable to those obtained from applying the mask on the proposed scheme.

Furthermore, it can be seen in Table 5.2 that our scheme can provide a large enough MC_{tot} compared to state of the art integrated photonic TDRC schemes, and the optimal MC_{tot} improves by 1 when using the masked scheme. For the majority of tasks considered here, the performance obtained for the unmasked reservoir is comparable to that of the masked reservoir, and even sometimes slightly outperforms it. However, for tasks which require a larger memory, the mask indeed provides a considerable improvement, e.g., in NARMA-5 and XOR-4,5. When comparing to examples from the literature, it is important to consider the different factors involved. For example, the number of nodes, training and test dataset lengths, regularization parameter(s), and post-processing are almost different in every case, which makes it difficult to have a completely fair comparison. Nevertheless, it is shown that the mask-less scheme achieves results that are within the vicinity of state of the art performance, and comparable to those obtained from applying the mask on the proposed scheme.

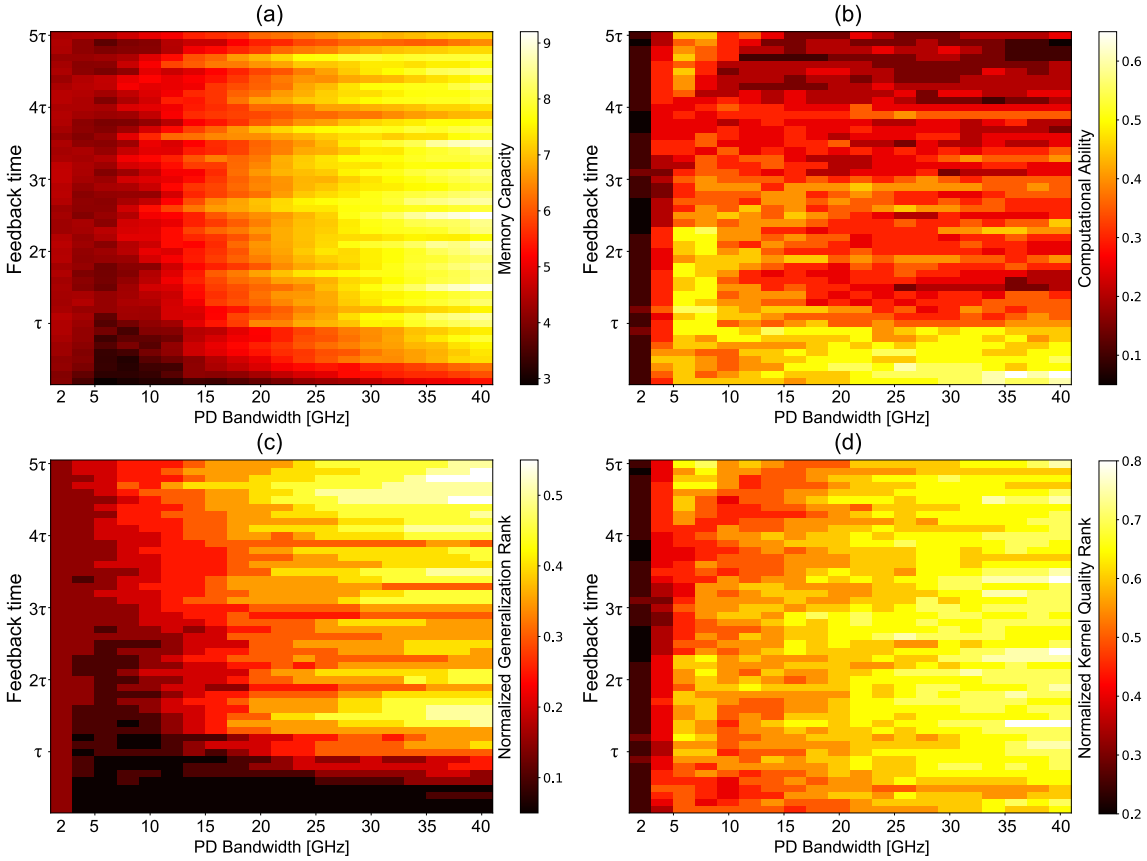

Fig. 5.5: Performance on the task-independent tests, with respect to photodetector bandwidth and feedback time [1].

5.3.5 Effect of post-filtering

In this section, we consider the effect of post-filtering at the readout. Filtering has been recently investigated [30] as a means of performance improvement for optoelectronic reservoirs. This is of particular interest for passive photonic reservoirs as the readout imprints features on the final output signal which need to be considered. Here, we sweep the photodetector's bandwidth (20 points), ranging from 2 GHz to 40 GHz, while considering the task-independent metrics and the NARMA-k task. The results are not independent of the applied phase shift. However, for visualization purposes, we choose one value of phase shift, and plot the results with respect to feedback time. The results on $\mathrm{MC_{tot}}$, as shown in Fig. 5.5 (a), show an overall increasing trend with increasing bandwidth. In Fig. 5.5 (b), there is a consistently high CA in the sub-τ region from 5 GHz and above. Thus, for tasks that require less $\mathrm{MC_{tot}}$ and more CA, one could in principle scale down the reservoir while being less constrained by readout electronics. For the most part, KQR increases quasi-monotonically with increasing bandwidth, similar to Fig. 5.5 (a). Also, it can be seen that smaller τ_{fb} are more tolerant to lower bandwidths in terms of CA. We

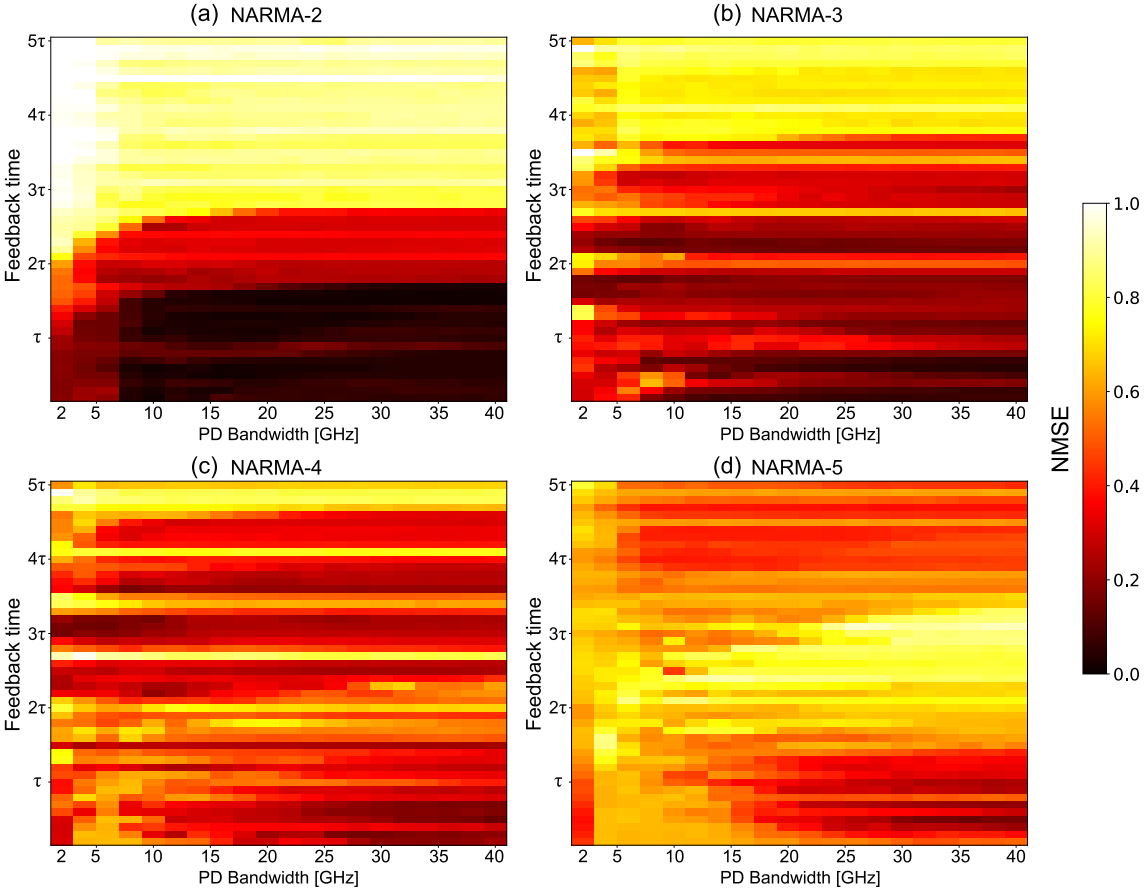

Fig. 5.6: Performance on NARMA-k benchmark task. For some feedback lengths, post-filtering assists in lowering the NMSE [1].

also consider the NARMA-k benchmark task, and report the behavior for different photodetector bandwidths, as shown in Fig. 5.6. For NARMA-2 and NARMA-3, excellent performance is obtainable in the sub-τ region, where there is a large degree of tolerance for lower bandwidths. On the other hand, NARMA-4 and NARMA-5 seem to benefit more from the slower dynamics of longer delay lines. We conclude that the output readout affects the dimensionality of the reservoir, mostly reducing the effective number of nodes through a combination of noise and linear filtering, which could essentially drown some of the independent features which become insufficiently distinguishable after passing through the output electronics. Thus, shorter feedback lengths, which provide stronger and more frequent dynamics, seem to be particularly robust to the readout's considered effects.

5.3.6 Effect of ENOB

ADCs in oscilloscopes typically have a lower effective number of bits (ENOB) than their designed resolution, especially when operating at high frequencies. Therefore,

it is interesting to see how the proposed RC scheme would perform with such low ENOB. We consider how the scheme performs in the sub-τ region, showing the optimum values obtained for some of the tasks presented earlier in the work under the harsh conditions of a single oscilloscope acquisition, which is equivalent of greatly reducing the considered SNR in this work. Thus, we consider the cases of 5-8 bits of ADC resolution to mimic the effect of a lowering ENOB. For XOR-k tasks we report the BER, while for NARMA-k and Santa Fe tasks we report the NMSE.

Table 5.3: Performance on lower ENOB

Task	5 bits	6 bits	7 bits	8 bits
XOR-1	0.0	0.0	0.0	0.0
XOR-2	8.3×10^{-3}	6.6×10^{-4}	0.0	0.0
NARMA-2	0.13	0.08	0.05	0.04
NARMA-3	0.29	0.21	0.16	0.12
SF	0.46	0.27	0.22	0.17

While the consideration of a single acquisition heavily degrades the signal quality, the proposed scheme can still solve the above tasks effectively, albeit with less accuracy. This degradation can be easily mitigated by averaging multiple signal acquisitions, which is usually done in practice and can improve the ENOB by multiple effective bits of precision, in accordance with the number of signals acquired and averaged. An important consideration here would be the signal stability during the total acquisition time of the multiple signals to be averaged. This is done in practice by using a temperature controller to keep the photonic circuit thermally stable during the total acquisition time. Usually, precision temperature controllers based on resistance temperature detector (RTD) in a feedback loop – typically proportional-integral-derivative (PID) – can achieve a thermal stability < 0.1°C. However, when considering real-time processing, one limitation is that averaging over multiple acquisitions would decrease the overall throughput of the system. Higher input optical powers would also further improve the SNR and minimize the needed amplification after detection, reducing their associated noise. Under these considerations, the simulation results obtained in this work should approximate the experimental performance.

5.3.7 Correlation between different tasks

The relation between task-specific performance and task-independent metrics has also been a topic of recent interest [31]. In this section, we shed light on the relation between different tasks through observation of the correlation between the

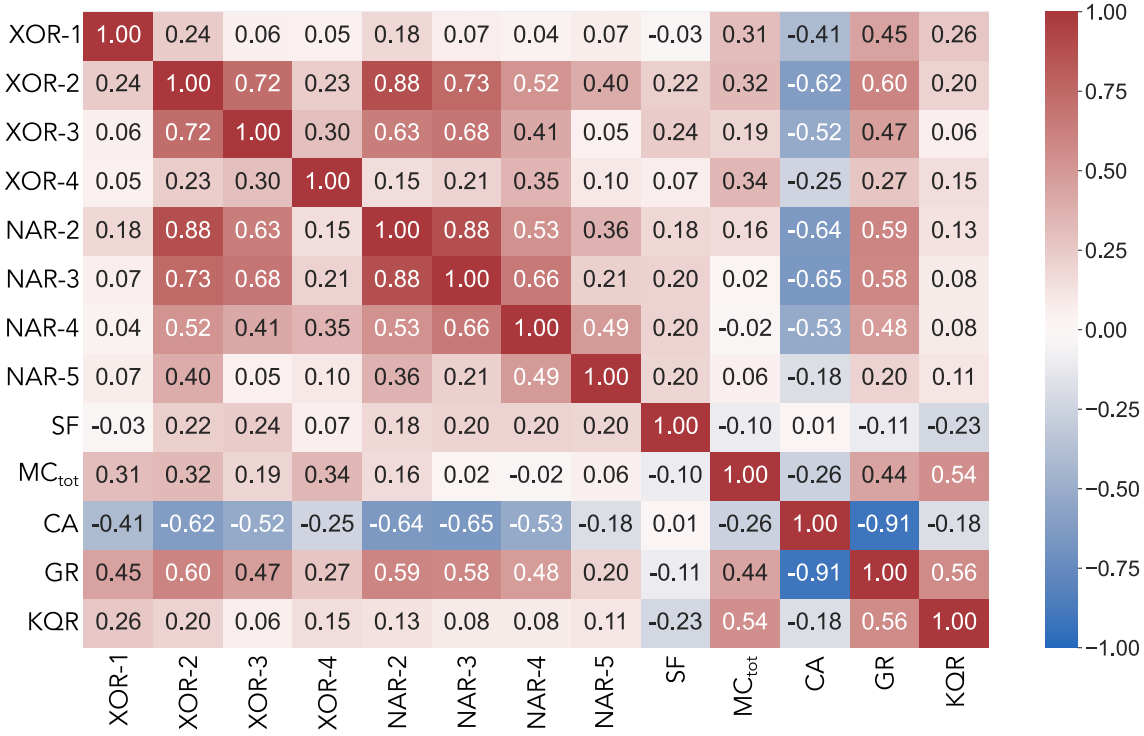

Fig. 5.7: Correlation matrix between different tasks, XOR-k, NARMA-k (NAR-k), Santa Fe (SF), and the task independent metrics [1].

reservoir's performance on different tasks over the chosen design space, where we consider the varying feedback length and applied phase shift, and a fixed photodetector bandwidth of 40 GHz. We employ the Pearson correlation coefficient (r_c) to identify linear relations between the performance on various tasks/metrics over the design space. We take into consideration all the presented tasks thus-far, namely the temporal bitwise XOR-k task (with k from 1 to 4 bits into the past), the NARMA-k task (with k from 2 to 5 inputs in the past), the Santa Fe chaotic laser prediction, as well as MC_{tot}, CA, and its constituent metrics (GR, KQR). In this case, for the XOR task, we also employ the NMSE as BER does not yield results falling within a normal distribution. When $0 < r_c \leq 1$, a positive correlation exists, and when $-1 < r_c < 0$ it is a negative correlation. A strong positive correlation indicates that two tasks perform similarly well (or similarly poorly) over the design space, while a strong negative correlation would mean that the tasks require opposite configurations. If a strong positive correlation exists between two tasks, we can infer that it is highly likely that the system can be configured in such a way to solve those two tasks at the same time, i.e. the tasks can be solved with the same hyperparameter configuration of the reservoir.

A strong positive correlation is observed between XOR tasks and NARMA tasks for some k. For example, $r_c(\text{xor2}, \text{narma2}) = 0.88$, which suggests that they can ben-

efit from sharing the same hyperparameter configuration within this design space, even though they are different families of tasks. The same applies to other cases: $r(\text{xor2}, \text{narma3}) = 0.73$, $r_c(\text{narma2}, \text{narma3}) = 0.88$ and $r_c(\text{xor2}, \text{xor3}) = 0.72$. These same tasks also exhibit a strong positive correlation with CA (the two quantities CA and NMSE improve in opposite directions). Looking at CA's constituent metrics, it is observed that this is mainly due to the stronger correlation with GR, as the correlation with KQR is generally weak with all benchmark tasks. This is due to the small variations in KQR over the explored design space, which was explained in section 5.3.2, and is the reason why CA itself does not show a strong linear relation with KQR. Similarly, for the Santa Fe task, weak relations are established due to minimal variation of NMSE over the design space. Beyond the sub-τ region, MC_{tot} does not change considerably, and hence the average or weak correlations with other tasks/metrics. Therefore, for this scheme, CA is a more relevant metric of the reservoir's performance on the different tasks than MC_{tot}, especially considering the fact that GR is in itself a test of memory. In fact, the individual memory components MC_k can give a better picture on how much the reservoir can remember for tasks that target specific inputs into the past, such as XOR-k. For other RC schemes, which can have sufficiently different types of nonlinearity, a stronger correlation with KQR over the design space would be expected. In any case, the correlation matrix not only shows inherent relations between different tasks over that space, but is also a useful tool to judge a reservoir's ability to solve multiple desired tasks on the same signal. It can also aid in the design process of any type of reservoir for the tasks of interest.

5.4 Conclusion

In this work, we have explored the potential for mask-less photonic TDRC using an all-optical, passive, integrated architecture. For applications with an already-incoming optical signal, dispensing with the mask enables bypassing the usual domain conversion(s) at the input layer. In addition to reductions in associated energy and complexity costs, this enables the processing of higher signal bitrates B_{m} as the mask normally requires $N \times B_{\text{m}}$ of signal generation/modulation bandwidth, thus restricting either the size of the reservoir or the overall speed of the system. We have explored this within the chosen design space of the minimum complexity architecture, focusing on the feedback length, and the applied phase shift which can be freely tuned post-fabrication. The best obtained results show small differences between masked and mask-less performance for tasks with moderate memory requirements.

Additionally, we considered the effects of post-filtering and showed that in the sub-τ region, some tasks can be performed even with very strong linear filtering, which degrades the linear independence of the sampled nodes. Furthermore, we have shown the correlation between the reservoir performance on the considered tasks as a way to understand the relation between the performance on different benchmark tasks and the task-independent metrics within the chosen design space. This would allow further optimization of the reservoir design to handle multiple tasks using the same configuration. This could be useful for other reservoir implementations and serve as a practical tool to aid in the design of multitasking RC. Future studies may consider nonlinear effects of the considered platform (e.g. χ^2 and χ^3 nonlinearities in LNOI), along with relevant parameters such as varying the SNR, waveguide losses, input wavelength, MZI arm lengths, and coupling ratio of the splitters, which would influence the dynamics of the proposed reservoir and the effective (usable) number of nodes. Moreover, the number of nodes required for certain tasks may be less than what is considered here, which means that they can operate at faster input clock cycles (e.g. tasks showing tolerance to lower bandwidth post-filtering). Finally, other metrics such as the information processing capacity (IPC) [9] may be applied to quantify the reservoir performance and provide insight into the computational power of the intensity-conversion nonlinearity used here and in similar passive photonic RC schemes.

Bibliography

[1] M. Abdalla, R. Cardoso, P. Jimenez, M. G. de Queiroz, A. Boes, G. Ren, A. Mitchell, I. O'Connor, and F. Pavanello, "Mask-less asynchronous time-delay reservoir computing using a passive photonic integrated circuit," *Journal of Lightwave Technology*, pp. 1–10, 2024.

[2] F. Stelzer, A. Röhm, K. Lüdge, and S. Yanchuk, "Performance boost of time-delay reservoir computing by non-resonant clock cycle," *Neural Networks*, vol. 124, pp. 158–169, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608020300125

[3] T. Hülser, F. Köster, L. Jaurigue, and K. Lüdge, "Role of delay-times in delay-based photonic reservoir computing [invited]," *Optical Materials Express*, vol. 12, p. 1214, 3 2022.

[4] J.-Y. Tang, B.-D. Lin, Y.-W. Shen, R.-Q. Li, J. Yu, X. He, and C. Wang, "Asynchronous photonic time-delay reservoir computing," *Optics Express*, vol. 31, p. 2456, 1 2023.

[5] A. Rodan and P. Tiňo, "Minimum complexity echo state network," *IEEE Transactions on Neural Networks*, vol. 22, pp. 131–144, 1 2011.

[6] A. Argyris, J. Schwind, and I. Fischer, "Fast physical repetitive patterns generation for masking in time-delay reservoir computing," *Scientific Reports*, vol. 11, 12 2021.

[7] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," *Optics Express*, vol. 29, p. 30991, 9 2021.

[8] M. Abdalla, C. Zrounba, R. Cardoso, P. Jimenez, G. Ren, A. Boes, A. Mitchell, A. Bosio, O. Ian, and F. Pavanello, "Minimum complexity integrated photonic architecture for delay-based reservoir computing," *Optics Express*, 3 2023.

[9] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Scientific Reports*, vol. 2, 2012.

[10] B. Vettelschoss, A. Röhm, and M. C. Soriano, "Information processing capacity of a single-node reservoir computer: An experimental evaluation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2714–2725, 2022.

[11] K. Harkhoe and G. V. der Sande, "Task-independent computational abilities of semiconductor lasers with delayed optical feedback for reservoir computing," *Photonics*, vol. 6, 2019.

[12] S. Tsunegi, T. Kubota, A. Kamimaki, J. Grollier, V. Cros, K. Yakushiji, A. Fukushima, S. Yuasa, H. Kubota, K. Nakajima, and T. Taniguchi, "Information processing capacity of spintronic oscillator," *Advanced Intelligent Systems*, vol. 5, no. 9, p. 2300175, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202300175

[13] R. Martínez-Peña, J. Nokkala, G. L. Giorgi, R. Zambrini, and M. C. Soriano, "Information processing capacity of spin-based quantum reservoir computing systems," *Cognitive Computation*, vol. 15, no. 5, pp. 1440–1451, Sep 2023. [Online]. Available: https://doi.org/10.1007/s12559-020-09772-y

[14] H. Jaeger, "Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the" echo state network" approach," 2002.

[15] A. Röhm, L. Jaurigue, and K. Lüdge, "Reservoir computing using laser networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–8, 2020.

[16] R. A. Legenstein and W. Maass, "2007 special issue: Edge of chaos and prediction of computational performance for neural circuit models," *Neural Networks*, vol. 20, pp. 323–334, 2007. [Online]. Available: https://api.semanticscholar.org/CorpusID:7249094

[17] Appeltant, *PhD Thesis: Reservoir computing based on delay-dynamical systems*, 2012. [Online]. Available: www.vubpress.be

[18] M. Dale, J. F. Miller, S. Stepney, and M. A. Trefzer, "A substrate-independent framework to characterize reservoir computers," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 475, 6 2019.

[19] A. Weigend and N. Gershenfeld, "Results of the time series prediction competition at the santa fe institute," in *IEEE International Conference on Neural Networks*, 1993, pp. 1786–1793 vol.3.

[20] K. A. Buzaverov, A. S. Baburin, E. V. Sergeev, S. S. Avdeev, E. S. Lotkov, M. Andronik, V. E. Stukalova, D. A. Baklykov, I. V. Dyakonov, N. N. Skryabin, M. Y. Saygin, S. P. Kulik, I. A. Ryzhikov, and I. A. Rodionov, "Low-loss silicon nitride photonic ics for near-infrared wavelength bandwidth,"

Opt. Express, vol. 31, no. 10, pp. 16 227–16 242, May 2023. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-10-16227

[21] A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (lnoi) for photonic integrated circuits," *Laser and Photonics Reviews*, vol. 12, 4 2018.

[22] L. Appeltant, G. V. D. Sande, J. Danckaert, and I. Fischer, "Constructing optimized binary masks for reservoir computing with delay systems," *Scientific Reports*, vol. 4, 1 2014.

[23] Y. Kuriki, J. Nakayama, K. Takano, and A. Uchida, "Impact of input mask signals on delay-based photonic reservoir computing with semiconductor lasers," *Optics Express*, vol. 26, p. 5777, 3 2018.

[24] K. Takano, C. Sugano, M. Inubushi, K. Yoshimura, S. Sunada, K. Kanno, and A. Uchida, "Compact reservoir computing with a photonic integrated circuit," *Opt. Express*, vol. 26, no. 22, pp. 29 424–29 439, Oct 2018. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-26-22-29424

[25] K. Harkhoe, G. Verschaffelt, A. Katumba, P. Bienstman, and G. V. der Sande, "Demonstrating delay-based reservoir computing using a compact photonic integrated chip," *Optics Express*, vol. 28, p. 3086, 2 2020.

[26] G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Optics Express*, vol. 30, p. 522, 1 2022.

[27] D. Bazzanella, S. Biasi, M. Mancinelli, and L. Pavesi, "A microring as a reservoir computing node: Memory/nonlinear tasks and effect of input non-ideality," *Journal of Lightwave Technology*, vol. 40, pp. 5917–5926, 9 2022.

[28] A. Katumba, M. Freiberger, P. Bienstman, and J. Dambre, "A multiple-input strategy to efficient integrated photonic reservoir computing," *Cognitive Computation*, vol. 9, no. 3, pp. 307–314, Jun 2017. [Online]. Available: https://doi.org/10.1007/s12559-017-9465-5

[29] J. Henaff, M. Ansquer, M. C. Soriano, R. Zambrini, N. Treps, and V. Parigi, "Optical phase encoding in a pulsed approach to reservoir computing," *Opt. Lett.*, vol. 49, no. 8, pp. 2097–2100, Apr 2024. [Online]. Available: https://opg.optica.org/ol/abstract.cfm?URI=ol-49-8-2097

[30] G. O. Danilenko, A. V. Kovalev, E. A. Viktorov, A. Locquet, D. S. Citrin, and D. Rontani, "Impact of filtering on photonic time-delay reservoir computing," *Chaos*, vol. 33, 1 2023.

[31] T. Hülser, F. Köster, K. Lüdge, and L. Jaurigue, "Deriving task specific performance from the information processing capacity of a reservoir computer," *Nanophotonics*, 3 2022.

Chapter 6

Experimental Work

This chapter reports the experimental work carried out during my stay at RMIT University, at the Integrated Photonics and Applications Center (InPAC), from February to September 2023. The aim of the reported experiments is to verify the proposed RC scheme in chapter 4, under mask-less conditions as discussed in chapter 5. The layout and experimental characterization were carried out by myself, while fabrication was done by the InPAC fabrication team, led by Dr. Guanghui Ren. The platform considered is the SiN-loaded LNOI, as shown in Fig. 6.1. SiN is a good optical load material due to a similar transparency and refractive index as $LiNbO_3$. Furthermore, SiN is CMOS-compatible, which allows standard deposition and etching to be used, which simplifies fabrication. The ridge waveguide (Fig. 6.1) enables lower propagation losses due to the minimal area of interaction between the optical mode and the sidewalls' roughness. This is because most of the mode is confined in the thin-film lithium niobate (TFLN) layer. However, due to weaker lateral confinement than wire waveguides (where the refractive index contrast is with air), these ridge waveguides require larger bending radii ($\sim 200\mu$m) to minimize associated radiative losses. However, for straight waveguides, losses as low as 0.3 dB/cm can be obtained. The following sections aim to document the activities I carried out, from the preparation of the chip layout using IPKISS[1], to the lab work encompassing device characterization and RC system testing.

[1]https://www.lucedaphotonics.com/luceda-photonics-design-platform

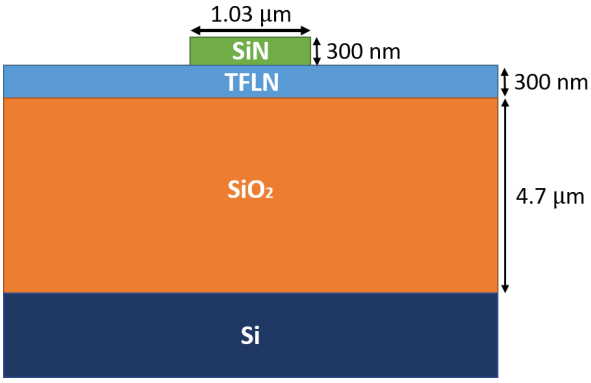

Fig. 6.1: Cross-section of the SiN-loaded LNOI platform. More information about the LNOI platform can be found here [1].

6.1 Chip Layout Design

The mask layouts were prepared using InPAC's LNOI process design kit (pdk) in IPKISS. The first layout is shown in Fig. 6.2. The grating couplers (GC) pitch was set to 127 µm for compatibility with the fiber array. The first and last GC were reserved for a reference loop, which estimates the GC insertion loss. Bend radii (including the concentric spiral waveguides) were set at 150 – 200 µm to ensure minimal bending losses. Heater pads were positioned in the horizontal direction for access to DC probes from below. On the other hand, the RF pads were positioned vertically for access from the left and right directions with respect to the chip. Three feedback lengths, previously considered in simulation in chapter 4 and chapter 5, were considered: 1.88 cm (R188), 2.84 cm (R284), and 4.55 cm (R455). For the devices only (without on-chip modulators), two instances of R188, one instance of R284, and one instance of R455 were placed on the layout. For devices with on-chip modulators, one instance each of R188 and R284 were placed. Furthermore, three test structures corresponding to the spiral lengths used were placed in the bottom half of the chip layout. These test structures allow the testing of the impact of bends on the spiral losses. Moreover, since LNOI is anisotropic, changing the direction of propagation along the chip affects which refractive index (ordinary or extraordinary) the field is experiencing, which contributes to the losses due to the modal mismatch. X-cut LNOI is usually preferred for electro-optic applications because it maximizes in-plane electric fields, utilizing the high electro-optic coefficients r_{11} and r_{12} for efficient modulation in integrated photonic devices. Design rule checking (DRC) was done manually, ensuring that:

- minimal amount of bends overlap with the electron beam lithography (EBL) grid lines to minimize stitching error;

- minimum distance between nearby waveguides is respected to minimize crosstalk;
- DC pad spacing (200 µm) corresponds to the pitch of the DC probe;
- RF pad spacing corresponds to the pitch of the RF probe.

The second layout (Fig. 6.3) was completed for a second fabrication run, which will take place at another facility, and which uses a different pdk developed by my colleagues at InPAC. In addition to the reference waveguides for estimating GC losses, additional reference waveguides were added to estimate the losses resulting from stitching errors on the new platform. Stitching errors in photonic chip fabrication refer to misalignments or discontinuities at the boundaries between the exposed sections during lithography, leading to misalignment and therefore optical losses. This layout does not include on-chip modulators and is thus focused on the devices themselves - namely the feedback structure used and its length, as well as the impact of the asymmetry of the MZI. Thus, the results reported in this chapter concern the first mask layout.

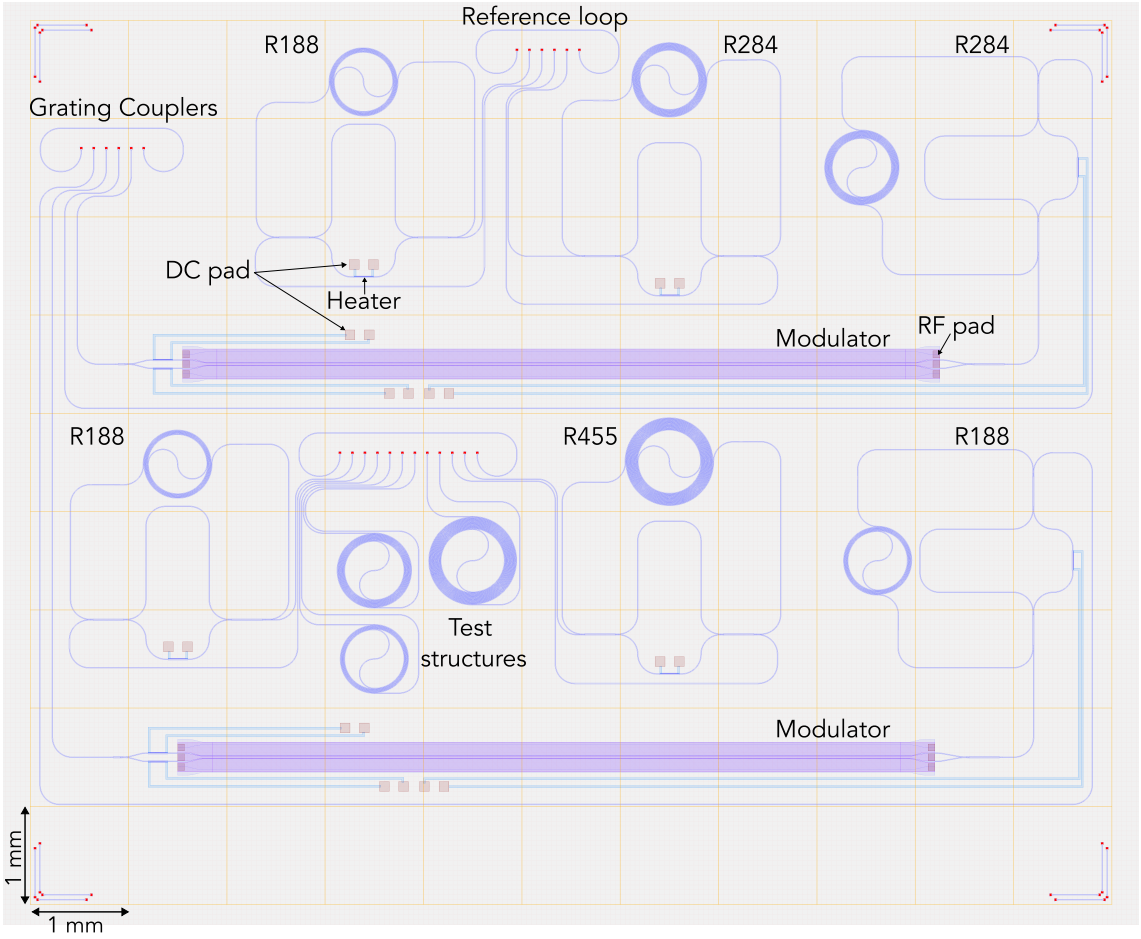

Fig. 6.2: The first layout includes test structures and test devices with and without on-chip modulators.

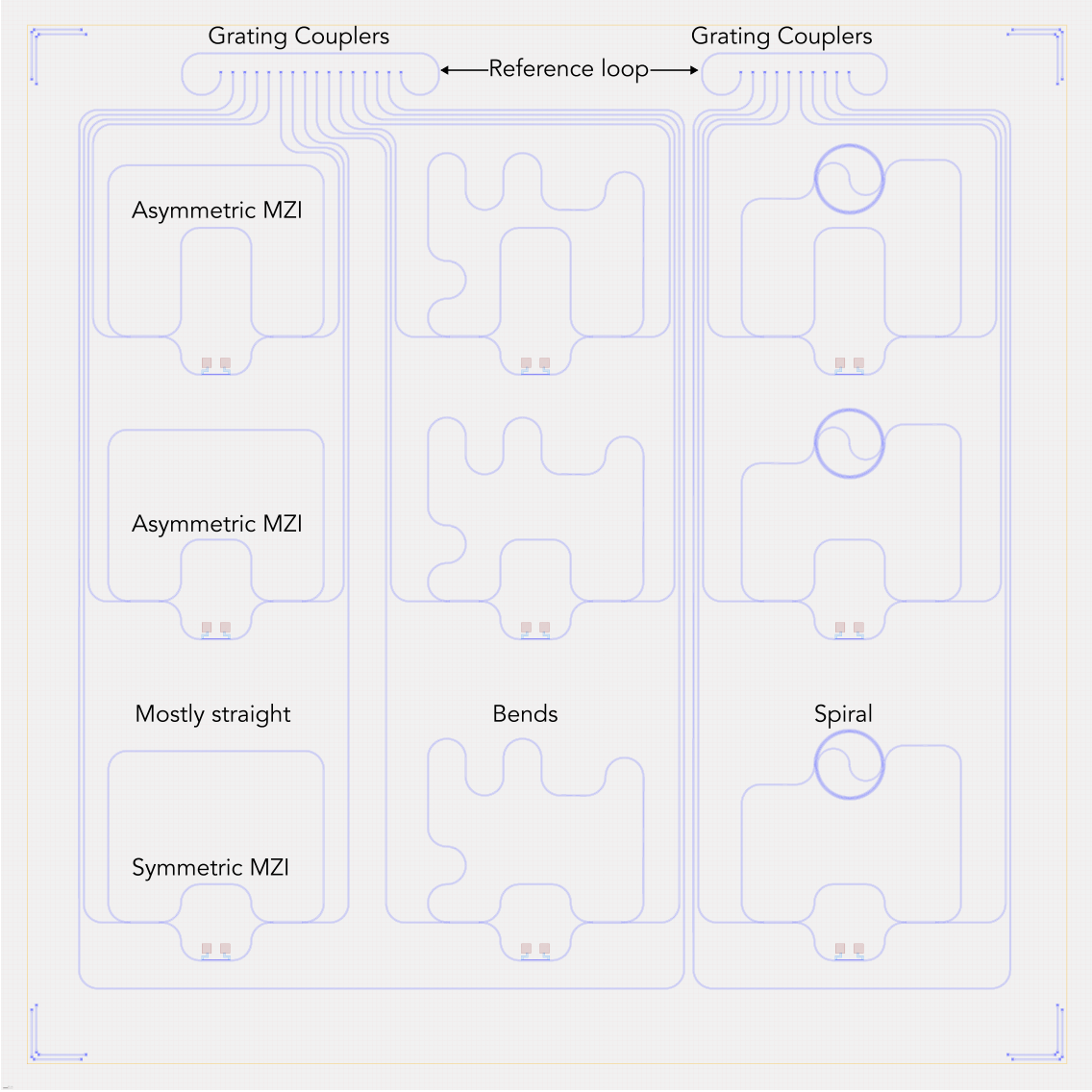

Fig. 6.3: The second layout allows the study of the impact of MZI asymmetry and the choice of feedback lengths and their chosen structures (mostly straight waveguides, meanders, and concentric spirals).

6.2 Device Characterization

In this section, the characterization of the grating couplers, reference spiral structures, and devices (R188, R284, R455) is reported. Extremely low losses were encountered for the devices with modulators, which are significantly lower than the noise floor of the setup, and therefore their performance is not reported here. The setup used is shown in Fig. 6.4. A tunable laser is connected to the fiber array which is aligned to the on-chip GCs which couple the light into waveguides (TE mode). The laser power is set at 0 dBm (1 mW).

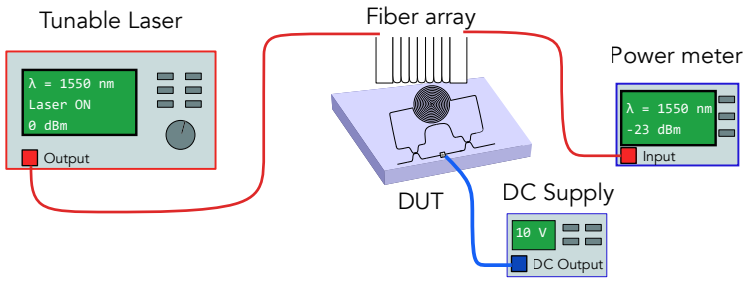

Fig. 6.4: Setup for loss characterization of the GCs, test structures, and devices.

The laser wavelength is swept and the power is recorded. The first measurements concern the reference loop which estimates the GC losses, as shown in Fig. 6.5. The results show that the GCs work at their designed wavelength, with a peak transmission at $\sim$ 1550 nm. However, it can also be seen that the insertion losses are high, with about 9.5 dB per GC for the upper devices and for the devices with modulators. For the bottom devices, the insertion losses are slightly more than 11 dB per GC. It is not unusual to have variability of performance over the dimensions of the chip in thin film platforms. This is due to the inconsistencies in TFLN layer thickness across the chip. The next measurement concerns the reference loops at the bottom of the chip, which share the same GC array as the bottom devices. It can be similarly observed that the GCs work at their designated wavelength, except for the one concerning the spiral used to form the feedback of the R455. Comparing to the values above, we can estimate that the spiral losses are the following:

- Loop188 $\approx$ 1.1 dB
- Loop284 $\approx$ 6.3 dB
- Loop455 $\approx$ 10.2 dB

The significant losses encountered can be attributed to modal mismatches between the ordinary and extraordinary refractive indices as the field continuously alters its

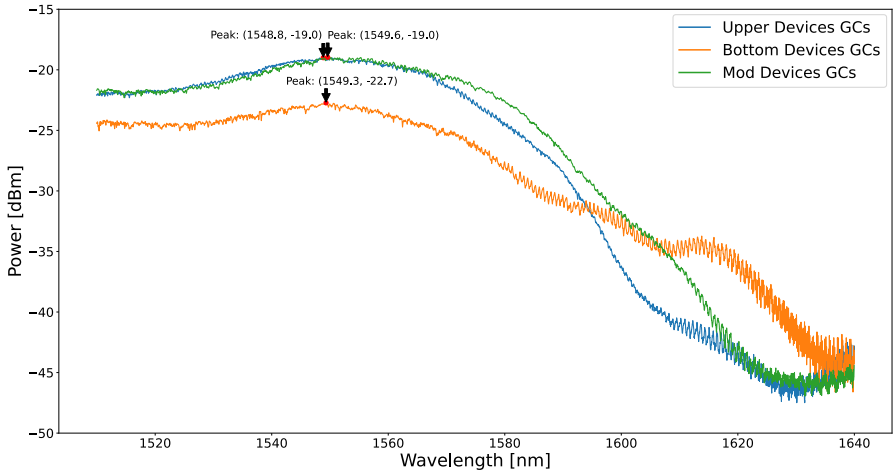

Fig. 6.5: Measurement of GC losses through the reference loop. Ripples are caused by the cavity formed by the fiber's cleaved end and the grating coupler.

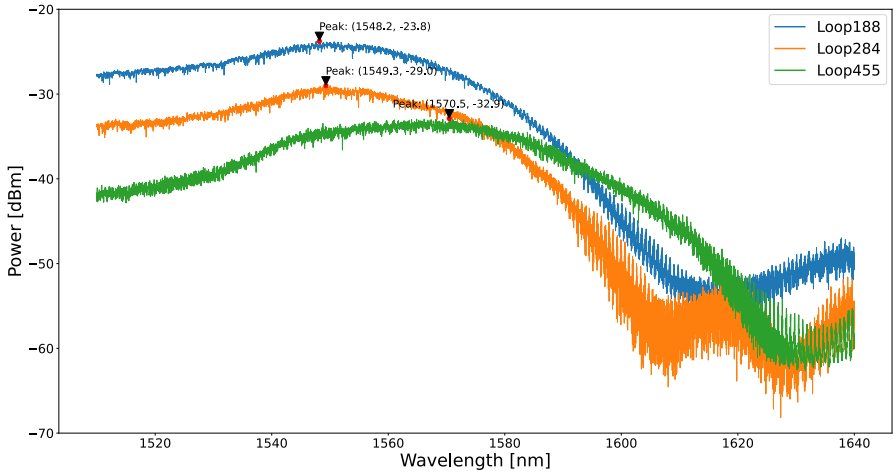

Fig. 6.6: Measurement of the feedback spiral losses through the reference structures.

orientation within the spiral. This does not exclude the possibility of losses due to local defects on the chip. After reporting the GC and spiral losses, we proceed to the device characterization. One can observe the envelope (FSR) created by the difference in MZI arm lengths. The smaller FSR is caused by the cavities formed by the feedback waveguide with the MZI arms. For larger L_{fb}, this FSR becomes smaller. Furthermore, the smaller ripples dampen with increasing L_{fb}, as a result of the higher losses encountered. Since for R455 the losses are very high, we proceeded with the system characterization of R188 and R284, excluding R455.

Finally, we proceed to the phase shifter characterization. We employ the DC supply, as shown in Fig. 6.4 to provide the voltage to the tungsten heater by means of a DC probe. The maximum operating voltage is around 9V, beyond which the heater can be permanently damaged. As shown in Fig. 6.8, the distance between two successive peaks on the MZI envelope is ≈ 650 pm, which means that a π phase shift is ≈ 325

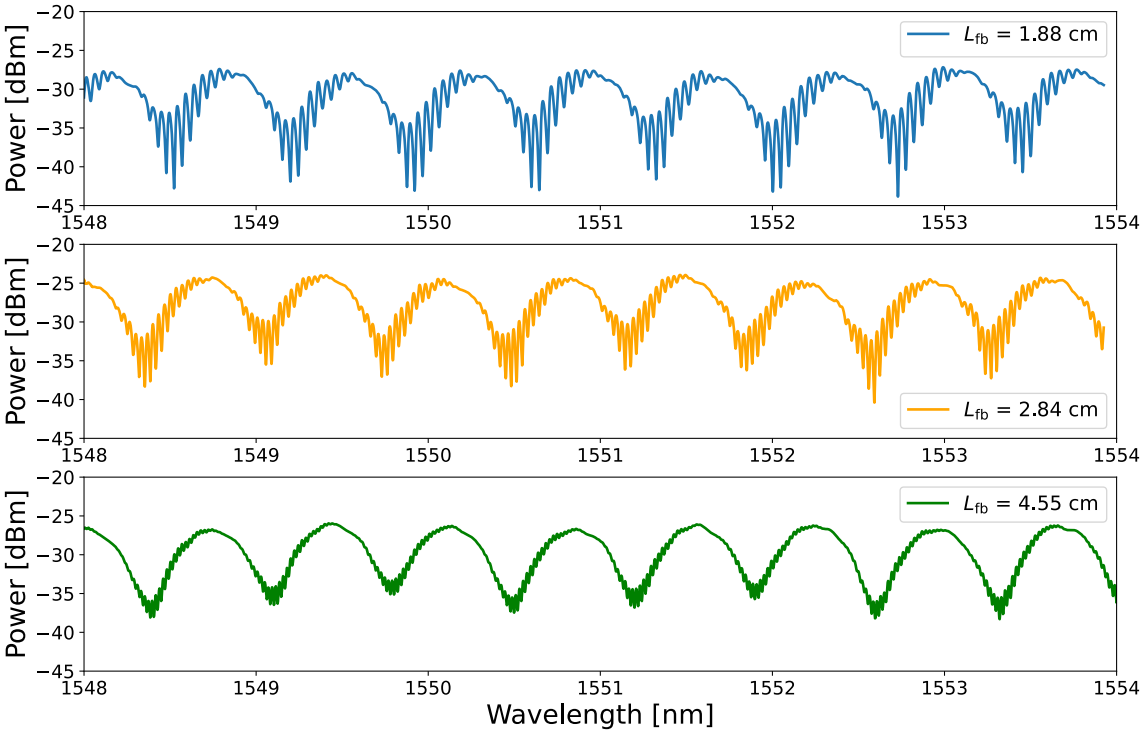

Fig. 6.7: Characterization for the devices with different feedback lengths.

pm. The full range of 228 pm thus provides $\approx 0.7\pi$ phase shift. This is a limitation in the fabricated chips that can be easily remedied in the future by increasing the heater length.

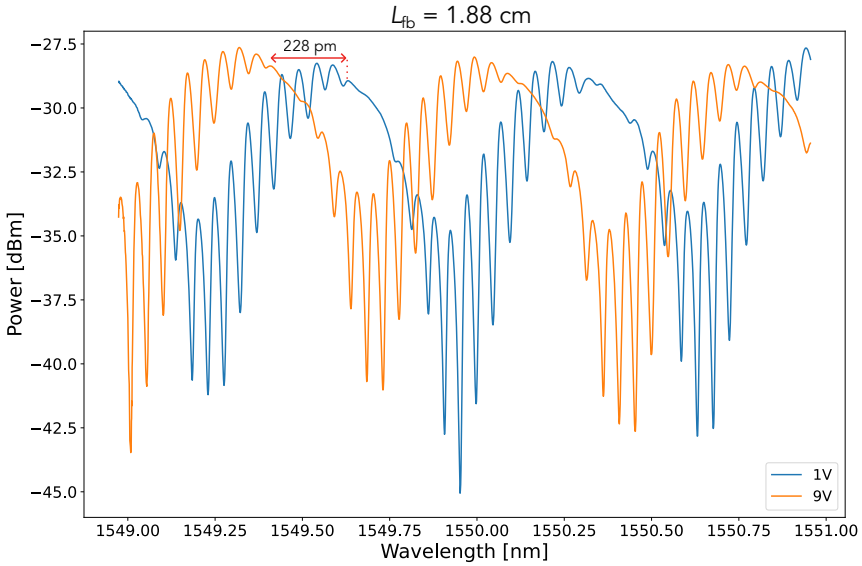

Fig. 6.8: Phase shifter characterization.

6.3 System Characterization

The system's temporal response is first captured using a non-return-to-zero (NRZ) input bitstream. A marker bit precedes the sequence and serves as the trigger level on the oscilloscope, as shown in Fig. 6.10 (a). The bulk modulator was also tested before the experiment, and it was found that a bias voltage of ∼14 V yielded the optimum swing. By creating a python script, based on an in-house PyVISA interface independent measurement control python package, the experiment is automated, i.e. the laser wavelength, DC supply voltage, oscilloscope acquisition could be controlled and swept. All functional instances of R188 and R284 were tested for 5 and 10 Gbit/s bit rates. Here reported are the results of one R188 instance, showing the temporal response to a 5 Gbit/s bit stream for various wavelengths $\lambda_1 - \lambda_6$ between 1549 nm – 1551 nm, and various V_ϕ between 5V – 9V, without (Fig. 6.11) and with (Fig. 6.12) a temperature controller (TC) for the device under test (DUT).

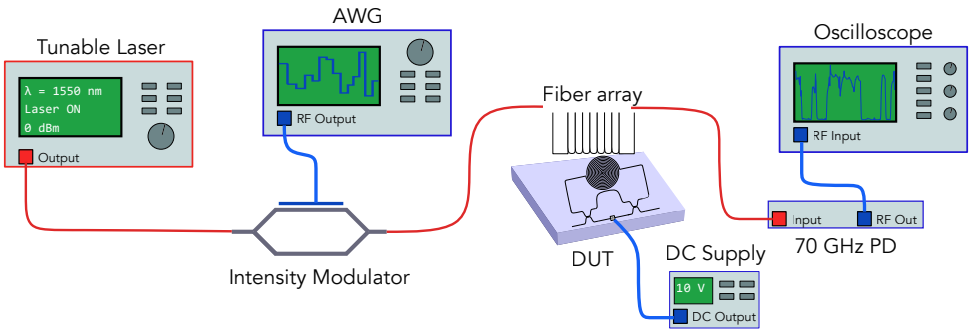

Fig. 6.9: Setup for RC system characterization. AWG refers to an arbitrary waveform generator. Erbium doped fiber amplifiers (EDFA) were also included before and after the device under test (DUT) to compensate for the losses. Additionally, an RF amplifier is used after the photodetector.

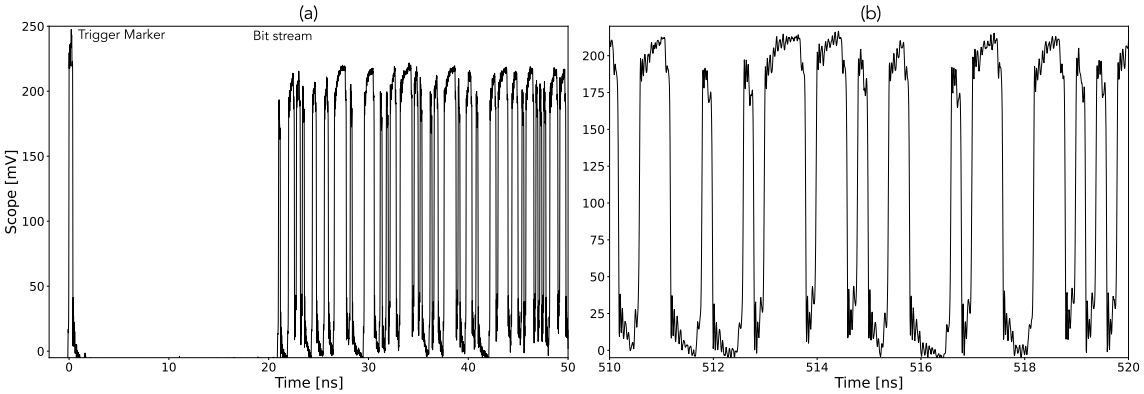

Fig. 6.10: AWG output of a 5 Gbit/s NRZ bit stream, measured on the oscilloscope.

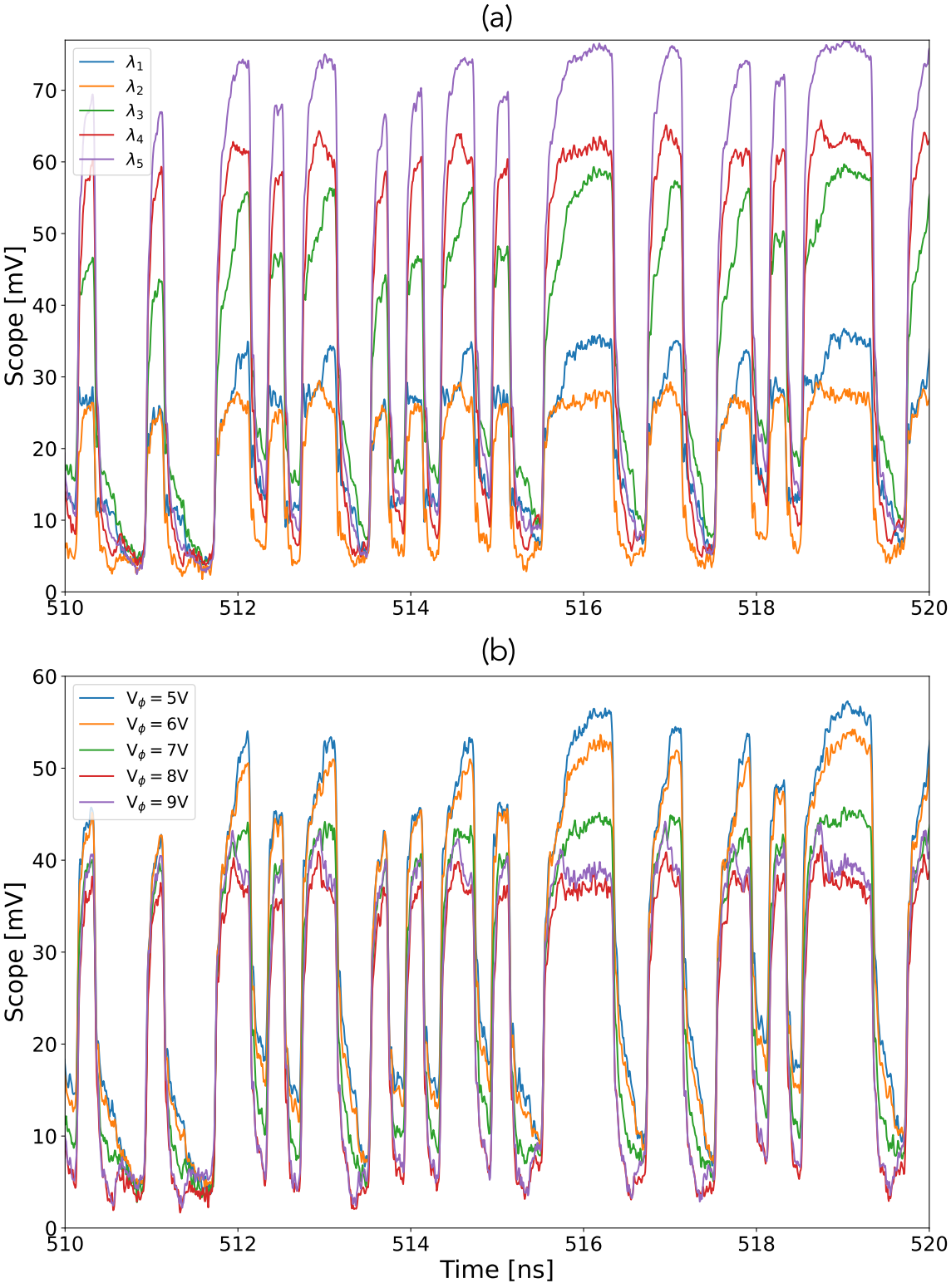

Fig. 6.11: Response of R188 to the input bitstream, using a TC. Varying dynamics for varying wavelengths and V_ϕ can be observed. Note especially the changing decay behavior.

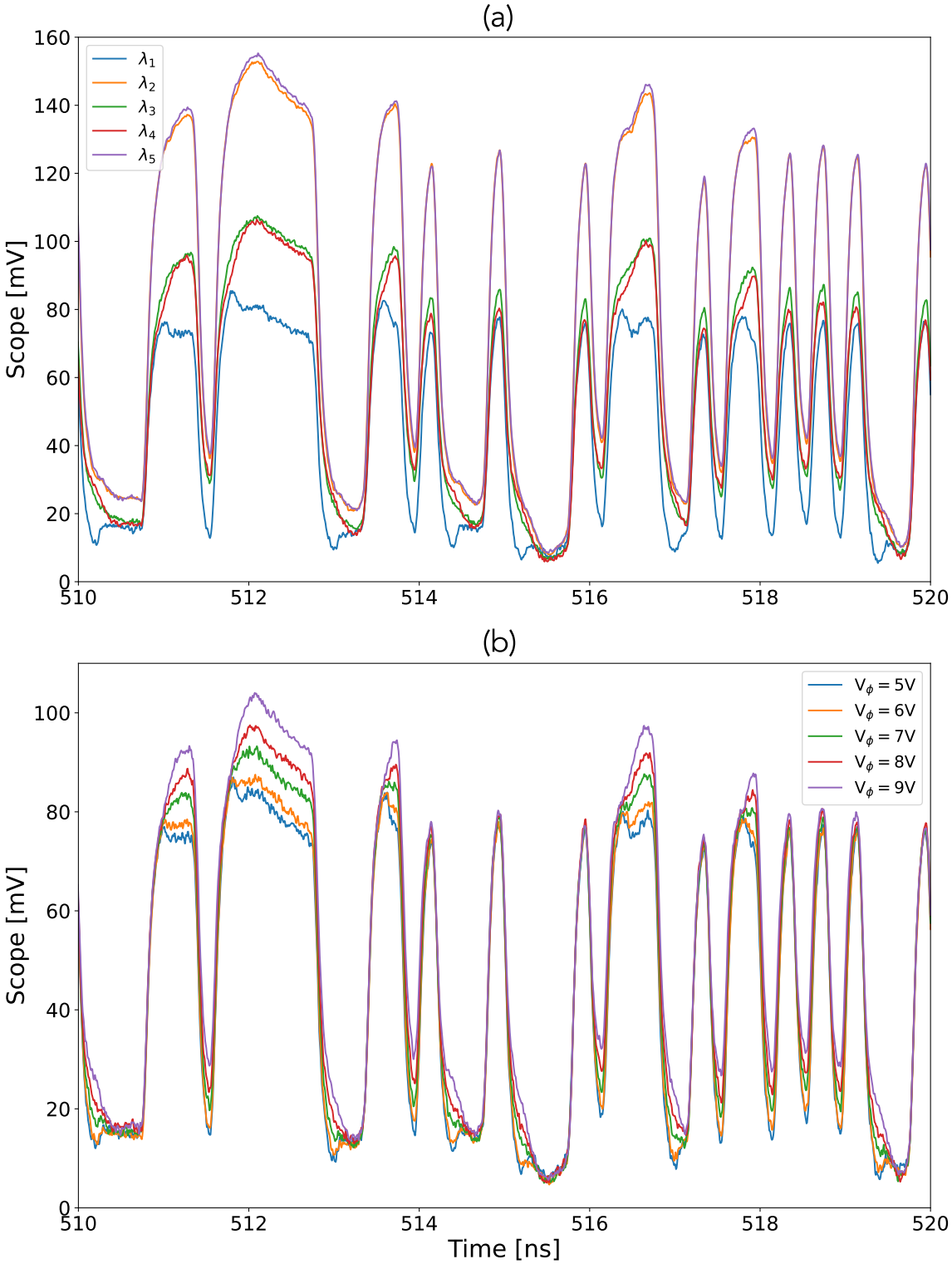

Fig. 6.12: Response of R188 to the input bitstream, using a TC. Similarly, varying dynamics for varying wavelengths and V_ϕ can be observed, although much less noisy.

The principle point of notice in the figures above is the changing amplitudes and decay times, signifying the different effective cavities being formed under the influence of the interference at the couplers, controlled by the heater.

6.4 Experimental demonstration

In this section, we attempt to train the reservoir on two tasks, the temporal bitwise XOR and the Santa Fe timeseries prediction. The responses obtained from R284 and R188 are used, with 6 voltage levels (4V - 9V) and 10 wavelengths (around 1550 nm) scanned per reservoir, yielding 60 different datasets for each. The input is mask-less and the clock cycle is 0.2 ns, corresponding to a data rate of 5 Gbit/s. For both tasks, ridge regression was used with the ridge parameter set to 10^{-4}.

6.4.1 XOR

The input is defined in a python script and exported as a .csv file to be sent to the AWG. The AWG has a maximum output sample rate of 64 GSa/s, which permits ~ 13 (~ 6) samples per input bit for a bit rate of 5 (10) Gbit/s. The input stream consists of 16384 NRZ bits, and is used to attempt the XOR-1 task. The BER threshold was scanned with 20 values from 0.05 to 0.95, and the best BER for each dataset is reported, as shown in Fig. 6.13. The obtained BER is very high, which means this task was not solved. This could be due to the 0 level being designated to 0 mV, which means a large portion of the gathered data is close to the noise floor of the system and/or limited by the photodiode sensitivity. This could be alleviated by setting a higher 0 level, while ensuring a sufficient swing between the two levels. Since the BER is already high when using all 16384 examples in the training set, a validation set was not used.

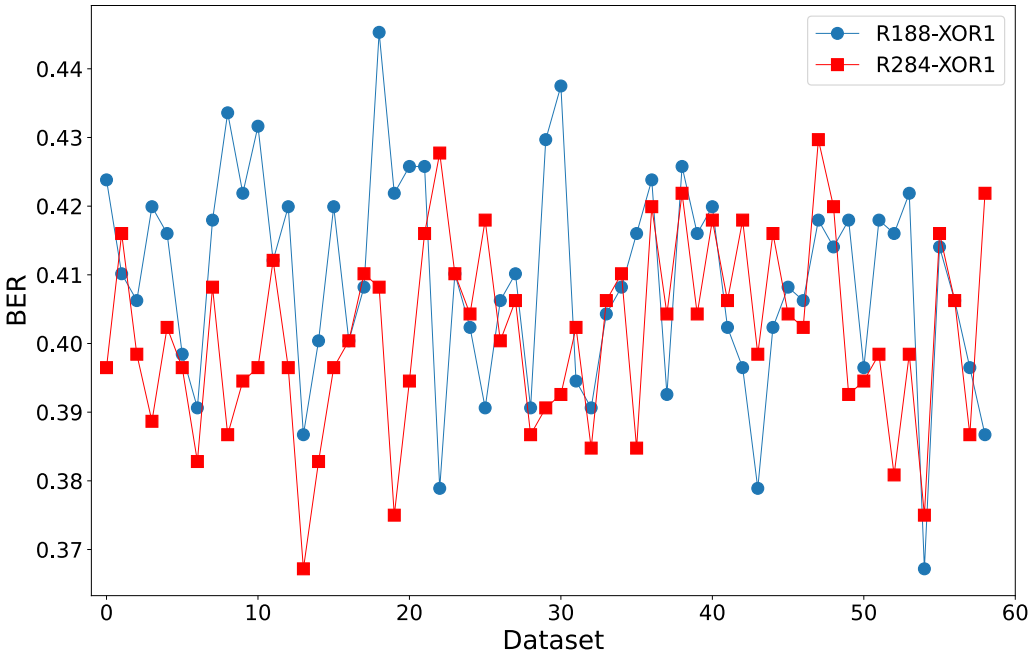

Fig. 6.13: BER obtained for the XOR task.

6.4.2 Santa Fe

The Santa Fe time series was also investigated. An input stream of 4000 training examples was employed, where 3500 where used for training and 500 for testing. In Fig. 6.14, the defined input and response are shown. We report excellent results on the 1-step ahead prediction, which is the commonly used version of this task. We also obtained good results for the 2-steps ahead prediction, while the performance was found to significantly degrade when considering 3 or more steps, as shown in Fig. 6.15 and Fig. 6.16.

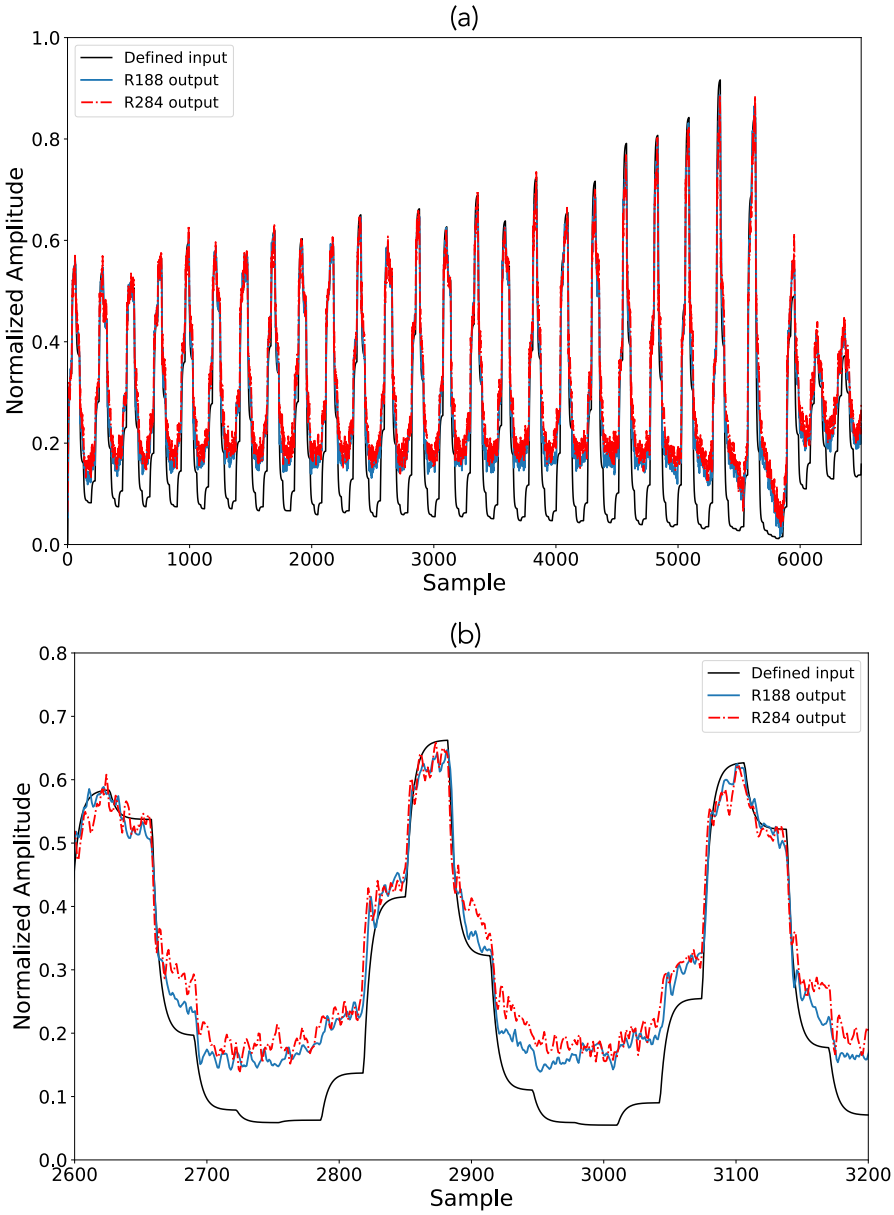

Fig. 6.14: (a) Zoomed out view and (b) zoomed in view of the Santa Fe input (software) and obtained responses from R188 and R284.

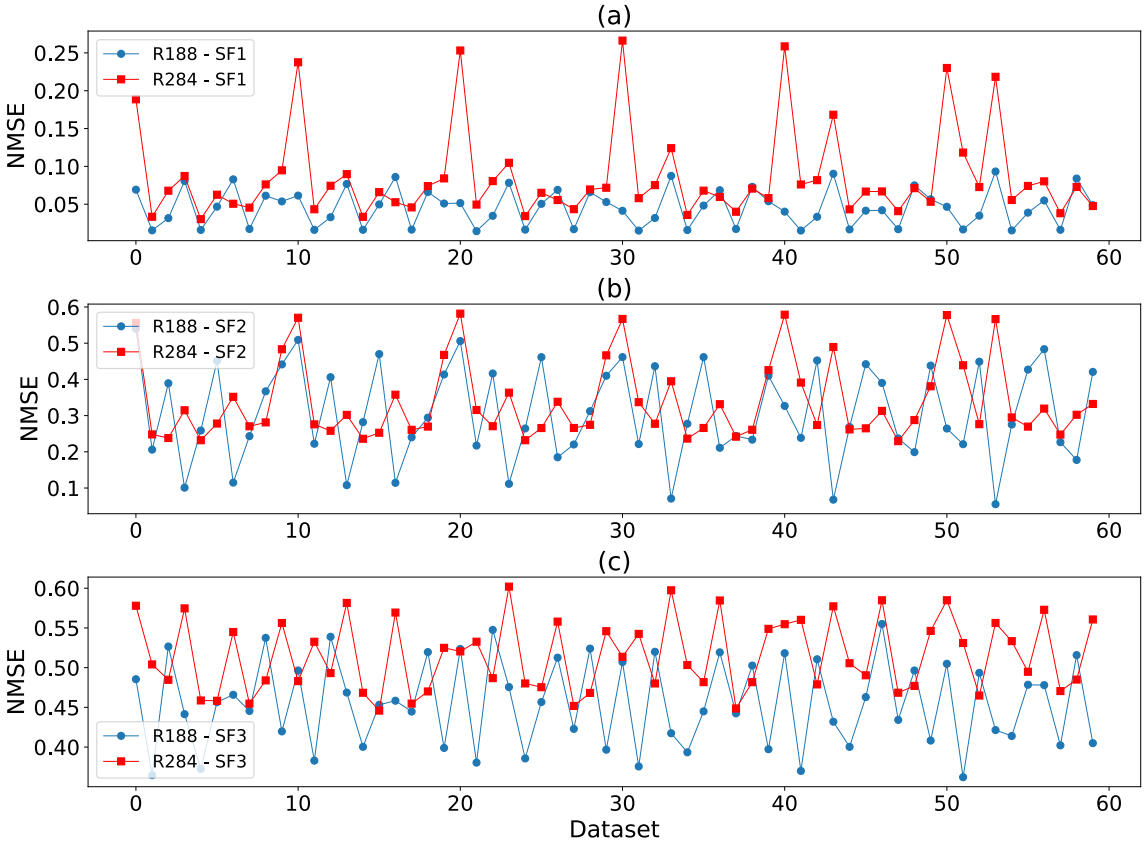

Fig. 6.15: NMSE obtained with the various datasets for the Santa Fe task for (a) 1 step ahead prediction, (b) 2 steps ahead prediction, and (c) 3 steps ahead prediction.

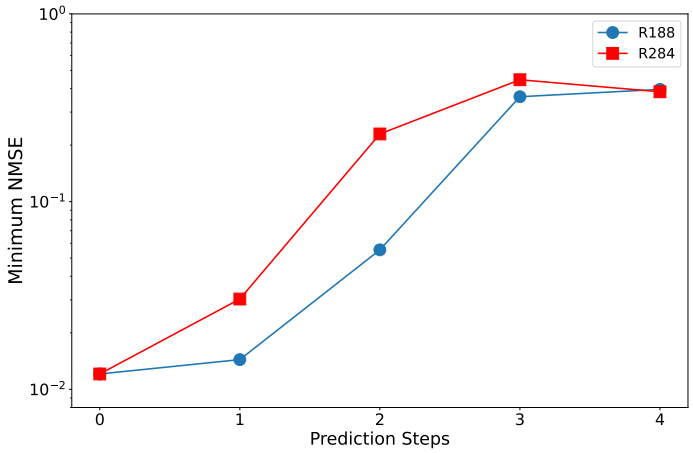

Fig. 6.16: Minimum obtained NMSE for R188 and R284 for various prediction steps, showing similar trends and performance for both reservoirs.

The 0^{th} prediction step is considered to ensure the temporal synchronicity between the training set and the obtained response. This ensures that the task is solved by the memory of the reservoir and not due to an already delayed input. Table 6.1 reports the minimum NMSEs obtained for the various prediction steps,

corresponding to Fig. 6.16. The achieved results are slightly better than simulations, and are among the best obtained in the relevant literature.

Table 6.1: Minimum obtained NMSE on the Santa Fe task (experimental)

Task	R188	R284
SF-1	0.01	0.03
SF-2	0.05	0.22
SF-3	0.36	0.44
SF-4	0.39	0.38

6.5 Conclusion

In conclusion, different instances of the proposed reservoir were designed, physically realized on the LNOI platform, and characterized for the relevant parameters. The results show the changing time dynamics for different V_ϕ, which confirms the operation principle of this architecture. The XOR task yielded a high BER, while the Santa FE time series was solved with excellent results. A fabrication based on the second layout (Fig. 6.3), which features shorter feedback waveguides, would allow higher data rates to be used, and possibly improve the results. Additionally, the use of a different cut of LNOI can minimize anisotropic effects and would further minimize spiral losses to values closer to those of straight waveguides (< 0.3 dB/cm), which would allow larger memory tasks to be solved.

Bibliography

[1] A. Boes, B. Corcoran, L. Chang, J. Bowers, and A. Mitchell, "Status and potential of lithium niobate on insulator (lnoi) for photonic integrated circuits," *Laser & Photonics Reviews*, vol. 12, no. 4, p. 1700256, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.201700256

Chapter 7

Conclusion

7.1 Summary

In this dissertation, the need for alternative computing paradigms was presented. The attractiveness of leveraging photonic integrated circuits towards that aim was discussed. The current trends in photonic neural networks were summarized and a detailed overview of the state of the art in photonic reservoir computing was provided. The main theme explored throughout this work is the minimization of the complexity of photonic RC schemes, while still achieving high performance. To that aim, we have launched multiple investigations in different directions. The first concerns proposing a novel integrated photonic RC architecture that is reconfigurable using a single element, in the form of a phase shifter, which controls two parameters simultaneously; the feedback strength and the feedback phase. Tuning the feedback strength in this manner does not require power to be coupled out of the system (lost), as is typically achieved by using an optical attenuator. The integration of the feedback waveguide is enabled by the low-loss LNOI platform or any other low-loss platform, such as silicon nitride. Our simplified design enabled good performance on some RC benchmark tasks, while at the same time significantly reducing the hardware complexity of the reservoir and auxiliary equipment. The second investigation concerns the further reduction of complexity by disposing with the input mask at the input layer and engineering the feedback length to prolong the dynamical system's transient response time, which is essentially the aim of the input mask. The reservoir's performance in the asynchronous regime was evaluated using a variety of task-independent metrics and benchmark tasks. The results showed comparable performance to the literature for the considered tasks. In summary, the results presented in this dissertation suggest that such minimal RC schemes can achieve on-par performance with schemes that either require costly auxiliary instruments,

or schemes that use a multitude of tunable parameters. The proposed implementations are less costly and more energy-efficient for a variety of tasks, especially if the inputs are already in the optical domain. Finally, the impact of bandwidth limitations and noisy readout on performance was examined across various tasks. It was concluded that the nature of the task determines its robustness to lower bandwidths (i.e. slower photodetectors). Finally, the experimental demonstration of the proposed design was presented, and the scheme was shown to solve the Santa Fe benchmark task with excellent results.

7.2 Perspectives

There are several key limitations of the proposed architecture, particularly at the proposed high speeds. The first is the need for high speed readout equipment to achieve higher processing speeds. This means high bandwidth photodetectors and driving/readout electronics, which are costly and energy consuming. Like other PIC-based schemes, our proposed scheme is limited to solving tasks with moderate memory requirements. For example, solving the NARMA-10 task with this scheme, especially when considering readout noise, becomes very difficult.

Large waveguide losses pose a significant challenge to this approach, restricting its implementation to inherently low-loss platforms. However, non-ideal conditions are inevitable and result in higher-than-expected losses. Although the system successfully solved one of the tasks, improved experimental performance over a wider variety of tasks could be achieved with reduced waveguide losses (< 0.5 dB/cm).

Furthermore, while we brushed on the topic of solving multiple tasks at the same time, further investigation is still needed. In particular, it would be interesting to investigate the use of multiple wavelengths, as could be done using an integrated micro-comb source, which is possible on the LNOI platform. Furthermore, the rich variety of optical nonlinearities accessible on LNOI can be explored and its RC performance compared against this minimalist scheme.

Finally, and not unique to this approach, photonic systems inherently suffer from wavelength drift due to refractive index changes caused by ambient temperature fluctuations. While compensation is feasible in a controlled lab environment, it becomes challenging in packaged systems for real-world applications. Several studies have investigated methods to mitigate this issue, as discussed in chapter 3, yet there remains significant potential for further improvement. The proposed architecture may handle this through either an FPGA or a co-packaged ASIC that implements a zero-calibration point algorithm to compensate for temperature drift. The ASIC

would take as input a portion of the detected output power, and accordingly adjust the phase shifter voltage to perform the compensation.

Appendix A

Synthèse en Français

A.1 Résumé

Cette étude concerne le calcul par réservoir à retard temporel, en anglais Time-Delay Reservoir Computing (TDRC) dans les plateformes de photonique intégré, en particulier la plateforme Lithium Niobate On Insulator (LNOI). Nous proposons une nouvelle architecture intégrée «tout optique », avec seulement un déphaseur comme paramètre modifiable pouvant atteindre de bonnes performances sur plusieurs tâches de référence de calcul par réservoir. Nous étudions également l'espace de conception de cette architecture et le fonctionnement asynchrone du TDRC, qui s'écarte du cadre plus courant consistant à envisager les ordinateurs TDRC comme des réseaux. En outre, nous suggérons d'exploiter le schéma tout optique pour se passer du masque d'entrée, ce qui permet de contourner la conversion Optique/Electronique/Optique (O/E/O), souvent nécessaire pour appliquer le masque dans les architectures TDRC. Dans des travaux futurs, cela pourra permettre le traitement de signaux entrants en temps réel, éventuellement pour des applications de télécommunication de pointe. Les effets de la lecture électronique de sortie sur cette architecture sont également étudiés. Aussi, nous suggérons d'utiliser la corrélation de Pearson comme une métrique nous permettant de concevoir un réservoir capable de traiter plusieurs tâches en même temps sur le même signal entrant (et éventuellement sur des signaux dans des canaux différents). Les premiers travaux expérimentaux menés à l'université RMIT sont également présentés. Par ces travaux, nous voulons étudier la performance de ces nouvelles architectures TDRC tout en ayant minimisant la complexité du matériel photonique. Pour cela on s'appuiera principalement sur les faibles pertes du LNOI qui permettent l'intégration du guide d'onde de rétroaction, et en utilisant uniquement l'interférence et la conversion d'intensité à la sortie (par le biais d'un photodétecteur) en tant que non-linéarité.

Cela constitue une base sur laquelle pourront s'appuyer de futurs travaux étudiant les gains de performance lorsque des non-linéarités supplémentaires sont prises en compte (telles que celles de la plateforme LNOI) et lorsque la complexité globale du système augmente par l'introduction d'un plus grand nombre de paramètres. Ces travaux portent donc sur l'exploration d'une approche informatique non conventionnelle particulière (TDRC), utilisant une technologie particulière (la photonique intégrée), sur une plateforme particulière (LNOI). Ces travaux s'appuient sur l'intérêt croissant pour l'informatique non conventionnelle puisqu'il a été démontré au fil des ans que les ordinateurs numériques ne peuvent plus être une solution unique, en particulier pour les applications émergentes telles que l'intelligence artificielle (IA). Le paysage futur de l'informatique englobera probablement une grande variété de paradigmes informatiques, d'architectures et de hardware, afin de répondre aux besoins d'applications spécialisées croissantes, tout en coexistant avec les ordinateurs numériques qui restent - du moins pour l'instant - mieux adaptés à l'informatique à usage général.

PARTIE I: CONTEXTE SCIENTIFIQUE

A.2 Introduction

L'essor de l'informatique au 20e siècle – partant des les matériaux de base jusqu'à l'échelle et l'étendue des applications - n'a pas été une simple percée technologique. Les progrès mathématiques du XIXe siècle avaient déjà transformé la logique philosophique en un langage formel qui a donné naissance à la logique symbolique, puis à l'algèbre de Boole, à la base de l'informatique numérique. En logique symbolique, une expression est évaluée pour sa valeur de vérité, produisant l'une des deux sorties possibles : Vrai ou Faux, qui peuvent en pratique être représentés comme les états « ON » et « OFF » d'un interrupteur. Les travaux pionniers de la théorie de la calculabilité dans les années 1930, catalysés par les travaux de Gödel sur la logique formelle, ont abouti à la machine de Turing : un ordinateur conceptuel qui manipule des symboles sur une longueur infinie de ruban suivant un tableau de règles. Cette machine simple peut mettre en œuvre, en principe, n'importe quel algorithme informatique. L'algèbre de Boole a également été largement utilisée par Claude Shannon pour développer la théorie de l'information, posant les bases de la théorie moderne de la communication.
D'un point de vue plus pratique, les progrès de la physique des semi-conducteurs et de la science des matériaux ont permis l'élaboration de la jonction p-n. Ainsi,

l'informatique moderne n'a pu voir le jour que quand les sciences théoriques et pratiques se sont rencontrées au cours de la première moitié du XXe siècle. Les ordinateurs à tube à vide ont été utilisés bien au-delà de la première moitié du XXe siècle. Cependant, ils étaient difficiles à entretenir en raison de pannes constantes (peu fiables) et fonctionnaient à des centaines de volts (gourmands en énergie). L'invention du transistor promettait (et plus tard réalisa) un changement d'état logique rapide, économe en énergie et fiable. Cela a permis aux ordinateurs d'augmenter leur nombre d'interrupteurs (grâce à la réduction de la taille des dispositifs et à l'intégration sur puce) et d'élargir le champ des applications, leur permettant d'être les machines performantes et omniprésentes que nous connaissons aujourd'hui.
L'incroyable succès de la microélectronique , stimulé par la loi de Moore, prévoyant que le nombre de transistors sur un circuit intégré doublerait environ tous les deux ans, a soutenu l'amélioration des performances informatiques pendant plus de 50 ans. Notons également les progrès continus au niveau des architectures des dispositifs permettant de réduire la taille des transistors à 2 nm (y compris Fin-Fet, GAAFET, etc.), mais au prix d'une augmentation des coûts de R&D et de fabrication. En outre, les ordinateurs d'aujourd'hui - qui utilisent essentiellement l'architecture connue sous le nom d'architecture « Princeton » ou plus communément de modèle « von Neumann » - souffrent de ce qu'on appelle un goulot d'étranglement.
Celui-ci se situe au niveau du transfert de données entre les unités distinctes de mémoire et de calcul. C'est pourquoi l'exploration de paradigmes informatiques non conventionnels a suscité un intérêt croissant, allant du calcul en mémoire (les opérations de calcul et la mémoire se situent dans la même unité), au calcul stochastique, en passant par le calcul neuromorphique et le calcul par réservoir, pour ne citer qu'eux. Le calcul optique a également été utilisé comme un moyen d'exploiter les avantages à travailler avec la lumière plutôt qu'avec des signaux à fréquence radio (RF), ou simplement comme base pour la mise en œuvre de certains des paradigmes susmentionnés.
La photonique peut faire la différence dans le domaine de l'informatique, notamment en termes de :

- **Largeur de la bande de fréquence** : la photonique fonctionne avec des ondes lumineuses (des centaines de THz) alors que l'électronique numérique fonctionne avec des ondes RF (GHz). La différence de largeur de bande atteint alors 5 ordres de grandeur. Un parallélisme massif peut être obtenu grâce au multiplexage des fréquences, les peignes en fréquence pouvant générer de la lumière dans environ 10^7 lignes de fréquence distinctes ; les données de chaque ligne pouvant alors être traitées en parallèle.

- **Dynamique à faible perte**: les photons peuvent se propager avec des pertes négligeables dans les configurations en espace libre et des pertes très faibles dans certaines plates-formes intégrées (par exemple, les pertes démontrées sur les plates-formes LNOI ou avec SiN sont de l'ordre de 0,2 dB/cm dans la bande optique C). C'est une autre raison pour laquelle, on s'attend à ce que les interconnexions optiques jouent également un rôle important dans la communication électronique intra-puce.

L'utilisation de plates-formes photoniques pour des applications en calcul est donc très prometteuse, en particulier pour les paradigmes émergents qui promettent une augmentation de la vitesse et de l'efficacité du calcul, par exemple dans le domaine du traitement du signal ou de l'intelligence artificielle.

A.3 Quelques éléments de photonique

Cette section présente un bref aperçu de la photonique, en mettant l'accent sur les circuits intégrés photoniques (PICs). De nombreuses recherches ont été menées pour trouver des matériaux adaptés aux dispositifs et circuits photoniques intégrés. Pour la plupart des plates-formes, la photolithographie est utilisée de la même manière qu'en microélectronique, pour la gravure et le dépôt de matériaux. Mais, contrairement aux circuits électroniques où les électrons existent facilement à l'intérieur des matériaux, les circuits photoniques ont besoin de sources de lumière, généralement obtenus par émission stimulée en générant des paires électron-trou dans un matériau à gain. La lumière a également besoin d'être détectée, la détection se faisant dans un semiconducteur par la combinaison de ces paires électron-trou.
D'un point de vue historique, la conception des dispositifs a précédé de quelques décennies les réflexions au niveau du circuit ou de systèmes. Cela a entraîné la spécialisation des plateformes pour ces différents dispositifs, par exemple : le phosphure d'indium (InP) pour les lasers, l'arséniure de gallium (GaAs) pour les détecteurs et le niobate de lithium LiNbO3 pour les modulateurs à grande vitesse. Chacun de ces dispositifs a été fabriqué selon des procédés de fabrication différents qui sont fondamentalement incompatibles avec ceux utilisés pour les autres dispositifs.
Si on se focalise maintenant au niveau de la conception et de la simulation des dispositifs et des systèmes photoniques il est nécessaire de revenir à l'optique fondamental. La physique de la lumière peut être décrite par différentes théories, qui suivent l'évolution de notre compréhension de la nature de la lumière. Par ordre chronologique, ces théories sont l'optique géométrique, l'optique des ondes scalaires, l'optique des faisceaux, l'optique de Fourier, l'optique électromagnétique

et l'électrodynamique quantique. Ces différentes théories de la lumière peuvent être utilisées pour modéliser la propagation de la lumière dans les circuits photoniques, selon que l'on souhaite une simulation au niveau du dispositif ou au niveau du système, et selon les types de phénomènes pris en compte. Pour les implémentations photoniques intégrées au niveau du système, on modélise la plupart du temps la lumière grâce à l'optique ondulatoire à ondes guidées (par exemple, la théorie des modes couplés). Les simulations au niveau du circuit effectuées dans cette thèse sont réalisées à l'aide de la méthode de la matrice S. Un exemple simple est présenté ci-dessous pour montrer comment cela fonctionne.
Considérons une source laser, décrite comme un champ électromagnétique cohérent et monochromatique. Le champ électrique E de cette source de fréquence $\omega = 2\pi f$, en supposant une onde plane (c'est-à-dire sans dépendance spatiale), peut être décrit par son amplitude A et sa phase initiale ϕ comme suit :

$$E_{\text{src}}(t) = A \exp(i\omega t + \phi) \tag{A.1}$$

Cette lumière peut être couplée à une puce photonique intégrée par divers moyens, le plus souvent par l'intermédiaire d'un coupleur à réseau. Lors du couplage, le champ est donné par :

$$E_{\text{gc}} = \gamma_{\text{gc}}(\lambda, \theta) E_{\text{src}} \tag{A.2}$$

où γ_{gc} est le coefficient de couplage du CG qui dépend de la longueur d'onde du laser λ et de l'angle θ entre la fibre et la normale. Après avoir été couplé dans le guide d'onde de la puce, il se propage sur une certaine distance, ce qui donne:

$$E_{\text{wg}} = E_{\text{gc}} \exp(i\beta L_{\text{wg}}) \exp(-\alpha_{\text{wg}} L_{\text{wg}}) \tag{A.3}$$

où L_{wg} est la longueur du guide d'ondes [m], α_{wg} est la perte du guide d'ondes [m^{-1}], et $\beta = 2\pi n_{\text{eff}}/\lambda$ [m^{-1}] est la constante de propagation du mode guidé avec l'indice de réfraction effectif n_{eff}. L'indice effectif décrit en première approximation l'indice que le champ subit lors de sa propagation dans le guide d'ondes sachant qu'il ne se déplace pas uniquement dans le cœur du guide. Cela signifie qu'une partie du mode s'échappe dans le milieu environnant, se faisant, le champ rencontre un indice « effectif » qui tient compte de cet effet. D'autres éléments composants les circuit photoniques peuvent être décrits de manière similaire qu'il s'agisse d'éléments actifs, passifs, linéaires ou non-linéaires. Nos simulations sont réalisées dans le domaine temporel en raison de l'aspect dynamique du calcul TDRC.

A.4 A propos du calcul par réservoir

Comme le calcul neuromorphique , le calcul par réservoirs (RC) est un type d'approche d'apprentissage automatique supervisé. Initialement connu sous le nom de « Echo-state networks » (ESN), le RC a été utilisé dans des implémentations logicielles au début des années 2000 pour contourner les problèmes de formation de grands réseaux neuronaux récurrents (RNN). Différents systèmes physiques peuvent être exploités pour mettre en œuvre les RC, notamment dans les domaines de l'électronique et de la photonique. L'ESN est un système dynamique non linéaire qui présente des caractéristiques particulières, à savoir

- la propriété d'écho-état (ESP) ; le système permet des réverbérations d'entrées passées qui diminuent avec le temps (mémoire déclinante),
- l'expansion de la dimensionnalité, généralement obtenue grâce à une fonction d'activation non linéaire.

L'ESP est une condition qui nécessite l'oubli des entrées passées, de sorte que le réservoir ne dépende pas de ses conditions initiales. Il s'agit d'une caractéristique essentielle du RC car elle permet au réservoir de répondre de la même manière à la même entrée. Cette mise en correspondance cohérente est essentielle pour qu'un modèle formé sur la sortie du réservoir puisse se généraliser. L'expansion de la dimensionnalité permet de séparer plusieurs classes avec de simples classificateurs linéaires, même pour des tâches qui nécessiteraient normalement des réseaux neuronaux plus vastes et plus profonds. L'idée principale est qu'en projetant les données dans un espace de dimension supérieure, il existe davantage d'hyperplans capables de bien séparer les différentes classes de données, ce qui permet d'obtenir une précision élevée.
Tout schéma de RC se compose des trois couches suivantes :

- La couche d'entrée : où les données sont potentiellement prétraitées puis injectées dans le réservoir,
- La couche réservoir : composée de nœuds non linéaires connectés les uns aux autres,
- La couche de sortie : où les états du réservoir sont collectés au moyen d'un mécanisme de lecture.

Chaque couche est reliée à la suivante par une série de connexions pondérées. La distinction entre l'informatique à réservoir des réseaux neuronaux temporels similaires, tels que les RNN, réside dans le fait que les poids entre la couche d'entrée et

la couche de réservoir peuvent être définis et fixés, de même que les interconnexions entre les nœuds non linéaires, alors que dans les RNN, toutes les connexions doivent être optimisées. Ainsi, dans le RC, le système s'entraîne uniquement sur la couche de sortie, ce qui signifie que seules les méthodes des moindres carrés sont nécessaires pour trouver les poids optimaux.
Le RC peut être réalisé spatialement ou temporellement, dans le cas des TDRC, si le réseau consiste en des multiplexés dans le temps. Dans ce dernier cas, il s'agit d'utiliser un seul nœud non linéaire et de l'échantillonner dans le temps pour obtenir les états souhaités du nœud. Notre travail s'est alors focalisé sur la conception de ces TDRC utilisant des circuits photoniques. Ce sujet a été exploré depuis 2011, et un chapitre détaillé sur l'état de l'art aborde les innovations les plus notables dans ce domaine.

PARTIE 2: CONTRIBUTIONS SCIENTIFIQUE

A.5 Calcul par réservoir photonique de complexité minimale

Notre première contribution s'appuie sur la littérature existante et sur le thème contemporain du calcul par réservoir dans les circuits intégrés photoniques, plutôt qu'en espace libre. Notre étude montre comment un dispositif simple (illustré sur la Fig. A.1), composé de guides d'ondes passifs et d'un déphaseur chauffant (pour la reconfigurabilité), peut déjà être aussi performant que certains schémas plus complexes. La non-linéarité se retrouve alors dans la non-linéarité de conversion d'intensité fournie au photodétecteur. Les performances ont été mesurées sur une variété de tâches standard de référence. Pour ce qui est de son fonctionnement, l'élément chauffant modifie la longueur du chemin optique dans le guide d'ondes inférieur, ce qui entraîne une modification de la dynamique de sortie (illustré sur la Fig. A.2). Une fois la dynamique souhaitée trouvée, le circuit peut être entraîné à résoudre des tâches de référence avec une grande précision, telles que les tâches XOR bit à bit (Fig. A.3), Mackey-Glass et Santa-Fe, et la capacité de mémoire (Fig. A.4). Nous avons également montré qu'il était possible que la rétroaction soit intégrée à la puce, grâce aux faibles pertes de la plate-forme LNOI. Cette étude montre en fin de compte ce qu'un simple schéma TDRC peut faire, et servira de référence lorsque des schémas plus complexes seront envisagés.

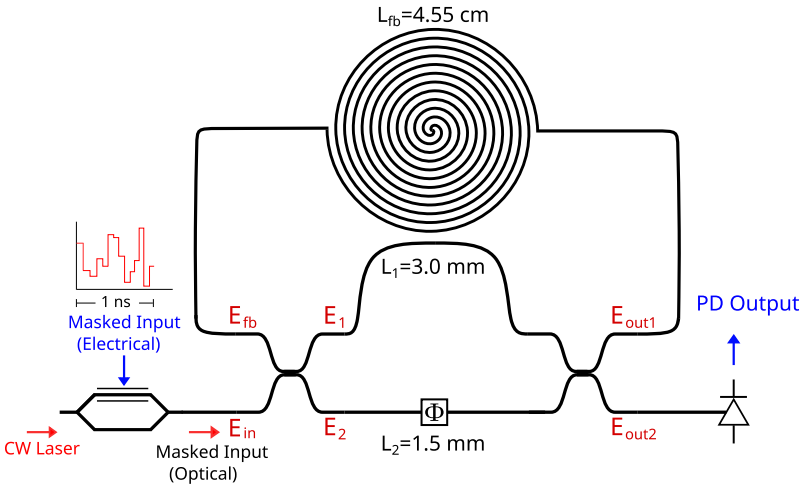

Fig. A.1: Architecture de calcul par réservoir photonique reconfigurable de complexité minimale.

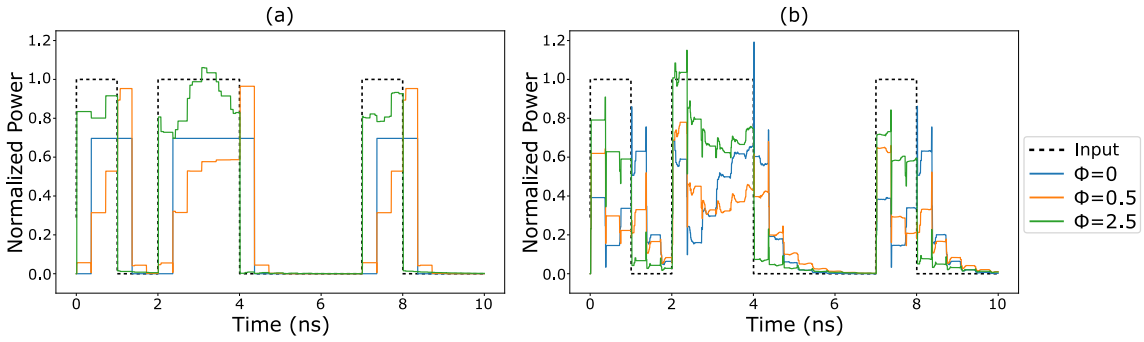

Fig. A.2: Sortie du système pour différents déphasages (tension de chauffage)

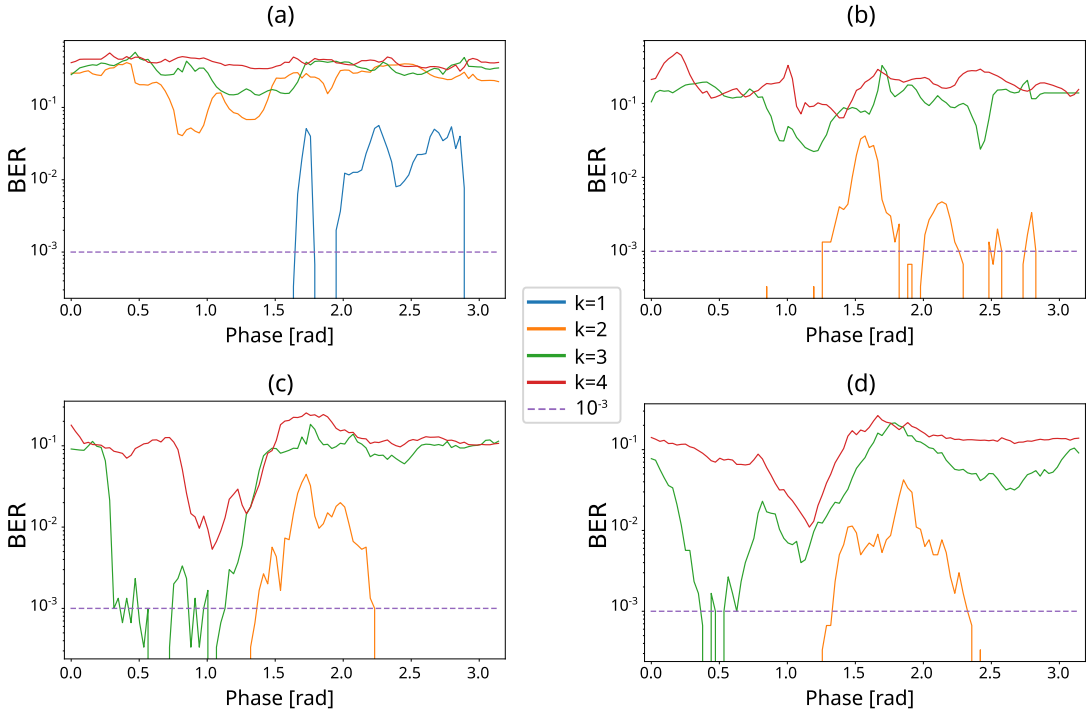

Fig. A.3: Performances du XOR pour différentes profondeurs de mémoire, sous différentes valeurs de déphasage.

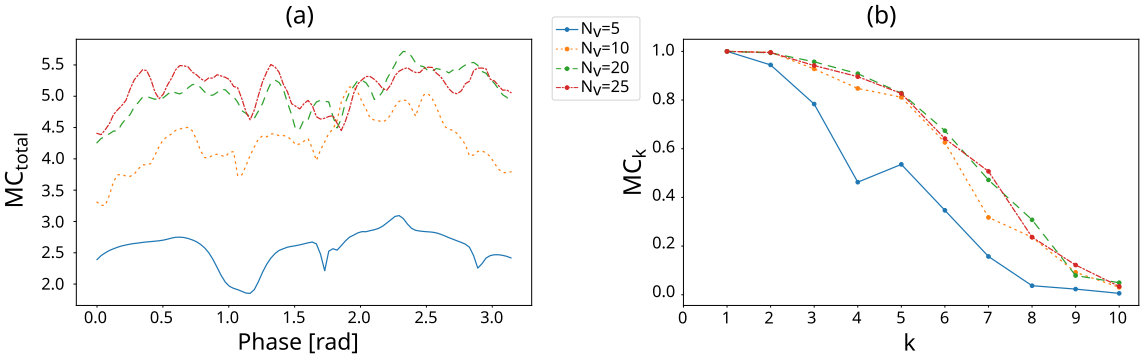

Fig. A.4: (a) Capacité totale de la mémoire et (b) capacités individuelles maximales

A.6 Calcul par réservoir photonique asynchrone sans masque

L'une des principales limites du TDRC se situe dans le multiplexage temporel des nœuds limitant la vitesse de traitement du système. Cette limitation est généralement accentuée par l'utilisation d'un masque d'entrée, qui réduit encore la vitesse de traitement effective du système. Ce masque empêche le système de tomber dans un état stable pendant la durée de l'échantillon d'entrée. De plus, ce masque ajoute des contraintes en termes de complexité et de consommation d'énergie puisqu'une conversion O/E/O peut être nécessaire pour appliquer le masque. Dans notre étude, nous explorons la possibilité d'utiliser le réservoir sans le masque d'entrée en utilisant la longueur du guide de rétroaction de telle manière à remplir le même rôle que le masque (comme illustré sur Fig. A.5). Le système n'a alors plus besoin de masque d'entrée accélérant ainsi son fonctionnement. Nous avons testé le réservoir sur une série de mesures ne dépendant pas de de la tâche choisie, ainsi que sur quelques tâches de référence. Les résultats numériques montrent de bonnes performances sur ces différentes tâches de référence (Fig. A.6), comparable à d'autres schémas dans la littérature (utilisant un masque).

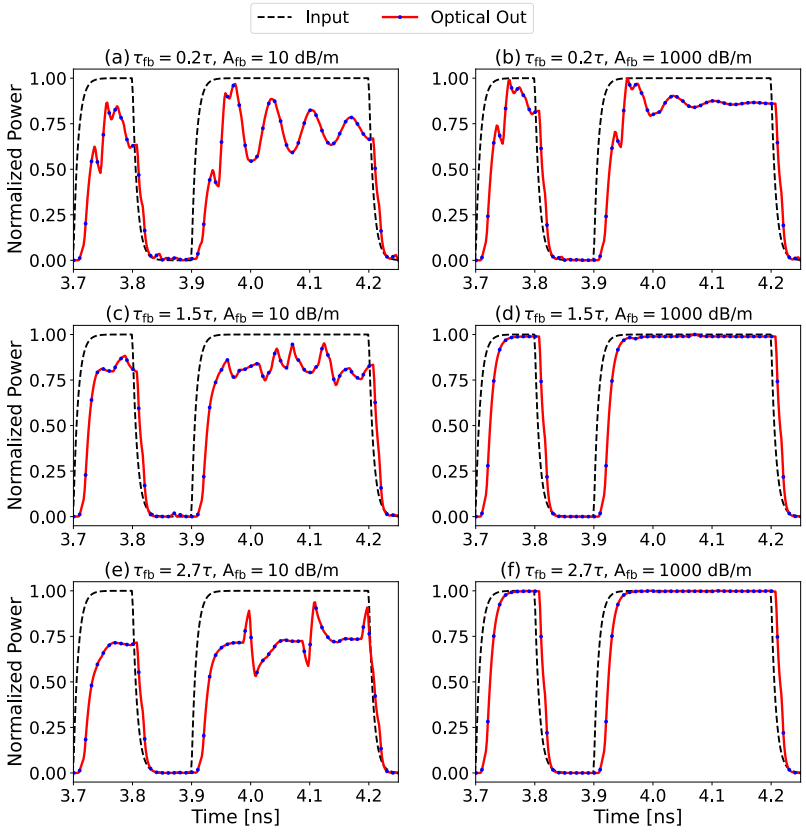

Fig. A.5: Des longueurs de rétroaction plus courtes permettent au signal d'interagir avec lui-même plusieurs fois, ce qui permet d'avoir le même effet qu'un masquage de l'entrée. Cela étant possible grâce aux faibles pertes de la plateforme LNOI.

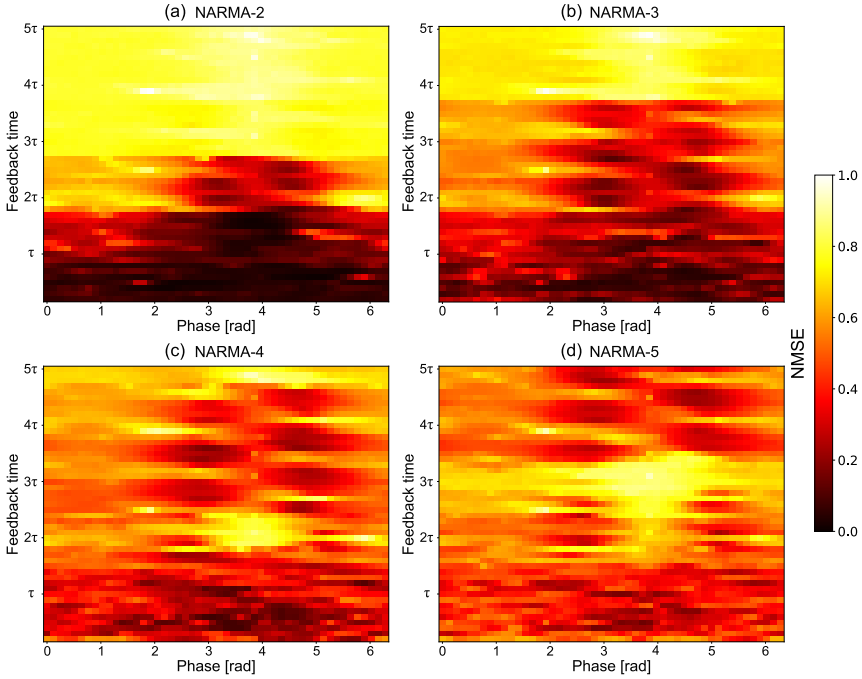

Fig. A.6: L'erreur quadratique moyenne normalisée (NMSE) est présentée pour une variété de tâches avec différents temps de rétroaction et déphasages.

A.7 Effet d'une lecture à largeur de bande limitée

Nous examinons également les effets de la lecture électronique sur les performances de cette architecture. La lecture joue un rôle central, étant l'intermédiaire entre le réservoir optique et le signal électronique de sortie. La lecture, composée d'une photodiode et des circuits associés, imprime des imperfections sur le signal qui sera utilisé pour l'apprentissage du modèle des moindres carrés. Pour cela, nous considérons une photodiode simple et un modèle de bruit, et nous tenons compte de la conversion analogique-digital (ADC) effectuée par l'oscilloscope numérique. Pour l'ADC nous avons utilisé une seule acquisition avec un nombre effectif de bits inférieur pour vérifier que le schéma peut fonctionner dans ces conditions pour quelques tâches. Les ENOB considérés sont généralement utilisés dans les oscilloscopes haut de gamme.

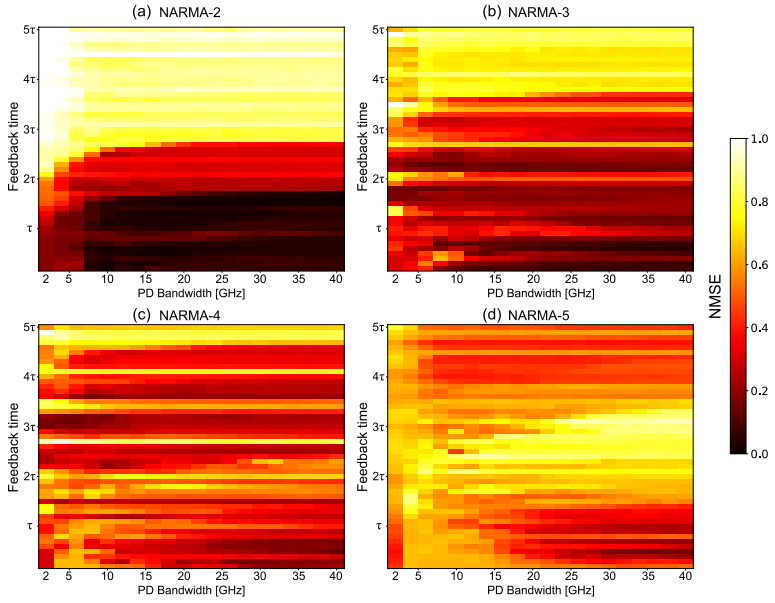

Fig. A.7: Les performances sont enregistrées pour différentes largeurs de bande du photodétecteur et différents temps de rétroaction, montrant une grande tolérance pour des largeurs de bande plus faibles pour certaines tâches.

Table A.1: Le système est capable de résoudre certaines tâches même en considérant une routine d'acquisition unique avec un nombre effectif de bits inférieur.

Task	5 bits	6 bits	7 bits	8 bits
XOR-1	0.0	0.0	0.0	0.0
XOR-2	8.3×10^{-3}	6.6×10^{-4}	0.0	0.0
NARMA-2	0.13	0.08	0.05	0.04
NARMA-3	0.29	0.21	0.16	0.12
SF	0.46	0.27	0.22	0.17

A.8 Démonstration expérimentale

Dans cette dernière étude, nous avons fabriqué et testé expérimentalement notre schéma de calcul par réservoir à complexité minimale sans masque sur la tâche de référence Santa-Fe, en obtenant d'excellents résultats. La puce a été fabriquée par l'équipe de fabrication du Centre intégré de photonique et d'applications du RMIT. Le design de la puce (comme illustré sur Fig. A.8) a été réalisé à l'aide d'IPKISS et du kit de développement de processus (pdk) d'InPAC. Différentes règles de conception ont été vérifiées manuellement pour s'assurer que :

- Un minimum de courbes se chevauchant avec les lignes de la grille de lithographie par faisceau d'électrons (EBL) ;
- la distance minimale entre les guides d'ondes soit respectée ;
- L'espacement des plots DC (200 µm) corresponde au pas de la sonde DC ;
- L'espacement des contacts RF corresponde au pas de la sonde RF.

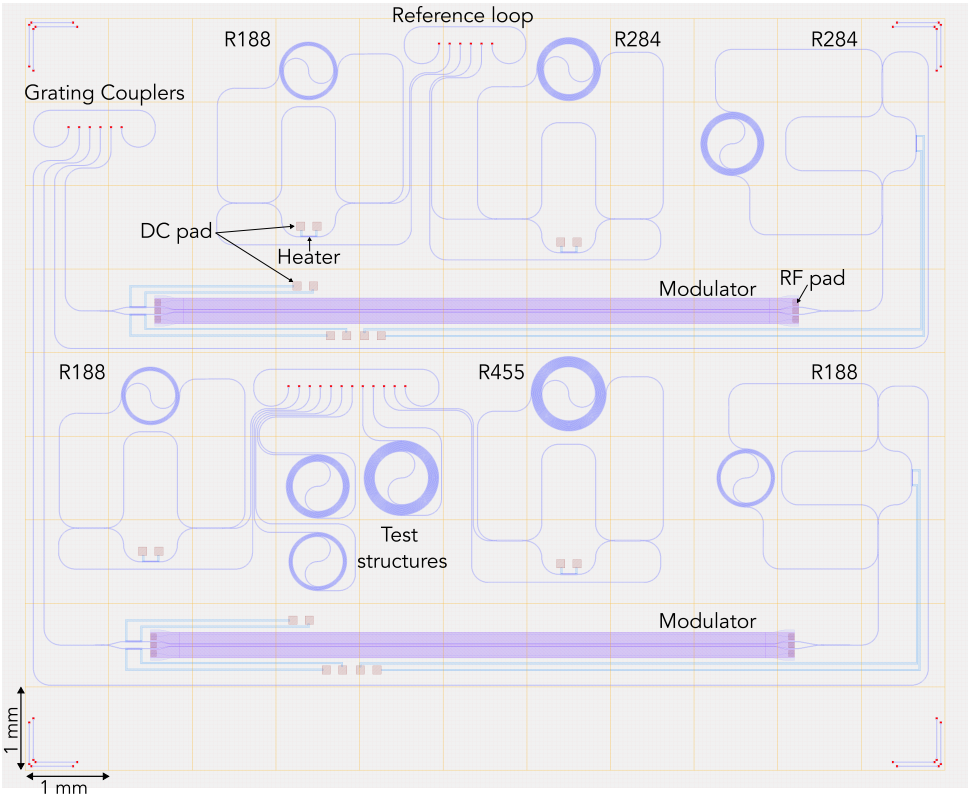

Fig. A.8: Disposition de la puce générée à l'aide d'IPKISS. Plusieurs dispositifs de test ont été réalisés avec des longueurs de rétroaction variables, ces dispositifs possèdent des modulateurs intégrés, ainsi que des structures de test.

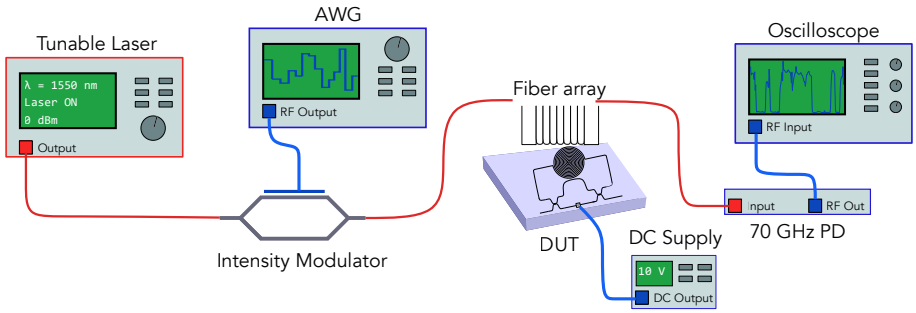

Fig. A.9: Illustration du dispositif expérimental.

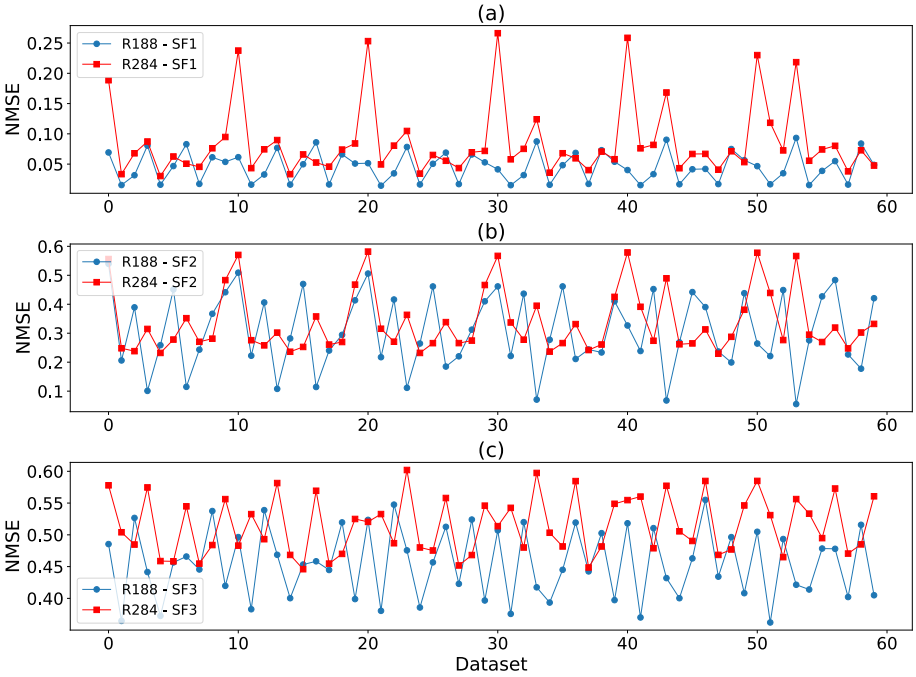

Fig. A.10: NMSE obtenu avec les différents ensembles de données pour la tâche de Santa Fe pour (a) la prédiction 1 pas, (b) la prédiction 2 pas, et (c) la prédiction 3 pas.

Le dispositif expérimental utilisé est illustré sur la Fig. A.9. Les résultats expérimentaux de la tâche de Santa Fe sont présentés sur la Fig. A.10 pour tous les ensembles de données obtenues par balayage en longueur d'onde et en tension au déphaseur. Nous avons considéré deux dispositifs, celui utilisant une boucle de rétroaction de 2,84 cm (R284) et celui utilisant une boucle de rétroaction de 1,88 cm (R188). Normalement, la tâche de Santa Fe est considérée pour la prédiction à un pas, mais nous avons également documenté les résultats pour la prédiction à 2 pas et à 3 pas. Chaque pas représente des exigences différentes concernant la mémoire du système de calcul par réservoir. On peut constater que R188 est plus performant que R284, en raison des pertes de rétroaction plus faibles rencontrées.

A.9 Résumé et conclusions

Au cours de ces travaux nous avons cherché à minimiser la complexité des schémas de RC photonique et avons obtenus de bons résultats. Pour cela, nous avons lancé plusieurs recherches. La première consiste à proposer une nouvelle architecture de RC photonique intégrée qui est reconfigurable à l'aide d'un seul élément de déphasage, qui contrôle deux paramètres simultanément : la force de rétroaction et la phase de rétroaction. Se faisant, la modification de la puissance de la rétroaction ne nécessite pas de couplage de puissance hors du système, comme c'est généralement le cas en utilisant un atténuateur optique. Les plateformes à faible perte (LNOI, SiN) permettent l'intégration d'un guide d'onde de rétroaction. Notre conception simplifiée a permis d'obtenir de bonnes performances pour certaines tâches de référence en matière de RC, tout en réduisant considérablement la complexité matérielle du réservoir et de l'équipement auxiliaire. La deuxième étude porte sur une réduction supplémentaire de la complexité en éliminant le masque d'entrée au niveau de la couche d'entrée et en concevant la longueur de la rétroaction pour obtenir une dynamique qui reste transitoire pendant la perturbation de l'entrée, ce qui était essentiellement l'objectif du masque d'entrée. Les performances du réservoir dans le régime asynchrone ont été évaluées à l'aide d'une variété de mesures (indépendantes de la tâche effectuée mais également sur de tâches de référence). Les résultats ont montré des performances comparables à celles de la littérature pour les tâches étudiées. En résumé, les résultats présentés dans cette thèse suggèrent que de tels schémas RC minimaux peuvent atteindre des performances comparables à celles des schémas qui nécessitent des instruments auxiliaires coûteux ou des schémas qui utilisent une multitude de paramètres. Enfin, les effets de la lecture à bande passante limitée ont été étudiés, en montrant comment le bruit et la limitation de la bande passante affectent les performances du système pour diverses tâches. Nous avons conclu que la nature de la tâche détermine la robustesse des performances du système pour des largeurs de bandes plus faibles. Enfin, la démonstration expérimentale a été présentée et le système a permis de résoudre la tâche de référence de Santa Fe avec d'excellents résultats. L'architecture proposée présente plusieurs limites essentielles, en particulier liées aux échelles de temps. La première est la nécessité d'un équipement de lecture à grande vitesse pour atteindre des largeurs de bande plus élevées. Cela implique des photodétecteurs et des oscilloscopes à large bande passante, qui sont coûteux et encombrants. Comme d'autres systèmes basés sur des puces, il est limité à la résolution de tâches requérant une mémoire moyenne ou faible. Par exemple, il est très difficile de résoudre la tâche NARMA-10 avec ce système, surtout si l'on tient compte du bruit de lecture. En

outre, bien que nous ayons abordé la question de la résolution de plusieurs tâches en même temps, des recherches plus approfondies sont encore nécessaires. Il serait notamment intéressant d'étudier l'utilisation de plusieurs longueurs d'onde, comme on pourrait le faire en utilisant une source de peigne en fréquence intégrée, ce qui est possible sur la plateforme LNOI. Aussi, il serait intéressant d'exploiter la grande variété de non-linéarités optiques accessibles sur le LNOI pour créer des systèmes de RC et comparer leur performance au schéma minimaliste. Enfin, et cela n'est pas spécifique à notre design, les systèmes photoniques souffrent intrinsèquement de décalages de la longueur d'onde due aux changements d'indice de réfraction provoqués par les changements de température. S'il est possible de compenser ce phénomène en laboratoire, cela n'est pas évident dans un environnement conditionné pour des applications réelles. Bien que plusieurs études aient exploré les moyens d'atténuer ce problème, des améliorations sont encore possibles. L'un des moyens de résoudre ce problème à l'aide de l'architecture proposée pourrait être l'utilisation d'un FPGA ou d'un ASIC cointégré qui met en œuvre un algorithme d'étalonnage pour compenser la dérive de la température. L'ASIC prendrait en entrée une partie de la puissance de sortie détectée et ajusterait en conséquence la tension du déphaseur pour effectuer la compensation.