Université du Québec
Institut national de la recherche scientifique
Centre Énergie Matériaux Télécommunications

RESOURCE-EFFICIENT SMART OPTICAL SIGNAL PROCESSING ENABLED BY NONLINEAR FIBER OPTICS AND MACHINE-LEARNING APPROACHES

Par

Bennet Fischer

Thèse présentée pour l'obtention du grade de
Philosophiæ Doctor, Ph.D.
en sciences de l'énergie et des matériaux

Jury d'évaluation

Examinateur externe	Prof. Pablo Bianucci Concordia University
Examinateur externe	Prof. Bhavin Shastri Queen's University
Examinateur interne	Prof. Shervin Vakili INRS-EMT
Directeur de recherche	Prof. Roberto Morandotti INRS-EMT

Remerciements

I would like to acknowledge Prof. Roberto Morandotti for the opportunity to conduct this exciting cutting-edge research in his group, and the academic freedom and support throughout my entire time.

I would like to thank Prof. José Azaña for his valuable feedback and discussions on signal processing techniques, as well as Prof. Tiago Falk for his valuable insights and input for machine-learning methods.

In particular, I would like to thank Dr. Mario Chemnitz for his thorough support, fruitful discussions and valuable feedback in all conducted experiments and especially providing his theoretical knowldedge in supercontinuum generation.

I am also thankful for the collaboration with Yi Zhu and his knowledge in digital signal processing as part of the INRS COVID grant.

I am particular thankful for the discussion with Dr. Piotr Roztocki and Benjamin MacLellan and their valuable feedback. I also thank our lab technician Robin Helsten for his equipment knowledge, lab expertise and thorough help with sometimes stubborn equipment.

I am also grateful to Prof. Sai Chu and Prof. Brent Little for providing the integrated samples used in this thesis.

In general, I thank the ultra-fast optical processing group and collaborators for discussions and the great time I had.

Finally, I like to thank my family and friends for their continuous thorough support throughout my time in Canada.

Résumé

En optique, l'utilisation croissante de techniques et d'algorithmes d'apprentissage machine (ML) impact de façon importante une variété de systèmes. Celle-ci a mené aux développements de nouvelles techniques d'analyse, de fonctionnalités inaccessibles à la conception classique de système et des réseaux de neurones optiques (ONN). Ces nouveaux concepts forment ensemble la photonique intelligente. Cette thèse, qui se compose de deux parties, portera tout d'abord sur l'étude de l'utilisation d'algorithmes d'optimisation guidée pour la réutilisation d'infrastructures existantes en route pour de nouvelles tâches ("L'apprentissage machine "pour" la photonique). Ensuite, elle abordera l'étude d'une nouvelle approche d'analyse de signaux optiques basée sur l'optique non linéaire ("L'apprentissage machine-photonique").

Dans la première partie, la mise en œuvre d'un conformateur d'impulsions autonome et reconfigurable basé sur une puce photonique interférométrique existante est étudiée pour la première fois. Le concept utilise la synthèse de cohérence temporelle sur puce pour générer des formes d'onde arbitraires dans le régime de la picoseconde. Pour le contrôle, une lecture simple est mise en œuvre à l'aide d'une méthode d'échantillonnage entièrement optique basé sur un mélange dégénéré à quatre ondes. De plus, deux algorithmes couramment utilisés sont comparés et évalués pour leurs performances en matière de mise en forme d'impulsions picosecondes. Enfin, le caractère évolutif et la précision de cette approche sont explorés par des simulations.

Dans la deuxième partie, la mise en œuvre d'un ONN basé sur une fibre standard, disponible sur le marché, est examinée. Cette approche exploite une dynamique non linéaire riche (c'est-à-dire la fission de solitons) qui se produit pendant la génération de supercontinuum (c'est-à-dire la génération de lumière non linéaire à large bande). Le système mis en œuvre est évalué selon différents tests de performance d'apprentissage machine et ses performances sont comparées à celles d'autres approches d'ONN existants. Enfin, le système est testé sur un exemple réel sous la forme d'une tâche de classification de la COVID-19.

Les résultats de cette thèse indiquent que la photonique non linéaire peut être un outil essentiel pour le contrôle de systèmes adaptatifs, les tâches d'inférence neuromorphique et le domaine de la photonique intelligente en général. De plus, il a été démontré que les composants existants peuvent en effet être réutilisés, en atteignant des performances remarquables et en fournissant ainsi une nouvelle stratégie pour l'implémentation opportune de la photonique intelligente.

Mots-clés photonique intelligente, mise en forme d'impulsions picosecondes, interférométrie sur puce, échantillonnage entièrement-optique, algorithme génétique, optimisation par essaims de particules, génération de supercontinuum, fission de solitons, machine d'apprentissage extrême, photonique neuromorphique

Abstract

The increasing use of machine learning (ML) techniques and algorithms in optics are having an large impact over a variety of systems, enabling new processing techniques, novel system functionalities beyond classical system design and development of optical neural networks (ONNs). These novel concepts are commonly referred to as *smart* photonics. This thesis consists of two parts: In the first part, the use of guided optimization algorithms is investigated towards the re-utilization of existing infrastructure en route for new tasks ("Machine Learning (for) Photonics"), and in the second part an entirely new approach or optical signal processing based on nonlinear optics is investigated ("Photonic Machine Learning").

In the first part, the implementation of an autonomous, reconfigurable pulse-shaper based on an existing interferometric photonic chip is investigated for the first time. The concept utilizes on-chip temporal coherence synthesis for the generation of arbitrary waveforms in the picosecond regime. For monitoring, a straight-forward readout is implemented utilizing an all-optical sampling scheme based on degenerate four-wave mixing, and two commonly used algorithms are compared and evaluated for their performance for picosecond pulse-shaping. Finally, the scalability and accuracy of this approach is explored via simulations.

In the second part, the implementation of an ONN based on a standard off-the-shelf fiber is examined. This approach exploits the rich nonlinear dynamics (i.e., soliton fission) occurring during supercontinuum generation (i.e., nonlinear broadband light generation). The implemented system is tested on various ML benchmarks and performance is compared to other existing ONN approaches. Lastly, the system is tested on a real-world example in form of a COVID-19 classification task.

The findings of this thesis indicate, that nonlinear photonics can be an essential tool for adaptive system control, neuromorphic inference tasks and the smart-photonic field in general. Moreover, it was demonstrated that existing components can indeed be re-utilized, achieving remarkable performance and thus providing a new strategy for timely *smart* photonic implementations.

Keywords smart photonics, picosecond pulse-shaping, on-chip interferometry, all-optical sampling, genetic algorithm, particle swarm optimization, supercontinuum generation, soliton fission, extreme learning machine, neuromorphic photonics

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Artificial intelligence (AI) and machine learning (ML) algorithms are currently revolutionizing several applications fields, promising the advent of a fourth industrial revolution [1, 2]. Unsurprisingly, ML techniques and algorithms are rapidly becoming the backbone of many everyday service and technology products, enabling novel devices and consumer services, such as smart-home devices and assistants, or autonomous cars. Moreover, over the past years these approaches have also become an essential tool in science and industry, for example in physical sciences such as microscopy and ultrafast optics [3, 4], the development of new materials and drugs [5–8] or medical diagnostics [9, 10], to name only a few.

Given the ever increasing amount of data that accompanies such a revolution, a research and application field that can undoubtedly benefit from this development is optical signal processing [11, 12]. Indeed, AI concepts (incl. ML algorithms and guided optimization techniques [13]) can help improve performance of telecommunication lines via advanced digital post-processing or ideally more efficient all-optical signal processing hardware [11, 12, 14, 15]. The merger of AI and ML techniques with photonics is often referred to as *smart* photonics, and can be usually divided into two main branches, as illustrated in Fig. 1.1:

- "*Machine learning for photonics*", describing the use of machine-learning and optimization techniques for photonics

- "*Photonic machine learning*", describing the implementation of optical machine learning systems.

However, it should be noted that this division is not mutually exclusive and overlaps exist between different concepts exist.

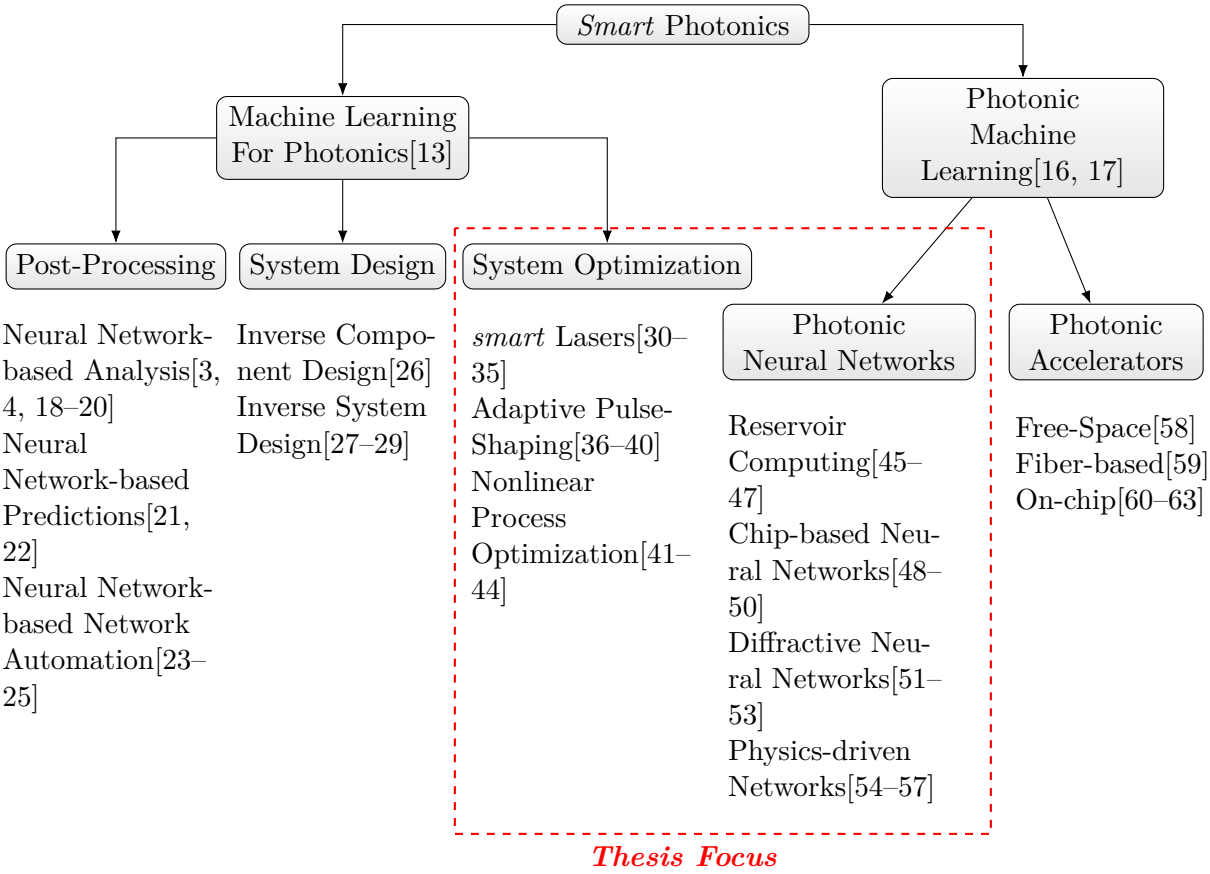


Figure 1.1: Overview of the smart photonics field including selected application examples, consisting of two main branches: machine-learning (for) photonics and photonic machine-learning as well as a few examples for each category. The dashed box denotes the focus of this work: Smart System Optimization in Chapter 2 and Photonic Neural Networks in Chapter 3.

Indeed, promising and rapid progress has been made in the *smart* photonic field [13], however implementations so far often utilize highly application-specific system designs or complex integrated photonic chips. Although these systems and photonic chips can reach remarkable performances, the development of such complex systems is often prohibitively resource, time, and cost expensive, hence only available to a handful of groups. Such advancement clearly limits timely progression, which in turn can be potentially detrimental for the research field as a whole [64]. Consequently, novel approaches for the *smart* photonics field are required, focusing on re-utilizing existing photonic

components and devices towards new applications, and utilizing optical effects available to such components, both in addition to specifically designed systems. This thesis aims at proposing and demonstrating a new strategy in the field. It poses the question to what extent off-the-shelf optical components and non-purpose optical devices can be reutilized and empowered by the use of machine learning algorithms in combination with nonlinear optics.
Strikingly, while nonlinear optical effects are extensively used in classical signal processing [65] and have been early on proposed in the context of photonic machine learning [66], they have been rarely exploited towards *smart* photonic implementations thus far. This in turn raises several questions:

- **Can nonlinear optics provide a versatile tool for the field of *smart* photonics?**
- What can machine learning do for photonics and what can photonics do for machine learning?
  - Can existing integrated components repurposed towards new application fields using machine learning approaches?
  - Can existing off-the-shelf photonic components reutilized towards novel (neuromorphic) computing concepts?

Optical signal processing may benefit two-fold from the implementation of ML techniques in conjunction with nonlinear optical effects: Firstly, they can enable a wider use of existing components by utilizing *smart* algorithms [13, 67], and secondly, enable novel, often unusual, optical processing schemes such as neuromorphic (i.e., brain-inspired) computing systems [68, 69], which are able to process information highly efficiently. Hence, ML methods and optimization algorithms can provide novel methodology and also improve existing systems beyond their intended use and design [15, 70].
However, a big challenge for out-of-the-lab applicability of such smart devices are efficient and inexpensive monitoring solutions. Indeed, most *smart* photonic implementations so far utilize standard, but often expensive and complex diagnostics such as full-field reconstruction [39, 71], ultra-fast or real-time oscilloscopes [34, 72], or radio-frequency analyzers [30, 31]. These techniques are also often limited in terms of bandwidth ($\ll 100\,\mathrm{GHz}$), hence necessitating the development and implementation of efficient, inexpensive but suitable on-the-fly measurement solutions. Such techniques are of high interest, as they are able to achieve light-speed processing rates without inefficient electro-optic

conversion in compact footprints [14, 65, 70]. Nonlinear optics in compact waveguides offers elegant and efficient solutions to overcome this limitation.

Indeed, the realization of novel computing schemes is a topic of high interest in recent years given the increasing environmental impact of digital AI and ML approaches [73]. In particular, machine learning approaches are often carried out using classical computer architecture, which can be very inefficient due to lacking parallelism or energy requirement for data storage and transfer, in turn leading to a poor use of resources in such software-based approaches. For example, training state-of-the-art neural networks, such for natural language processing, can produce as much CO_2 as several flights from New York to San Francisco [73]. This development clearly limits the scope and scale of the problems to which machine learning can be applied to, e.g., in the context of 'big data' or portable applications. Photonic neuromorphic (brain-inspired) hardware may play an important role in overcoming current limitations by providing energy-efficient [69], ultra-fast information processing at the speed of light as it can perform certain mathematical operations fully passively or at extremely low power (femto to attojoule per operation [70]). Moreover, the use of photonics enables a high degree of parallelism exploiting various degrees of freedom available to photonics [74, 75]. While numerous promising approaches have been developed [16, 76–78], many solutions are often overlooked in favor of existing conventions from fields of which new methods are borrowed (e.g., here computer science), often resulting in in designed solutions with ever growing complexity rather than in the search for innovation and uncommon principles [79].

Thesis Structure

The goal of the thesis is to investigate the utilization of nonlinear optical effects towards the implementation of *smart* photonic devices for efficient and scalable all-optical signal processing. In particular, the work focuses on exploring the use of nonlinear optical techniques for optical monitoring and inference, with emphasis on repurposing and reutilizing existing integrated and fiber-based components. Towards this, an optical sampling monitoring solution based on degenerate four-wave mixing is investigated to enable an autonomous on-chip pulse-shaper for the picosecond domain (Chapter 2). Subsequently, a simple yet powerful in-fiber implementation of an all-optical neural network emulator based on soliton fission is presented and studied (Chapter 3). Both distinct applications have been enabled by the use of very similar equipment and methods, i.e., highly non-

linear fibers, programmable filters, and guided optimization routines, ultimately demonstrating the economical reconfigurability of common equipment. Each chapter contains a separate introduction into the research field with a detailed literature review, and the introduction of necessary scientific background to comprehend the systems under study, as well as an intermediate conclusion.

In detail, Chapter 2 focuses on the realization of an autonomous optical pulse-shaper operating the picosecond domain based on the temporal coherence synthesis approach. For this implementation, an existing optical chip is repurposed utilizing two common guided optimization techniques such as genetic algorithms and particle swarm optimization. A key functionality for the successful implementation is an all-optical sampling scheme based on degenerate four wave-mixing in a highly nonlinear fiber, which allows for straight-forward measurement of the different optical waveforms. Finally, the performance of the pulse-shaper is investigated in detail and a comparison between two commonly used optimization algorithms is performed.

Chapter 3 focuses on the implementation of an all-optical, in-fiber, transient neural network based on the extreme-learning machine framework. The system is first described on a theoretical level and implemented using only off-the-shelf fiber components, re-utilizing developed methodology from Chapter 2. The system uniquely exploits complex soliton fission dynamics as a computational resource in an efficient and compact design. The performance of the system is subsequently demonstrated performing several neural network benchmark tasks, as well as a real-world application in the form of a COVID-19 classification task.

Finally, Chapter 4 provides a brief summary of the presented work. In an outlook, shortcomings of the current approaches in the field are discussed and potential solutions presented.

Chapter 2

Smart Optimization for Photonics

Summary: This chapter focuses on the use of guided optimization methods and techniques in combination with optical systems to allow increased performance or functionality beyond *in silico* design only. In detail, the use of different meta-heuristic optimization techniques, namely the well known genetic algorithm (GA) and the less popular particle swarm optimization (PSO), in the context of all-optical signal processing is discussed. The difference between both algorithms will be investigated under different aspects, such as performance or convergence time, via simulations and experiments using an integrated interferometric platform to realize an autonomous optical pulse-shaper operating in the *ps*-domain.

The results presented in this chapter have been published in the following:

Journals:

B. Fischer, M. Chemnitz, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "Autonomous on-chip interferometry for reconfigurable optical waveform generation," Optica **8** (10), pp. 1268-1276 (2021).

Conferences:

B. Fischer, M. Chemnitz, B. Wetzel, P. Roztocki, B. MacLellan, C. Reimer, B. Little, S. Chu, E. Viktorov, D. Moss, M. Kues, J. Azaña, A. Pasquazi, M. Peccianti, and R. Morandotti, "User-friendly, reconfigurable all-optical signal processing with integrated photonics," 3rd URSI AT-AP-RASC 2022, Paper We-D-AI-PM4-1 (2022), *invited* .

B. Fischer, M. Chemnitz, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. Little, S. Chu,

D. Moss, J. Azaña, and R. Morandotti, "Scalable, Autonomous On-Chip Picosecond Pulse-Shaping Enabled by Smart Optimization," Conference on Lasers and Electro-Optics - CLEO, Paper STh2E.5 (2022).

M. Chemnitz, **B. Fischer**, B. MacLellan, P. Roztocki. R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "On-Chip Temporal Coherence Synthesis for Classical and Quantum Waveform Processing," Conference on Lasers and Electro-Optics - CLEO, Paper SF3G.2 (2022).

B. Fischer, M. Chemnitz, B. MacLellan, P. Roztocki, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "Picosecond Pulse Shaping via On-Chip Interferometry," Frontiers in Optics Conference, Paper FM1B.5 (2021).

M. Chemnitz, **B. Fischer**, B. MacLellan, P. Roztocki. R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "All-optical sampling for Adaptive On-Chip Picosecond Pulse-Shaping," OSA Nonlinear Optics Topical Meeting, Paper NF2A.1 (2021).

Patents:

B. Fischer, M. Chemnitz, B. Maclellan, P. Roztocki, J. Azaña, Y. Jestin, and R. Morandotti, System and method for self-adaptive picosecond pulseshaping, US 63/201,378 (patent pending).

'*Machine-learning (for) Photonics*' focuses largely on exploiting machine-learning concepts and guided optimization techniques to increase the performance or providing new functionalities to existing systems and concepts. Some examples for this advancement include the use of neural networks in optical microscopy to increase the resolution or help with identifying features [3, 18, 20], for the reconstruction of ultra-fast pulses [4], the understanding and prediction of ultra-fast and nonlinear dynamics [21, 22, 80], or customized nonlinear light generation schemes [42, 43, 81] using neural network or guided optimization strategies, to name only a few. Besides these more application-oriented implementations, machine-learning and optimization techniques are also increasingly investigated towards the (inverse) design of optical components [26, 82–87] and can be used to predict even complete experimental systems [27–29], offering innovative and often non-intuitive designs, otherwise difficult to obtain from pure theoretical and analytical considerations. Indeed, the optimization of system parameters using algorithms such as the genetic algorithm [88] are nowadays widely used in experimental implementations directly impacting the system performance as opposed to the aforementioned examples which are largely post-processing techniques. In the following, the imple-

mentation of system optimization schemes and techniques as illustrated in Fig. 1.1 will be discussed in detail.

Figure 2.1 shows a general feedback control system commonly used in conventional system control [89], as well as for smart photonic system implementations. However, contrary to classical systems, the optimization is mediated by computer algorithms rather than PID (proportional, integral, derivative) controllers, which can be difficult to tune (e.g., using methods such as Ziegler-Nichols tuning [90]).

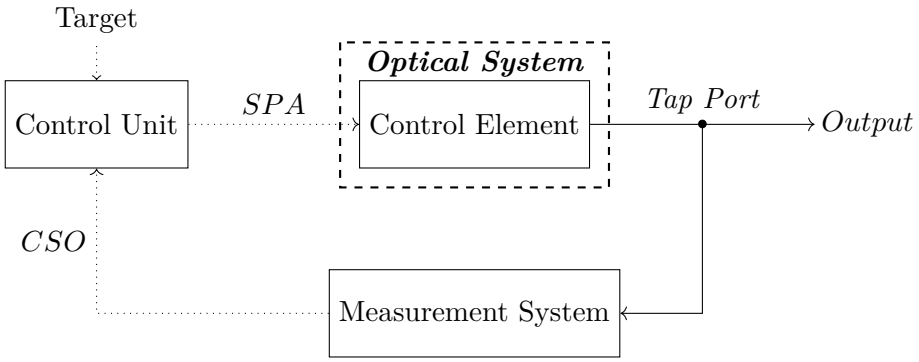


Figure 2.1: Illustration of a general adaptive control system for smart-photonic implementations. The control unit executes an algorithm to determine the required feedback signal SPA (system parameter adjustment) for the control element inside the optical system (dashed box) based on the current system output CSO (current system output) and the desired target signal. Solid and dotted lines represent optical and electrical paths, respectively.

Generally, any optical system requires three different elements to be used towards active (i.e., beyond advanced post-processing) 'smart' applications:

- Tunable control element inside the optical system that will *deterministically* change the system output depending on an electrical signal provided by the digital control unit
- Readout (measurement) of the feature that will be optimized
- Digital control unit to mediate the optimization or training, e.g., computer, field-programmable gate array (FPGA), microcontroller

Basically, any system containing reconfigurable or electrical addressable components can be used for the realization of a 'smart' photonic systems. The control element is in most cases a personal computer, but applications requiring, e.g., energy efficiency can be also realized using microcontrollers or field-programmable gate arrays (FPGAs) depending on the complexity of the chosen

task [58, 91]. The electrical addressable elements can be widely available devices and components such as spatial light modulators (SLM) [36, 92, 93], electric polarization controllers (EPC) [34], (pump) laser diodes [30, 32], fiber heaters [94], or more sophisticated designed components such as deformable mirrors [71], concatenated on-chip interferometers [40, 42] or complex integrated circuits [95].

However, to efficiently implement such schemes in a robust and sufficiently performant way (e.g., in terms of resolution, speed, etc.) detection arrangements for the respective optimization target feature (e.g., temporal or spectral optical power, temporal waveform profiles) are indispensable. Demonstrations so far include the use of standard optical spectrum analyzers (OSAs) [30, 42, 92, 96], autocorrelators [32], radio-frequency spectrum analyzers [30, 96], dispersive Fourier transform and frequency-to-time mapping techniques [34, 35], or straight-forward power measurements (including signals from second- and third-order nonlinearites) [36, 72, 97, 98]. Corresponding applications are, for example, the optimization of temporal pulse duration to achieve ultra-short pulses [36, 37, 97], adaptive phase control of ultra-short pulses [39, 71], the optimization of nonlinear propagation in highly-nonlinear fibers for customized spectral outputs [41–44, 83], free-space beam- and wavefront shaping [99, 100], or the realization of 'smart'- or 'intelligent' lasers'[30–35, 101, 102].

Especially the development of 'smart lasers' have gained a lot of interest over the past years as the use of optimization algorithms allow to efficiently navigate large parameter spaces during mode-locking procedure. The most prevalent approach is based on EPCs [101] inside the active fiber cavities, where the optimization algorithm controls the polarization in a nonlinear-polarization rotation mode-locking scheme. Besides, also other feedback approaches such as spectral filtering [92] or gain modulation [32] have been investigated. Overall, these specific implementations have led to several works demonstrating that besides the 'automatic' mode-locking approach, optimization algorithms can be used to customize the spectral output [34], achieve different operational regimes [30, 31] or pulse configurations [32], or can be used to generate specific outputs for applications in nonlinear optics such as breather solitons[1] [103, 104].

[1] see Chapter 3 for more details on soliton dynamics

2.1 Optimization Algorithms

In general, optimization algorithms can be used for various tasks in photonic implementations ranging from inverse component design, over 'smart' devices, to the training of optical neural networks (see also Chapter 3 for more detail). Often, optimization algorithms are preferred due to their simplicity yet sufficient performance compared to, for example, deep learning approaches such as deep reinforcement learning [105]. Therefore, the primarily focus in most of the photonic community is on two algorithms[2]: The genetic algorithm (GA) for experimental implementations, and the particle swarm optimization (PSO) algorithm for inverse design. The reason of this split is two-fold. On the one hand, the GA was already developed around the 1970's [107, 108], while the PSO is a more recent concept dating back to 1995 [109]. Thus, for the GA, many issues regarding the algorithms performance have been already solved, by implementing sophisticated operations such as multi-point cross-over or inversion techniques [110], before the introduction of the PSO. Therefore, one of the first experimental implementation of an optimization algorithm in optics dates back to 1997 by Baumert et al. [36] using a genetic algorithm for pulse-duration optimization. On the other hand, the PSO as a more recent invention faced several drawbacks such as premature convergence or insufficient accuracy [111]. This was only solved recently by implementing methods such as the adaptive PSO [112–114], where parameters can adapt during the optimization resulting in significantly better performance. Especially, in inverse design, the PSO has found a widespread use as it features less computational complexity which can drastically improve convergence for large-scale or complex problems [82, 84–87]. But it is also of increasingly interest for experimental implementations in recent years given its simplicity and performance [40, 99, 115, 116].

2.1.1 Genetic Algorithm

The genetic algorithm (GA) was developed around 1970 by the group of John Holland at the University of Michigan [108] and is inspired by the evolution of genes in nature.
The underlying principle of the genetic algorithm is the exploitation of functionality that is inherent to evolution [88], such as the creation of off-springs of an initial population in favor of improving the biological system, i.e., the objective of the system. A user-defined objective function[3] can be

[2] other methods such as the Rosenbrock algorithm [106] are also used in photonic systems [31] but not as wide-spread

[3] an objective function can be either designed to minimize (referred to as cost or loss) or maximizing (referred to as reward or fitness) a given function

constructed almost arbitrarily to solve specific problems, however parameter dependencies should be avoided in favor of multi-objective optimization approaches. Through the iterative generation and testing of different gene sets (populations), the algorithm is capable of finding optimal points in high-dimensional search spaces.

Fig. 2.2 illustrates the basic working principle of the GA.

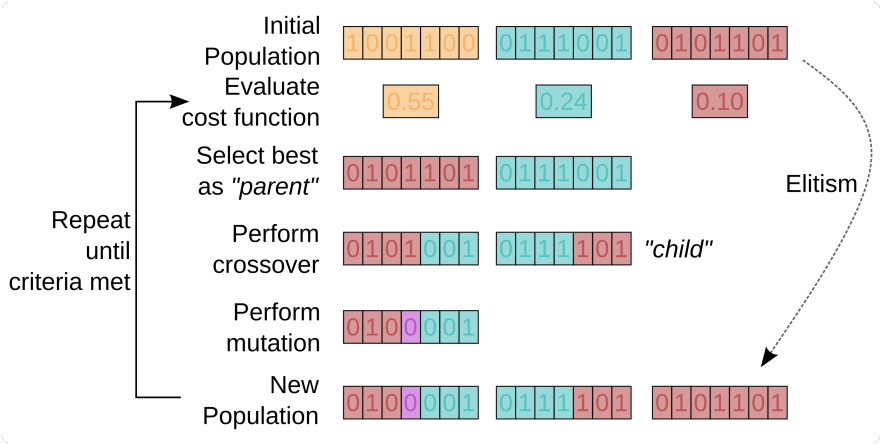


Figure 2.2: Illustration of the GA principle. The initial population is evaluated for a given objective or cost function, and then the best performing individuals are selected for subsequent genetic operations. The operations are repeated until the stopping criteria is met.

At the beginning, a random population is generated where the problem parameters are represented as individual genes[4]. Subsequently, each population is evaluated for the user-defined cost or objective function and the best performing individuals are selected as "parents" for the genetic operations if the termination criteria has not been met. Based on the chosen cross-over operation (e.g., single-point or multi-point [88, 110]) new off-springs ("children") are generated. The cross-over is essentially interchanging the genes between the selected parent individuals[5]. However, the simple interchanging of genes cannot guarantee a high population diversity (i.e., sufficient exploration of the search space), and therefore additional operations such as mutations are necessary. After generating the children, each new individual has a low probability to undergo mutation where one gene is randomly flipped, which ensures genetic diversity of the population. Further, best-performing individuals of the initial population can 'survive' (instead of being replaced by a child) - a feature commonly referred to as 'elitism'. With these newly generated individuals, the algorithm now iterates and re-evaluates the cost function, repeating the evolutionary operations for a specified number of iterations or until a user-defined termination criteria is fulfilled.

[4]for simplicity a binary encoding is shown, but also real-valued GA implementations are possible [88]
[5]additional operations such as inversion are also possible to increase diversity in the populations [110]

For the practical implementation of the GA, several hyper-parameters have to be carefully adjusted to allow for efficient convergence [117]. Moreover, certain parameters feature trade-offs between each other. For example, small population sizes would lead to small population diversities, however, combined with high cross-over and low mutation rate (or vice versa) a good convergence can be achieved as premature stopping of the algorithm is prevented [117]. However, the mutation rate should not be too high, otherwise the algorithm performs more of a random guessing [118, 119] rather than a guided optimization [117].

2.1.2 Particle Swarm Optimization

The particle swarm optimization (PSO) algorithm was developed around 1995 and is inspired by the social behavior of animal swarms [109]. The algorithm works similarly to how schools of fish or swarms of bird scour for food. The PSO is of particular interest in optics, as it can optimize large parameter problems without knowledge of gradients at minimal computational cost compared to other algorithms, and is usually more intuitive to use compared to, for example, GAs. In contrast to the GA, the PSO relies on straight-forward vector multiplications, and not on elaborate operations as the GA (i.e., mutation or crossover).

Initially, a swarm $\vec{x}$ of size M is defined, which will be optimized in a N-dimensional search space. Here, N describes the numbers of optimization variables and the particles x_i ($i \in [1, M]$) the potential solutions that are evaluated for a given objective function at iteration step n. Thus, the swarm will be represented by a $M \times N$ matrix. During the algorithm initialization (i.e., at $n = 0$), the particles will be randomly distributed in the search space with random inertia weights ($\vec{w}^n$) for each particle, which define the particles initial position and movement. To control the convergence behavior of the algorithm, several hyper-parameters can be tuned. These parameters can control the trade-off between a fast exploration of the search space (i.e., effectively finding where the minimum is) or a more accurate targeting of a found minimum (exploitation) [120], and thus directly impact the convergence performance and runtime.

The update rule of the particles $\vec{x}_i$ per iteration step n can be mathematically described by equation 2.1:

$$\begin{aligned} \vec{x}_i^{n+1} &= \vec{x}_i^{n} + \Delta\vec{x}_i^{n+1} \quad \text{with} \\ \Delta\vec{x}_i^{n+1} &= \underbrace{w_i^n \Delta\vec{x}_i^n}_{inertia} + \underbrace{u_{self}\left(\vec{x}_{i,best} - \vec{x}_i^n\right)}_{self\ confidence} + \underbrace{u_{social}\left(X_{best} - \vec{x}_i^n\right)}_{swarm\ intelligence}. \end{aligned} \tag{2.1}$$

Each particle's position ($\vec{x}^{n+1}$) will be iteratively updated based on the own individual-best position ($\vec{x}_{best}$, 'self confidence'), and on the knowledge from the swarm collective ('swarm intelligence') through a parameter X_{best}, which describes the overall-best position of the swarm. The significance of the values for the individual- and overall-best parameters are additionally weighted by self- (u_{self}) and social (u_{social}) adjustment factors, respectively. The updating of an individual particle in a two-dimensional search space is illustrated in Figure 2.3 (Figure shows adaptive inertia (i.e., velocity vector can change from one iteration to next)).

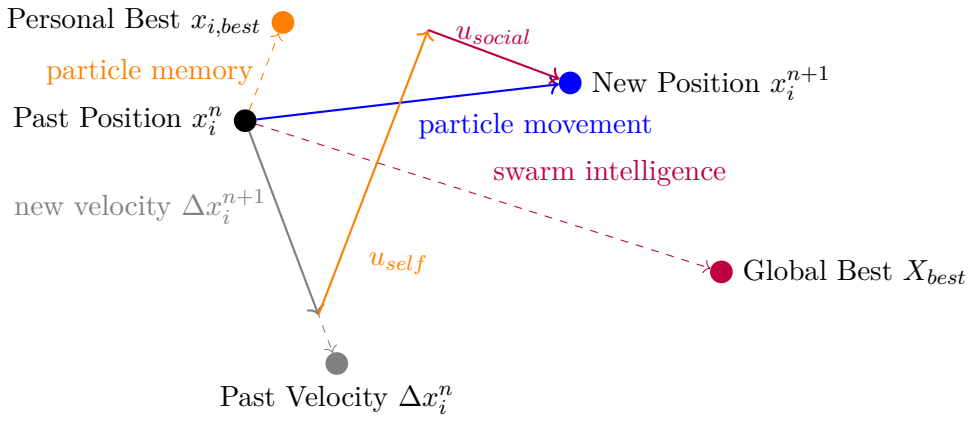


Figure 2.3: Illustration of the particle swarm optimization algorithm in a two-dimensional search space for a single particle. The calculation of the new velocity (and position) is mainly influenced by the current velocity, each particles own individual best solution (personal influence or cognitive acceleration), and the swarms intelligence (social influence or social acceleration).

One of the most effectual hyper-parameters are the self- and social adjustment factors, u_{self} and u_{social}, respectively. These factors weight the self confidence and swarm intelligence. In particular, it is useful to compare three cases: (i) a large self and low social adjustment factor, (ii) a small self and large social adjustment factor, and (iii) similar self and social adjustment factors.

In the case of (i) low convergence can usually be observed as the particles rely mostly on themselves and thus are barely impacted by the behavior of other particles in the search space. Thus, each particle only moves solely along its own trajectory preset mainly by the inertia term for

the direction, however, exploitation is predominant. The global best solution might not be found in this case within a reasonable iteration time. In the second case (ii), the opposite behavior can be observed. The particles will barely rely on their own best position and mostly on the global swarm leading to large 'swings' (or jumps) in the search space. However, the global best solution might be found here as a result of a higher exploration rate. Lastly (iii), a careful complementary adjustment of both factors will generally lead to the best convergence behavior as particles feature reasonable exploration (social adjustment) and accurate targeting of a minimum (self adjustment). Parameter values in the range $[1, 2]$ are often found as default values for both adjustment factors, and a value of $u_{self} \cdot u_{social} > 4$ is often considered a good starting point [111]. Further fine tuning into both directions of each factor might help improve convergence.

Notably, the swarmsize can affect the convergence behavior similarly to the social adjustment factors. Indeed, a smaller swarmsize usually features larger deviations to the global minimum, and thus will explore the solution space faster as the effective 'step-size' $\Delta\vec{x}_{n+1}$ increases (i.e., a faster scattering of particles). Contrary, a large swarmsize features smaller deviations towards the overall-best position, and as such the effective step-size decreases. Moreover, for only a few particles (i.e., a very small swarmsize) efficient convergence might not be possible as there are too little particles available in order to explore the search space sufficiently fast.

Lastly, convergence is impacted by the inertia factor w_n. Here, larger inertia can lead to divergence as particles move faster, smaller inertia to better convergence but if too small also increases time (comparable to a step-size in other algorithms). In general, the inertia can be understood as a way to further balance self confidence and swarm intelligence.

While the careful consideration of these parameters can already lead to good results, they often do not converge efficiently if the hyper-parameters are fixed throughout the optimization process. Indeed, the use of the PSO algorithm increases significantly with the introduction of adaptive parameters, which can change every iteration [112–114].

2.2 Picosecond Optical Pulse-Shaping: State-of-the-art

The control over the temporal properties of optical pulses is of great interest for many different applications as they can drastically impact the systems performance [121, 122]. In fact, the effi-

ciency of nonlinear interactions strongly depends on the pulse duration and shape (among other factors such as phase-matching or peak power [123]). This increase in efficiency is especially interesting for applications such as all-optical signal processing in the telecommunications domain [65], where the efficiency of e.g., nonlinear parametric amplification processes depend strongly on the signal waveform [124] or for the generation of parabolic pulses that experience less signal distortion during optical data transmission [125, 126]. Other examples include applications for improved spectroscopy [127], the optimized generation of Terahertz radiation [128, 129] or in quantum optics, were the tailoring of optical pulses allows to increase the coupling efficiency to optical cavities [130–132]. This can be of great interest for compact integrated platforms, which use for example microring-resonators as quantum state sources [133, 134].

Initially, the control over the temporal pulse envelope was accompanied by the concept of space-to-frequency mapping [135, 136]. These early works focused primarily on spatially dispersing the incoming pulse and then subsequent manipulation of individual spectral components (also known as Fourier synthesis as illustrated in Fig. 2.4) to achieve the desired pulse-shapes by using, for example, spatial light modulators [136, 137]. Although this concept found its use in commercial devices[6] [138], it is limited towards the femto and low picosecond regime due to the increasing demands on the dispersive elements, practicability (i.e., size), and cost. Therefore, a technological interest exists towards achieving picosecond waveform shaping in efficient and compact platforms, in order to improve the performance of existing and future applications.

To date, the majority of picosecond pulse-shaper implementations have focused primarily on fiber and electro-optic schemes. Existing approaches include for example, the use of birefringent crystals [139], frequency-to-time mapping techniques [126, 140], the use of large dispersion in combination with spectral shaping [126]or phase modulation [141]. Notably, contrary to the former approaches the concatenation of free-space interferometers has also been demonstrated as a viable option for pulse-shaping in a complete passive way, known as *temporal coherence synthesis* [142, 143]. While these approaches can partially solve the issue of picosecond pulse shaping, they are often bulky, use expensive large bandwidth equipment, and are difficult to scale towards more efficient and compact on-chip designs.

Therefore, recent progress targets arbitrary picosecond pulse shaping using various on-chip platforms and designs [144–149]. Implementations based on on-chip approaches are of particular interest,

[6]commercial devices are for example the Waveshapers: `https://ii-vi.com/product/waveshaper-b-series-cl/`

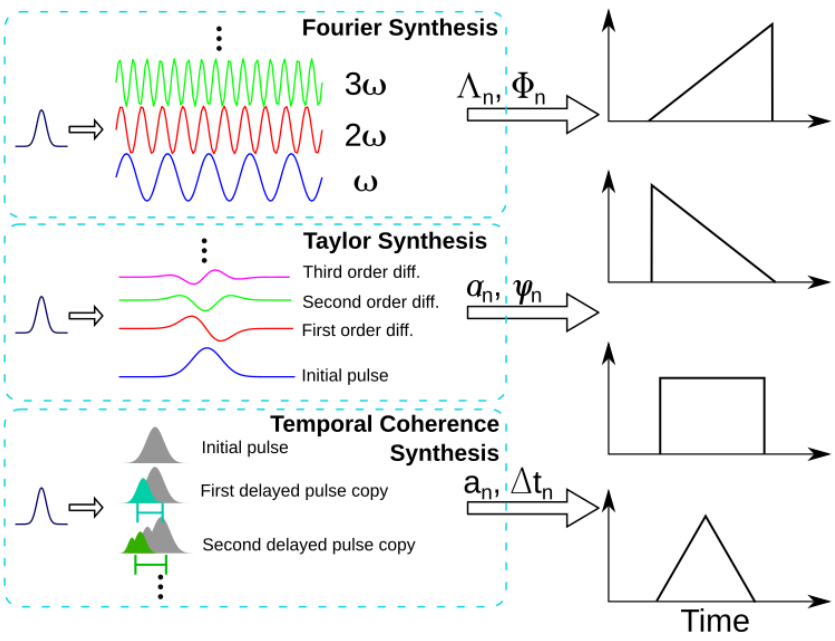


Figure 2.4: Illustration of different pulse-shaping concepts. Top: Fourier synthesis, where the waveform is constructed from different spectrally weighted components. Middle: Taylor synthesis, where the waveform is constructed using higher-order differentiation of the input pulse. Bottom: Temporal coherence synthesis, where the waveform is constructed via coherent overlapping of delayed pulse copies.

as they promise high performance, environmental stability, compact footprints, as well as potential mass-producability, hence ultimately reducing costs. Thus far, the majority of on-chip picosecond pulse-shaping implementations make use of Fourier [145] and Taylor synthesis methods [146] (see Fig. 2.4). The on-chip Fourier synthesis can be mathematically described [145] by the following transfer function

$$H_0(\omega) = \sum_{n=1}^{\infty} \Lambda_n \cdot e^{n\omega\tau + \Phi_n}. \tag{2.2}$$

Here, Λ_n and Φ_n are the amplitude and phase coefficients of the n^{th} angular frequency ω, respectively, and τ is the time-delay between two adjacent channels. Here, the required dispersion to 'access' individual frequency components can be achieved either by dispersive elements (such as arrayed waveguide gratings [150]) or multi-mode interference (MMI) couplers [145]. For the Taylor-synthesis, the output waveform $a(t)$ can be described in a more straight-forward manner as following:

$$a(t) = \sum_{n=0}^{\infty} \alpha_n g^{(n)}(t) e^{i\varphi_n}, \tag{2.3}$$

with α_n and φ_n as the amplitude and phase coefficients, $g^{(n)}(t)$ as the n^{th} differentiation of the input pulse and $\sqrt{-1} = i$ as the imaginary unit. Here, the Taylor-synthesis approach offers advantages over the Fourier-synthesis (or frequency-to-time mapping) concept, as no dispersive components or MMI couplers are required, and the implementation of optical differentiators can be straight-forwardly realized by microring resonators [125, 146, 151, 152]. However, for multiple differentiators the chip design becomes increasingly complex, placing a high demand on fabrication tolerances for optimum performance.

Although, these approaches are very promising, the chosen platforms often suffer from high optical losses, still require active modulation (resulting in a reduced energy efficiency) or have only low flexibility in terms of input and output pulse durations due to a limited number of waveguides or on-chip components (e.g.,Refs. [145] and [146] use four and three channels, respectively). Moreover, all current implementations use pre-defined weights for amplitude and phase from simulations to be transferred to the physical system. This approach often comes with limitations in achievable accuracy as not all experimental fluctuations can be considered in the simulations, such as cross-talk, fabrication tolerances [153, 154] or different environmental conditions can have a drastic impact on the performance of the designed circuits [155]. Therefore, methods to overcome these restrictions are of high interest for practical applications. Contrary to these *in silico* designs and methods, the use of machine-learning and optimization algorithms can potentially allow for flexibility and reconfigurability in optical pulse shaping as demonstrated in earlier concepts [36, 38, 156–158].

An alternative approach for the Fourier- and Taylor-synthesis methods is pulse-shaping based on temporal coherence synthesis [142, 143] (TCS). Here, the target waveform is generated by coherent overlapping of delayed and amplitude weighted (Δt_n and a_n in Fig. 2.4) pulse copies. An advantage of TCS is, that it can work without specifically designing a system as opposed to the aforementioned methods, given that TCS utilizes interferometers. Only access to relative time-delays and amplitude control are necessary (in free-space). This was first realized in 2007 [142] using a cascade of free-space Michelson interferometers [159]. TCS can be described in the frequency domain for M number of concatenated two-arm interferometers as [142]

$$A(\omega) \approx W(\omega)E(\omega) \prod_{n=1}^{M} (1 + \exp(-i(\omega' + \omega_c)\Delta t_n)), \tag{2.4}$$

with ω_c and ω' being the central carrier and base-band frequency, respectively ($\omega = \omega_c + \omega'$) and Δt_n being the relative time-delay between pulse copies. $W(\omega)$ is an amplitude weight matrix and $E(\omega)$ is the electrical field of the input pulse. The time-domain TCS can equivalently be written as

$$a(t) \approx w(t)e(t) \otimes (\delta(t) + \delta(t - \Delta t_1)) \otimes (\delta(t) + \delta(t - \Delta t_2)) \otimes ... \otimes (\delta(t) + \delta(t - \Delta t_m)). \tag{2.5}$$

Here, $a(t)$ and $e(t)$ are the output and input fields, respectively, $\delta(t)$ is the dirac delta function, and $\otimes$ stands for the convolution operation.

However, moving this concept to on-chip approaches is challenging given imperfect designs/fabrication, as well as additional impact by the phase (e.g., through dispersion). Thus, in the following the application of TCS in an integrated platform will be studied.

2.3 Autonomous Picosecond Optical Pulse-Shaper

In the following, a novel approach for the implementation of an integrated picosecond pulse-shaper is presented. The proposed device is based on a temporal coherence synthesis scheme in combination with an optimization algorithm to generate the desired waveforms. Contrary to previous implementations, this approach utilizes on-chip concatenated Mach-Zehnder instead of Michelson interferometers. Fast and efficient diagnostics is crucial for the implementation of such *smart*-photonic processors. In order to avoid the use of slow and expensive measurement techniques in the picosecond domain, an all-optical sampling scheme was implemented. This monitoring solution was realized based on degenerate four-wave mixing in a highly non-linear fiber and allows for straight-forward fast and ambiguity-free measurements of the picosecond waveforms.

2.3.1 Programmable On-Chip Split-And-Delay Line

Fast switching in light-confining waveguides, promises efficient and novel applications in terms of smart photonics with a very compact footprint through the use of electronically interconnected, integrated photonics [160]. The photonic chip used in this work was initially designed as an optical switch [161], and was subsequently re-purposed, in combinations with optimization techniques, to allow for customized nonlinear light generation [42].
The optical chip is based on Hydex, a proprietary material with comparable properties to silicon-oxy-nitride (SiON) [162, 163], and also a complementary-metal-oxide-semiconductor (CMOS) compatible material platform that permits mass-producible devices with vast functionalities (such as thermo-optical switches [161], microring resonators [162], or long spiral (>50 cm) waveguides [163]) on a small footprint. Remarkably, waveguides based on SiON provide exceptional low linear (<0.06 dB/cm) and nonlinear losses[7], and sufficiently high refractive index contrast with $n_{Hydex} = 1.7$, in comparison to fused silica ($n_{SiO_2} = 1.45$), SiN ($n_{SiN} = 2.0$), and Si ($n_{Si} = 3.0$) crucial for robust light guidance. Both are essential to achieve multiple, picosecond-scale delays on a single optical chip. Compared to waveguide platforms based on materials such as silicon-on-insulator, SiON offers the additional benefits of (i) smoother waveguide roughness, which reduces propagation losses [164], and (ii) the absence of nonlinear losses, such as two-photon absorption, which can lead to significant losses in silicon around 1550 nm and hence usually requires elaborate counter mechanisms such as p-i-n structures [165, 166]. While the former is important for the implementation of longer delays, the latter is of importance if one considers the shaping of high-power or femtosecond pulses.
The most relevant chip parameters are summarized in Table 2.1.

The specific chip used for the optical pulse-shaping based on temporal coherence synthesis features a cascade of on-chip balanced and unbalanced Mach-Zehnder interferometers (MZIs) as illustrated in Fig. 2.5.

Mach-Zehnder interferometers are a type of interferometer consisting of two couplers/splitters, one for splitting and the other for recombination – instead of a single splitter as for example used in Michelson interferometers [159]. One of the two arms of the interferometer usually features either a fixed or variable a delay (also referred to as unbalanced interferometer, conversely no path difference between both arms is known as a balanced MZI), which allows to produce either a specific pulse

[7] negligible nonlinear losses up to peak intensities of 25GW·cm^{-2}[162]

Parameter	**Value**	**Unit**
Waveguide refractive index	1.70	[1]
Cladding refractive index	1.45	[1]
Waveguide dimensions	1.5×1.5	[μm^2]
Group velocity	5.87	[ns/m]
Linear loss	0.10	[dB/cm]
Nonlinear loss (@1550nm)	-	[dB/cm]
Insertion loss per facet	1.40	[dB]
Nonlinear Coefficient γ	233	[$W^{-1} \cdot km^{-1}$]
Dispersion Coefficient β_2	-2.87	[ps^2/km]
Dispersion Coefficient β_3	-0.0224	[ps^3/km]
Min./Max. path lengths	5.0/9.5	[cm]
Minimum waveguide bend radius	35.0	[μm]

Table 2.1: Material parameters for the split-and-delay line chip used for the pulse-shaping [42, 161, 163].

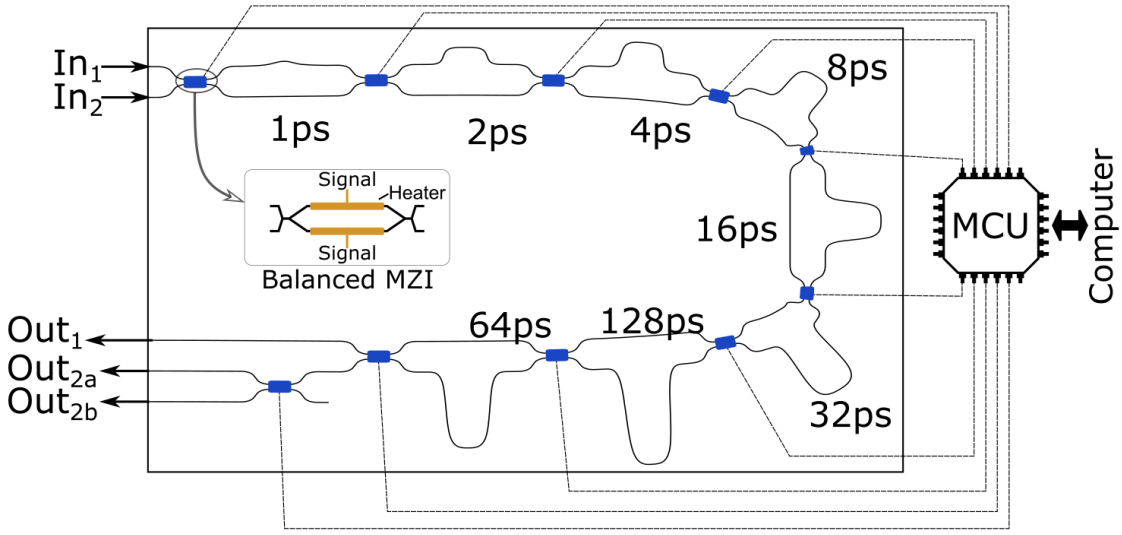


Figure 2.5: Illustration of the split-and-delay line chip layout. The sample features ten tunable balanced MZIs (blue boxes), eight unbalanced MZI with binary increasing delays, two input and three output ports which are all fiber-coupled. The balanced MZIs are controlled through a computer via a microcontroller unit (MCU). Each of the short paths features an additional tap port (≈ 5%) but is omitted in the sketch for clarity (see Appendix A.2).

splitting (e.g. to achieve coherent generation of double pulses if the input pulse duration is less than the MZI delay) or interference effects (for pulse durations longer than the MZI delay). Optical switching is enabled by controlled, voltage-dependent coupling between two neighboring waveguides based on the relative phase difference between them, which can be altered by applying a voltage on the gold electrodes [159]. In such a configuration, the controlled current heats the electrodes which in turn alter the samples refractive index via the thermo-optic effect. Optical propagation can

be described by coupled-mode theory applied to two parallel waveguides that constantly exchange power during propagation. For a given length, the input phase determines the power exchange until it reaches a coupling minimum or maximum. The specific chip used for the demonstration has a bit-wise increasing delay architecture

$$\Delta t_n = T \cdot 2^{n-1}, \tag{2.6}$$

with T=1 ps and integer $n \in [1, 8]$. Thus, the design allows for up to 255 ps in delays with a resolution of 1 ps (see Appendix A.2). The sample features a total of ten tunable balanced MZIs in a push-pull configuration ($\approx$ 100 ms switching time [161]) forming eight unbalanced MZIs (i.e., the individual delays) and one variable optical attenuator (VOA) for ports $\mathrm{Out}_{2a,b}$. The thermal heaters on the balanced MZIs are wire-bonded with gold wires and electrical addressable through a MCU (Atmel Atmega 328P), which can be additionally interfaced to a standard computer, see Fig. 2.6. By applying a voltage to the heaters of the balanced MZIs, the transmission can be tuned from long to short path of the unbalanced MZI. A characteristic curve for this switching behavior is shown in appendix A.2. At each balanced interferometer, the incoming pulse is split with a particular amplitude ratio (mediated by a microcontroller unit, MCU). Subsequently, one copy of the pulse travels along the long (i.e., delay) path, while the other copy travels a fixed short path. After propagation at the next MZI, both pulse copies interfere with each other generating a new waveform.

Indeed, cascading MZIs is advantageous to reduce optical losses. In such configuration, the optical pulse energy does not leave the coupled waveguide system until the optical pulse reaches the last coupler. Hence, the device loss, and thus the system's energy efficiency, is determined by the coupling coefficients for the last coupler of the setup only (besides low propagation and fiber-to-chip coupling losses). Lastly, for increased environmental stability, the chip was glued inside an aluminum holder using a thermally conductive adhesive.

2.3.2 Integrated Split-And-Delay Line for On-Chip TCS

The actual design of the chip consists of balanced and unbalanced interferometers as depicted in Fig. 2.5 and Fig. 2.7a. Each delays coupling ratio can be tuned by the balanced interferometer

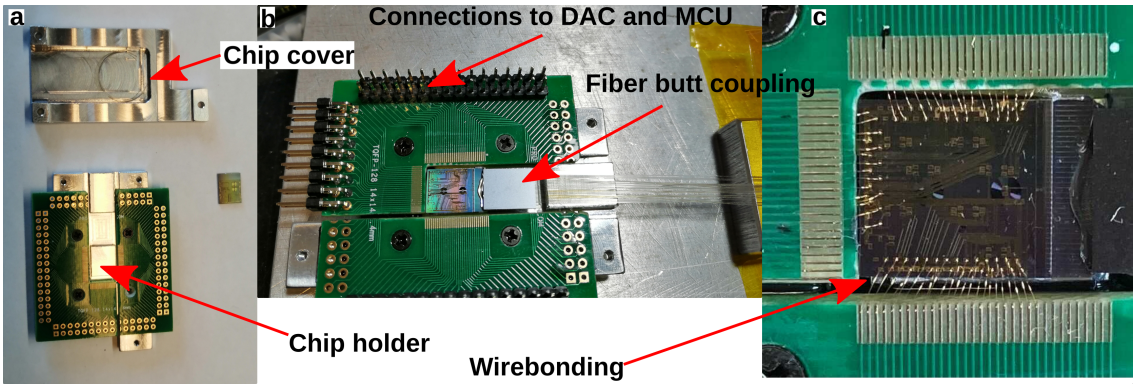


Figure 2.6: Picture of the split-and-delay line chip. a) Overview of different components. From top to bottom: Chip cover, photonic chip sample (11.75×9.00 mm^2), base-plate with mounted PCB (printed circuit board). b) Assembled and fiber-coupled chip incl. pin connections to a 16-bit digital-to-analog converters (DACs, Analog Devices AD5370) and MCU. c) Chip sample after wirebonding.

section. As the sample was initially designed for a different application [161], the exact lengths (and thus phase) of the balanced interferometer section is unknown but constant throughout the experiment. As shown in previous work [167], such a configuration can be simplified to a tunable coupler (see Fig. 2.7b).

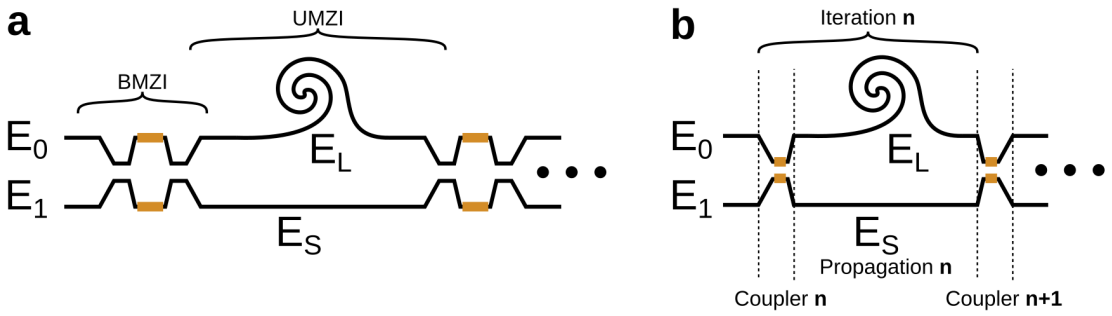


Figure 2.7: Illustration of the integrated spit-and-delay line. a) Accurate description of a single delay (unbalanced Mach-Zehnder interferometer, UMZI), which can be controlled by means of two balanced Mach-Zehnder interferometers (BMZIs). b) Simplification for simulation purposes. The functionality of the individual delays can be likewise described by means of tunable coupler and delay matrices [167]. This simplifies the simulation from multiple elements per delay section n to only two per delay section. E_0 and E_1 are the input fields to each coupler, while E_L and E_S describe the propagation through the long and short delay, respectively.

With this simplification in mind, the behavior of tunable couplers as in Fig. 2.7b can be straightforwardly described using a matrix approach [159, 168]. Modeling the chip using a matrix approach significantly simplifies simulations opposed to previous implementations of TCS (see equation 2.4 and 2.5). For MZIs, the output fields E_L and E_S for the long and short path can be described as a matrix product of a tunable coupler and propagation for given input fields E_0 and E_1:

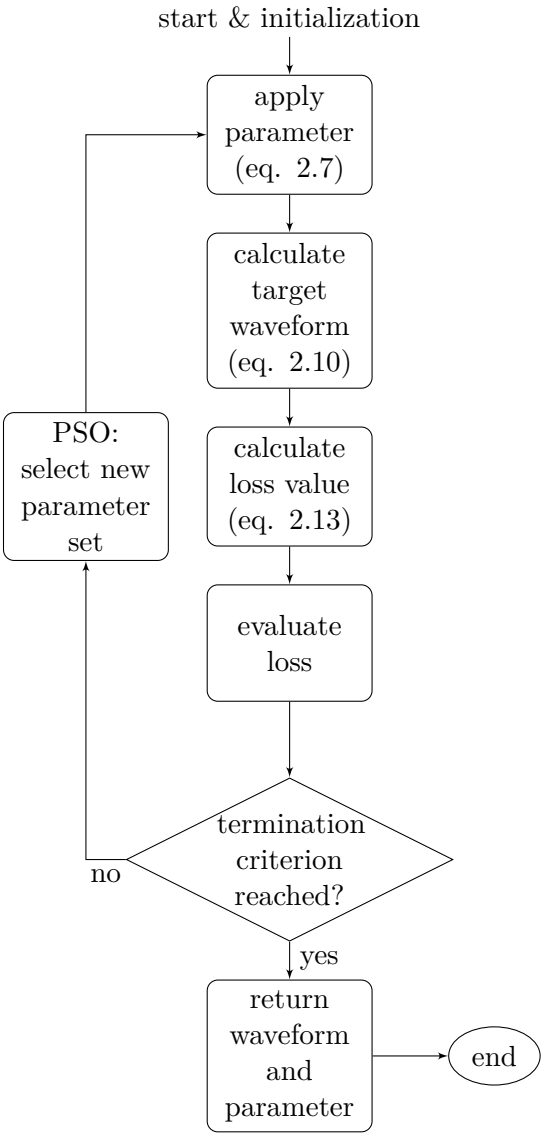
$$\begin{bmatrix} E_L \\ E_S \end{bmatrix} = \prod_{n=1}^{N} \left\{ \underbrace{\begin{bmatrix} exp(i\varphi_n) & 0 \\ 0 & 1 \end{bmatrix}}_{Propagation} \underbrace{\begin{bmatrix} \sqrt{1-\varepsilon_n} & i\sqrt{\varepsilon_n} \\ i\sqrt{\varepsilon_n} & \sqrt{1-\varepsilon_n} \end{bmatrix}}_{Coupler} \right\} \begin{bmatrix} E_0 \\ E_1 \end{bmatrix}, \tag{2.7}$$

with $\varphi_n = \frac{2\pi}{\lambda} n_{eff}(\lambda) \Delta L_n + \Delta\Phi_n$, where λ is the central wavelength, $n_{eff}(\lambda)$ is the effective refractive index of the fundamental mode of the waveguide, and $\Delta L_n = \frac{c_0}{n_{eff}} \cdot 1\text{ps} \cdot 2^{n-1}$ is the length of the n^{th} long path. Since the calculation of the input and output field is a straight-forward multiplication in the Fourier domain (see Fig. 2.8b), φ_n can be written as $\varphi_n = \omega \cdot 2^{n-1} \cdot 1\,\text{ps} + \Delta\Phi_n$ with ω as the angular frequency of the light. ε_n denotes to the amplitude ratios which can individually controlled by the heaters, and term $\Delta\Phi_m$ is an additional phase offset from the delay and UMZI. In an optimal design (see section 2.3.5), each delay would feature a phase-shifter that allows additional control over this offset phase. The flowchart for the simulation algorithm and a pseudo code for the pulse-splitter are illustrated in Fig. 2.8.

It is important to note, that the matrix approach is only valid for optical pulses featuring low peak powers (i.e., low optical power and long optical pulses). In such case, nonlinear and dispersion effects during propagation can be neglected. This was the case for the experimental results in this work (pulse durations >10 ps and average powers of around 2-3 mW). In case of higher peak powers (or platforms with higher nonlinearity such as silicon), the propagation requires modeling using an approach based on coupled mode theory and the generalized nonlinear Schrödinger equation [42].

2.3.3 All-Optical Sampling Feedback System

Crucial for the implementation of smart photonic applications is the efficient and ultra-fast readout of the system output that can be used towards the optimization of a target output. For (arbitrary) pulse-shaping applications the unambiguous readout of the temporal information with sufficiently high accuracy is required. For this, only a limited number of techniques exist.
Several techniques have been developed in order to measure optical pulses, ranging from straight-forward detection using photodiodes and analog-to-digital converters, to complex full-field reconstruction of optical pulses [123]. While the former is limited to costly state-of-the-art radio-frequency



(a) Flowchart for the pulse-shaping simulation. Inside the optimization algorithm, the output for the current parameter set (= split ratios) is calculated (see Figure 2.8b) and subsequently evaluated.

Input: Chip parameters, Input pulse parameters, Target waveform, Algorithm hyper-parameter
Output: Chip settings, Output waveform, Figures
initialization;
while *Pulse-shaping simulation runs* **do**
  select parameter set (chip setting & pulse);
  perform FFT of time-domain pulse;
  for *1:number of switches* **do**
    `/* forward propagation */`
    calculate current phase φ_n for delay n;
    evaluate field for long and short path ; `/* eq. 2.7 */`
  end
  save short path in temporary variable;
  for *number of switches:1* **do**
    `/* backward propagation */`
    calculate current phase φ_n for delay n;
    evaluate field for long and short path ; `/* eq. 2.7 */`
  end
  perform IFFT of Fourier domain signal;
  return *Chip Settings, Output Waveform*;
end

(b) Pseudo code for the pulse-shaping simulation, step 'apply parameter' which emulates the pulse-splitter.

Figure 2.8: Flowchart of the pulse-shaping optimization algorithm. a) Flowchart, b) Pseudo-code for the 'apply parameter' step.

equipment[8] (≈110 GHz analog bandwidths), the latter is a rather complex technique which requires additional retrieval algorithms [169], which can result in slow measurements. Moreover, full-field reconstruction methods work often only in limited time-spans, generally in the femtosecond domain up to a few picoseconds (< 10 ps). Nevertheless, earlier pulse-shaping implementations often

[8] to capture the true pulse shape, detectors with response times significantly shorter than the measured pulse width are required [159]

utilized such techniques, primarily towards the manipulation of femtosecond pulses [39, 71]. Alternatively, other methods such as autocorrelation [159] are inexpensive and straight-forward to implement and have thus been used for machine-learning applications especially in the context of *smart* lasers [32, 72]. However, an autocorrelation always returns a symmetric signal and hence features waveform ambiguity, i.e., different output waveforms may result in the same or very similar autocorrelation traces.

Fig. 2.9 illustrates this issue. While the autocorrelation is suitable for symmetric pulses (the black Gaussian pulse), for example in the context of laser optimization [32, 92] where one aims to generate Gaussian or Lorentzian pulses, it leads to identical autocorrelation signals in the case of asymmetric input pulses. For example, the autocorrelation of positive and negative sawtooth pulses (Fig. 2.9, blue and magenta) are indifferent and provide an inadequate solution for pulse-shaping applications. On the other hand, the cross-correlation with a sampling signal (the red Gaussian pulse) returns a more accurate representation of the true input waveform *if* the sampling signal is temporally short enough. The non-zero angle of the vertical side of the cross-correlation plots (blue and magenta) is caused by the sampling pulse, and a longer pulse-duration will cause a more blurred signal (i.e., much lower angle), whereas a shorter pulse-duration will result in a 'sharper' edge (i.e., more vertical).

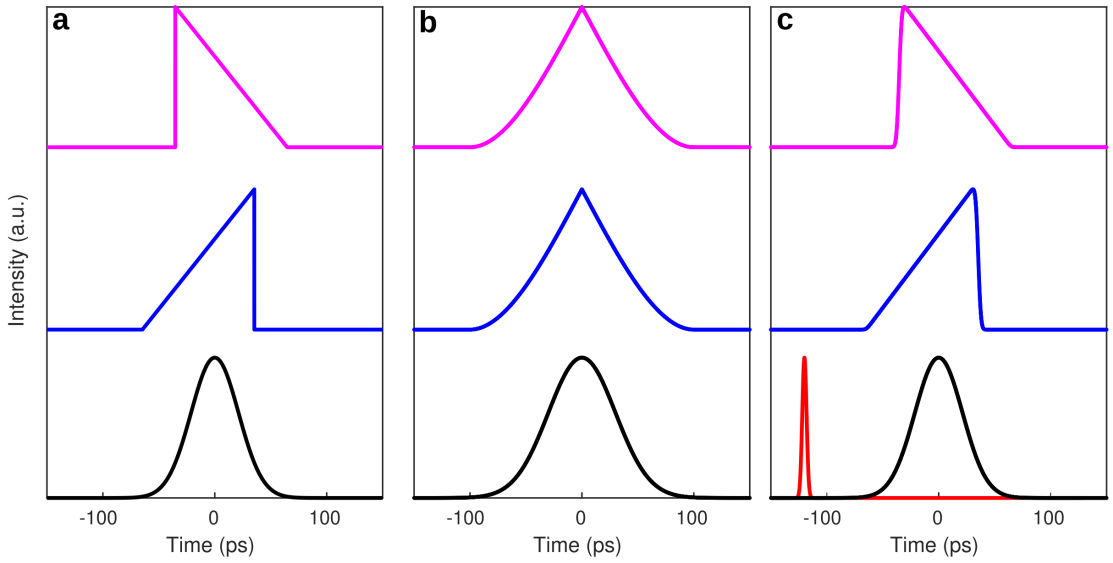


Figure 2.9: Comparison between auto- and cross-correlation between different input waveforms. a) Input waveforms (all 50 ps full width at half maximum, FWHM): Negative sawtooth (magenta), positive sawtooth (blue), Gaussian (black). b) Autocorrelation. c) Cross-correlation between input (a) and the red sampling signal (4.5 ps FWHM).

Therefore, methods such as autocorrelation are unsuited towards monitoring arbitrary waveforms. Conclusively, in order to efficiently utilize machine-learning techniques in experimental platforms, fast yet simple, unambiguous and accurate detection schemes are required.
Given these constraints, few techniques remain that are suitable for efficient on-the-fly machine-learning implementations:

- Indirect methods recording the final output of, e.g., a nonlinear process, and maximizing the resulting signal (i.e., conversion efficiency) [36, 97]
- Direct temporal measurements for longer pulses ($> 50\,\mathrm{ps}$) with state-of-the-art ultra-fast oscilloscopes ($\approx$110 GHz analog bandwidth)
- Temporal measurement using time-lens techniques [18, 170] for shorter pulses $< 20\,\mathrm{ps}$ with low phase distortions
- Temporal measurement using sampling techniques such as electronic sampling oscilloscope [145, 146] or all-optical sampling techniques[171–174]

Here, each of these methods have their own advantages and disadvantages in terms of experimental complexity and accessibility, measurement accuracy, achievable span/bandwidth, chip integratability, as well as costs. Implementations that fulfill most of these conditions are time-lens and optical sampling techniques.
The former, which was developed from space-time duality [175], has shown remarkable results in the past towards pulse characterization [176–178] and ultra-fast photonic oscilloscopes [179], but it comes with certain drawbacks. In fact, the implementation of a time-lens requires large dispersion in order to stretch the pulse wide enough to be measurable with low and more accessible bandwidth ($\ll 50\,\mathrm{GHz}$) detection equipment. Moreover, the waveform under investigation should not feature too large phase irregularities, as non-uniform phase can show strong aberrations arising from self-interference between different spectral components [180].
As outlined earlier, (all-) optical sampling is a viable alternative to measure picosecond pulses independent of the pulse phase. Due to the simplicity of these setups and the large achievable detection bandwidth ($\gg 100\,\mathrm{GHz}$), they have been heavily investigated as receivers in telecom applications [171, 181]. However, they have not been investigated yet for use in *smart*-photonic implementations.

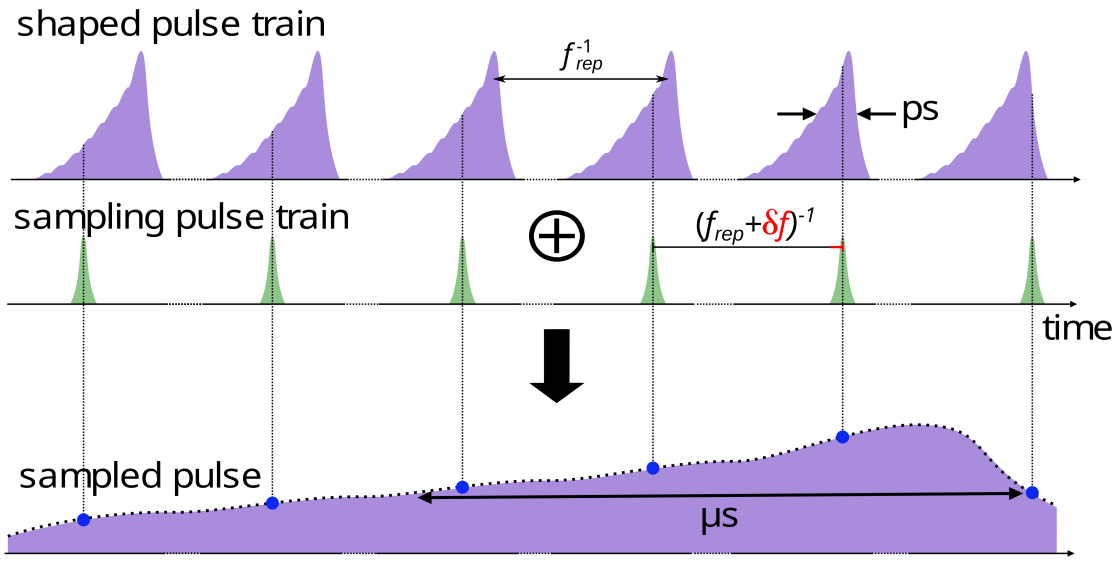


Figure 2.10: Illustration of the all-optical sampling principle. A probing pulse with a duration of $\Delta\tau_{Sampling} \ll \Delta\tau_{Target}$ is used to sample the target waveform. It is important, that both the sampling source and the target source feature a repetition rate offset δf.

The optical sampling follows a simple, yet effective principle as illustrated in Fig. 2.10. The core element for the sampling scheme is an (optical) AND-gate that combines a sampling signal with the signal under test. Commonly in optics, a sampling pulse and the target pulse are overlapped inside a nonlinear medium and parametric processes generate a new signal which is proportional to the target signal. In order to reconstruct the original pulse, the sampling pulse has to be scanned over the target pulse. This can be either achieved by tuning a delay or more straight-forwardly achieved by detuning the repetition rates of the target signal compared to the sampling pulse, so-called asynchronous sampling [171, 173, 174]. The latter was used in this work, where the sampling source and the shaping source were detuned by a multiple integer plus a small difference. The newly generated signal is generated at a much lower rate, which corresponds to the detuning between the target and sampling pulse (or multiples plus a small difference [174]). Thus, the target pulse can be measured with inexpensive and low bandwidth electronics ($\ll$GHz).

An optical AND-gate can be realized in various ways, such as highly nonlinear fiber using parametric effects, e.g., four-wave mixing [171, 172], cross-phase modulation [182], or various effects in semiconductor optical amplifiers [183–187] (e.g., cross-polarization and -gain modulation, or two-photon-absorption).

For the implementation in this work [40], four-wave mixing in a highly nonlinear fiber was utilized. The fiber parameters are summarized in Table 2.2.

Parameter	**Value**	**Unit**
Fiber length	1015	[m]
Effective area	11.6	[μm^2]
Dispersion D	0.08	[ps/nm/km]
Dispersion Slope S	0.017	[ps/nm^2/km]
Zero Dispersion Wavelength	1545	[nm]
Total Loss (@1550nm)	0.99	[dB]
Polarization Mode Dispersion	0.06	[ps/$\sqrt{km}$]
Nonlinear Coefficient γ	11.3	[$W^{-1}{\cdot}km^{-1}$]
Dispersion Coefficient β_2	-0.102	[ps^2/km]
Dispersion Coefficient β_3	0.0278	[ps^3/km]
Dispersion Coefficient β_4	$4.0{\times}10^{-5}$	[ps^4/km]

Table 2.2: Fiber parameters of the highly nonlinear fiber used for the all-optical gate [42, 188].

Exploiting the parametric nonlinearity of the highly nonlinear fiber, two pump photons and a signal photon can be converted into an idler photon symmetric to the pump as illustrated in Fig. 2.11. The generation of the idler at a given sampling point N is proportional to the square of the pump power and can be mathematically described as follows [189–191]:

$$P_{i,N} \propto \gamma^2 P_s(t) P_p^2(t - N\tau), \tag{2.8}$$

with $P_{i,N}$, $P_s(t)$ and $P_p(t - N\tau)$ being the optical power of the generated idler, the seed signal and the temporally shifted pump, respectively. γ is the nonlinear coefficient of the chosen nonlinear medium (see Table 2.2) and τ the temporal resolution of the sampling scheme.
The sampling resolution depends on both, the full-width at half maximum (FWHM) of the seed

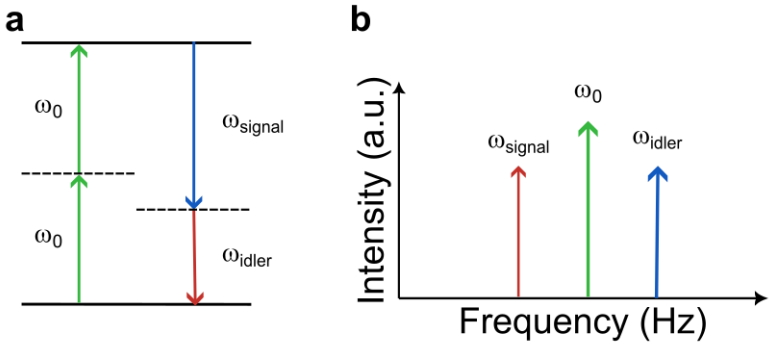


Figure 2.11: Illustration of the degenerate four-wave mixing process. a) Energy level diagram. Two pump photons ω_0 are converted to an idler photon ω_{idler} with the aid of a signal photon ω_{signal}. It holds $\omega_{signal} < \omega_0 < \omega_{idler}$. b) Frequency-domain representation. The pump and signal photons create an idler photon symmetric to the pump field.

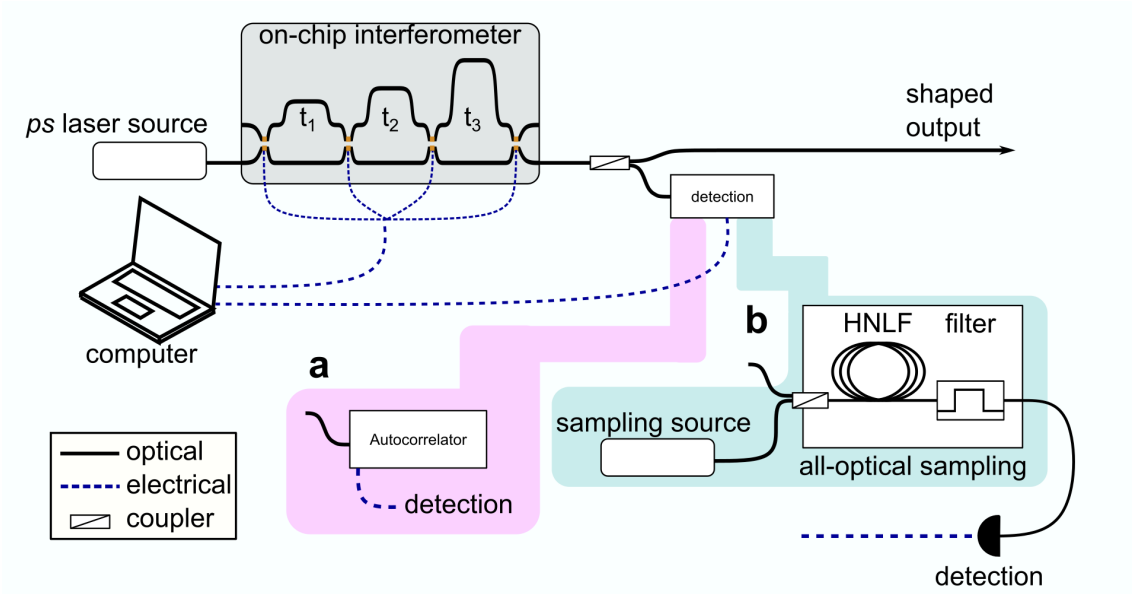


Figure 2.12: Characterization setup for the all-optical sampling scheme. Top: Simplified setup consisting of the shaping source and the integrated pulse-splitter. A tap port is implemented after the pulse-splitter for the detection scheme. a) Autocorrelation measurement. b) All-optical sampling scheme. The sampling source emits a pulse train which is a multiple (n=25) plus a small difference (<MHz) of the shaping source. The highly nonlinear fiber (HNLF) is used as the optical AND-gate and a subsequent filter is used to separate the generated idler.

signal (requiring $\Delta\tau_{Sampling} \ll \Delta\tau_{Target}$), see appendix A.1, as well as the difference in repetition rate δf between the seed signal and the pump signal. The temporal resolution τ can be calculated as follows given the repetition rate difference δf:

$$\tau = (n f_{rep,1})^{-1} \cdot V^{-1} \quad \text{with} \quad V = \frac{n f_{rep,1}}{\delta f}. \tag{2.9}$$

Here, $f_{rep,1}$ is the repetition rate of the shaped pump, n is an integer, V is a stretching factor to correct the retrieved time-scales, and $\delta f = |n f_{rep,1} - f_{rep,2}|$ is the difference between the shaped pump and the sampling source. In the chosen configuration, the temporal resolution is $\tau \approx 4.3\,\text{ps}$.

Before the implementation of the smart pulse-shaper system, the all-optical sampling scheme was characterized, using the setup shown in Fig. 2.12. The specifically implemented all-optical sampling scheme (Fig. 2.12b) consists of the sampling source (Menlo FC1500-250-WG), a highly nonlinear fiber (OFS-Fitel) as the optical AND gate, a tunable optical filter (Finisar Waveshaper 4000X) to separate the idler, and the detection electronics (Photodiode: Finisar XPDV2120R, Oscilloscope: Agilent DSO-X 92804A). The parameters for the used laser sources are summarized in Table 2.3.

For the characterization, a picosecond pump source (Pritel FFL) with a pulse duration of ≈20 ps was used. The obtained signal from the realized optical sampling scheme was subsequently compared to a commercial autocorrelator (Femtochrome FR-103XL) to evaluate its performance.

Parameter		**Pump source** Pritel FFL	**Sampling source** Menlo FC1500-250-WG
Central wavelength	[nm]	1544.6	1560
Pulse duration	[ps]	21.85	4.4
Repetition rate	[MHz]	10	250.27

Table 2.3: Overview over the two laser sources used for the pulse-shaping experiment.

Prior to the temporal pulse detection, the AND-gate operation was first optimized for maximum efficiency of the FWM process. For this, the optical spectrum was measured between the HNLF and the optical filter, and the four-wave mixing gain (i.e., the signal-to-noise ratio of the generated idler) was adjusted by carefully tuning the pump wavelength (Pritel FFL) in respect to the zero dispersion wavelength (≈1545 nm) and the wavelength and bandwidth of the signal[9] source (Menlo FC1500-250-WG). Fig. 2.13 shows the characterization results of the implemented all-optical sampling technique. Fig. 2.13a shows the optical spectrum before filtering. The signal from the sampling source was set to a central wavelength of 1557.9 nm with a 4.8 nm bandwidth using a tunable optical filter (Santec OTF350, not shown in Fig. 2.12). Moreover, to allow the continuous scanning over the target pulse, the repetition rate of the seed/sampling laser was set to 250.27 MHz, which is a factor 25 above the pump pulse (10 MHz) plus the small difference of Δf =0.27 MHz (see Appendix A.1 for the source characterization). After the HNLF, the generated idler signal was filtered using second tunable filter (Finisar Waveshaper, 1532.4 ± 3.8 nm) and detected.

As it can be seen from the illustration and results in Fig. 2.10 and Fig. 2.13b, respectively, the returned sampling waveform is discrete. As a monitor for pulse-shaping applications, a continuous target waveform will be required, so that the algorithm can continuously compare measured data points and the targeted points. Thus, an envelope retrieval is required to convert the discrete signal into a continuous. In this work, a software-based Hilbert transform [192] is used, but also direct implementation with electrical circuits is a low-cost possibility, avoiding computational expensive overhead, with minimal electronic bandwidth constraints [193].
From the measured sampling signal, the Hilbert transform was applied to reconstruct the pulse envelope, and the square root of the transformed signal was taken to obtain P_p. This is necessary

[9] pump and signal source refer to the FWM process and correspond to the shaped and sampling source, respectively

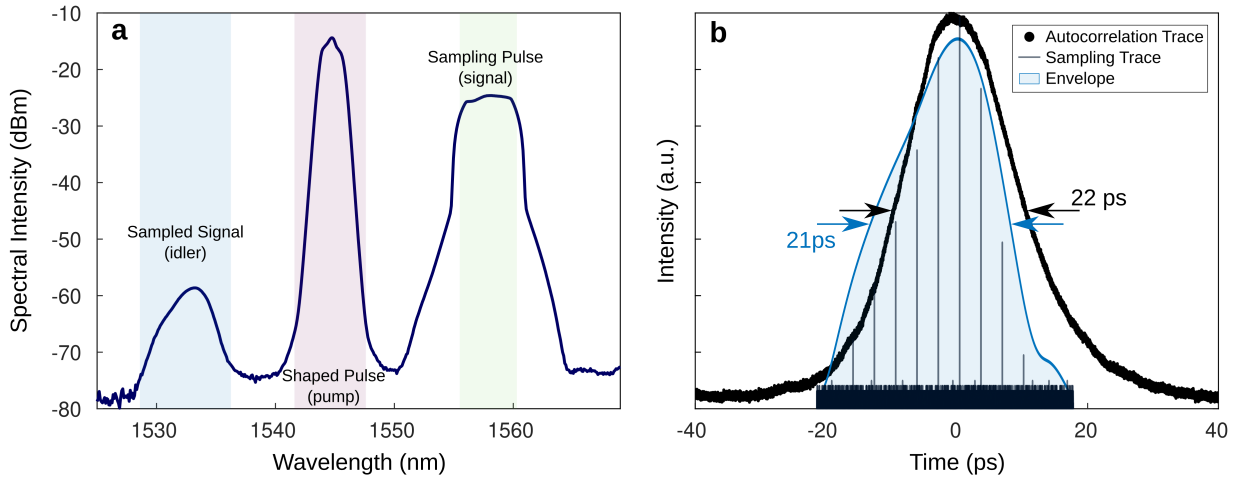


Figure 2.13: Results of the all-optical sampling characterization. a) Optical spectrum measured after the highly-nonlinear fiber. b) Autocorrelation measurement (black), retrieved sampling signal (grey) and the square-root of the sampling signal envelope (blue). For both, autocorrelation and optical sampling the correction factor was applied to retrieve the true pulse-width.

as the sampling signal returns the squared pump signal P_p^2 (see eq. 2.8). Since the down-sampling results in a different time-scale (ms instead of ps), the measured signal requires recalibration (similarly to an autocorrelator correction factor). This recalibration was performed by applying the stretching factor V (see equation 2.9). After applying the correction factor for both, the autocorrelation and the optical sampling, the retrieved pulse widths at full width half maximum match very well ($(\Delta\tau_{FWHM,1,AC} = 21.85\,\text{ps}$ and $\Delta\tau_{FWHM,1,Sampling} = 21.26\,\text{ps})$).

For an increased stability against environmental changes, the two laser sources (pump and seed) were locked to one another. In detail, the repetition rate of the pump source (10 MHz) was used as an external radio-frequency reference for the seed laser to lock the repetition rate. Thus, in case of any drifts in the repetition rate of the pump laser, the repetition rate of the seed laser changes accordingly, and the difference δf in repetition rate is at all times constant. In this configuration, the experiment was running stable for several weeks.

2.3.4 Particle Swarm Optimization-based Smart Interfacing

Equation 2.10 shows the specific targets for each of the implemented waveforms, where $H(t)$ refers to the heaviside step function[10], t_c is the temporal offset from zero (determined as the center of the returned waveform), I is an amplitude factor, and $mod(a, b)$ is the modulus operation.

[10]here, the half-maximum convention H(0)=0.5 was used

$$a_{target}(t) = \begin{cases} I \cdot \mod(t - t_c, \Delta\tau) H(t - t_c + \Delta\tau) \cdot (1 - H(t - t_c)) & \text{Positive Sawtooth} \\ I \cdot \mod(-t - t_c, \Delta\tau) H(t - t_c) \cdot (1 - H(t - t_c - \Delta\tau)) & \text{Negative Sawtooth} \\ I \cdot H(t - t_c + \frac{\Delta\tau}{2}) \cdot (1 - H(t - t_c - \frac{\Delta\tau}{2})) & \text{Flattop} \\ I \cdot \left(1 - \frac{|t - t_c|}{\frac{\Delta\tau}{2}}\right) H(t - t_c + \frac{\Delta\tau}{2}) \cdot (1 - H(t - t_c - \frac{\Delta\tau}{2})) & \text{Triangle} \end{cases} \tag{2.10}$$

For the experiment and simulation, two different use cases were implemented: Firstly, for the experiment, the FWHM ($\Delta\tau$ in equation 2.10) for all target waveforms were unconstrained, thus allowing to find the best performing waveform (i.e., the highest waveform fidelity) of the system independently of the waveform FWHM. Secondly, for the simulations in Section 2.3.5, the FWHM was fixed to arbitrarily chosen values within a given range in order to evaluate the system performance.

In order to evaluate the system's performance, different metrics (or estimators) can be used. In machine-learning, metrics are usually referred to as error or loss functions and calculate the difference between a given target (or label) and the current system output. For optimization tasks, L_1 ($||x||_1$) or L_2 ($||x||_2$) error metrics are often preferred due to their simplicity and thus reduced computational complexity. Popular choices include the Manhattan distance and the Euclidian distance (or mean square error, MSE) for L_1 and L_2, respectively:

$$\begin{aligned} Loss_{Manhattan} &= \sum_{n=0}^{m} |Y_{Target,n} - Y_{Output,n}| \\ Loss_{Euclidian} &= \sqrt{\sum_{n=0}^{m} (Y_{Target,n} - Y_{Output,n})^2} \\ Loss_{MSE} &= \frac{1}{n} \sum_{n=0}^{m} (Y_{Target,n} - Y_{Output,n})^2 \end{aligned} \tag{2.11}$$

Here, $Y_{Target,n}$ is the n^{th} element of a target vector $\mathbf{Y_{Target}}$ of length m, and $Y_{Output,n}$ denotes the n^{th} element of the system output vector $\mathbf{Y_{Output}}$.

However, for the implementation of the pulse-shaper a cosine similarity was used as a loss function (see Section 2.3.5 for a comparison). The cosine similarity $\cos\theta$ measures the overlap between two vectors $\mathbf{A}$ and $\mathbf{B}$. In the experiment, the vector $\mathbf{A}$ and $\mathbf{B}$ represent the targeted and returned waveform (i.e., intensity vs. time), respectively, in form of a one-dimensional vector:

$$\cos\theta = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\cdot\|\mathbf{B}\|} = \frac{\sum_{n=1}^{m} A_n B_n}{\sqrt{\sum_{n=1}^{m} A_n^2}\sqrt{\sum_{n=1}^{m} B_n^2}}. \tag{2.12}$$

In general, the cosine similarity $\cos\theta$ returns a value of -1 for opposite similarity while the function returns 1 for complete similarity. For a complete dissimilarity (i.e., without correlation), the returned value is 0.

Since optimization algorithms commonly minimize a given loss function (i.e., optimization towards $f_{loss} = 0$), the cosine similarity requires modification: Firstly, the cosine similarity was subtracted from unity, which provides a minimum only for complete similarity while the other cases return values > 0. Secondly, the function was used as a power of ten to increase the convergence speed as smaller changes in the waveform will still have significant impact on the change in loss function:

$$f_{loss} = 10^{(1-\cos\theta)}. \tag{2.13}$$

The implementation of these modifications allow the optimization algorithm to converge more quickly and robustly.

Lastly, as the waveform fidelity can change drastically while have similar (or small changing) loss values, an additional metric[11] is required to allow for a direct comparison of different algorithms and loss functions. For this, a quality (Q) factor is defined as follows:

$$Q = sgn(\cos(\theta)) \cdot 10^{|\cos(\theta)|}. \tag{2.14}$$

The Q-factor returns values in the range of -10 (inverted similarity) to 10 (perfect similarity).

2.3.5 Results

System Implementation

The complete setup for the smart pulse-shaper is depicted in Fig. 2.14. The laser source is a commercial fiber laser (Pritel FFL) and the sampling source is a commercial frequency comb (Menlo

[11] this metric was not used inside the algorithm but only for comparison of achieved results

FC1500-250-WG). The output of the laser source is directly input to one of the input ports (In_1) of the optical chip. Given the short input pulse duration ($\approx 20\,\text{ps}$) compared to the available delays, the amount of used delays was limited to $N = 5$ (i.e., six switches corresponding to 1 ps to 16 ps) and a fiber-coupled Faraday mirror was used to reflect the waveform at the chip output (Out_1), i.e., the waveforms propagates through the chip twice [139]. This in turn increases the available delay range and combinations in the range of the pump duration. The used chip connections are indicated as defined in Fig. 2.5. After exiting the sample (In_2), the shaped pulse is combined in a standard 2×2 3 dB fiber coupler with the signal from the sampling source before entering the highly nonlinear fiber. Finally, the generated idler is filtered and amplified before detection.

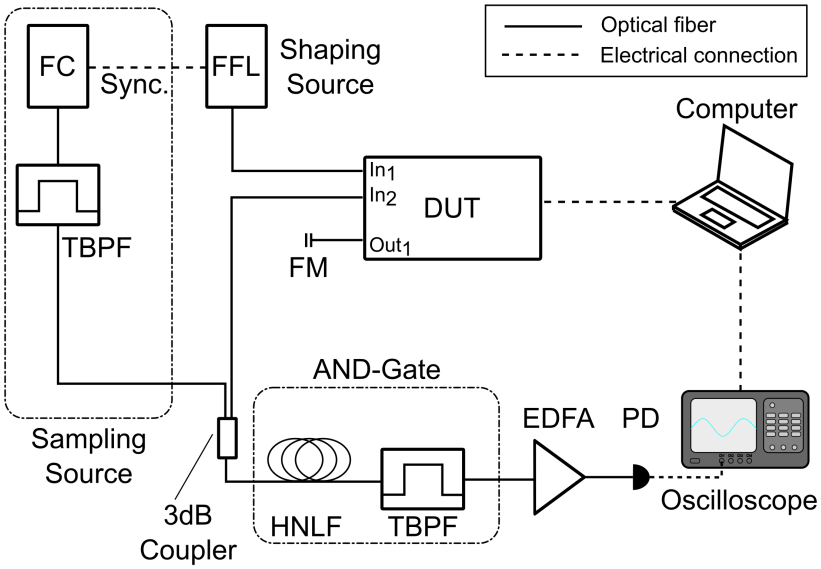


Figure 2.14: Complete Setup for all-optical pulse-shaping. FC = frequency comb, FFL = fiber laser, DUT = device under test (chip), FM = Faraday mirror, HNLF = highly nonlinear fiber, TBPF = tunable bandpass filter, EDFA = erbium-doped fiber amplifier, PD = photodiode.

As outlined in Section 2.3.3, the use of the frequency comb laser allows for straight-forward synchronization of both sources. This is advantageous for increased system stability, as any drift in the repetition rate of the shaping source would directly result in the same change in repetition rate of the sampling source thus ensuring a constant sampling calibration factor V.

Reconfigurable, and Scalable Picosecond Waveform Generation

The experimental results for the case of unconstrained waveforms (i.e., the FWHM of the waveform can vary) are shown in Fig. 2.15. In this configuration, the algorithm returns the waveform that

features the best overlap to the target function regardless of the pulse FWHM. The settings used for the PSO algorithm are summarized in Table 2.4.

Parameter	swarmsize	max. iterations	self adjust. factor	social adjust. factor	max. stall iterations	function tolerance
Value	60	60	1.49	1.49	12	1e-6

Table 2.4: Overview of the PSO parameters used for the pulse-shaping experiment.

The 'maximum number of stall iterations' corresponds to how long the algorithm will 'stall' for (i.e., return loss values that do not change by more than a defined tolerance value) before forcing a termination of the run.

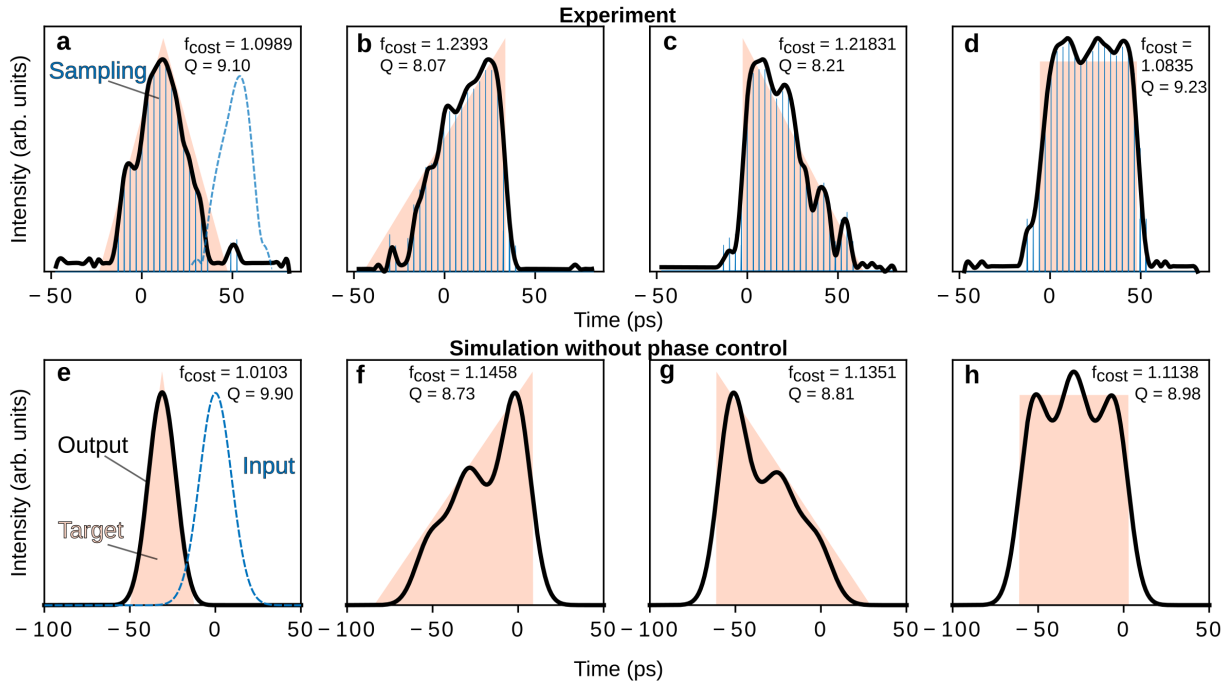


Figure 2.15: Shaping Results Experiment and Simulation. Top row shows the experimental results. Bottom row shows accompanying simulations. Black = returned waveform, pink = target waveform, solid blue line (a-d) = sampling signal, dashed blue line (a,e) = input pulse.

As shown in Fig. 2.15, a wide range of target waveforms can be achieved with the used chip sample, ranging to over 50 ps (in the case of Fig. 2.15h) even using only five delays. Even without access to the phase ($\Delta\Phi_m$ in eq. 2.7), the results for experimental and simulated output pulses match the targeted waveforms reasonably well. This is even more compelling given the simplicity of this chip featuring a limited number of delays (compared to previous chip implementations [145, 146]), as well as the fact that this specific sample was designed for a different task [161]. However, the lack of additional phase control leads to notable modulations in the returned waveform envelopes

(see Fig. 2.21). It should also be noted, that the current implementation lacks sensitivity in case of unconstrained FWHM. This can be seen from Fig. 2.15e, as no shaping occurred for the returned waveform (i.e., the sample only passed the pulse through without any splitting). In this case, the loss function does not change significantly for slight changes in the waveform. However, this is a common issue for (evolutionary) optimization algorithms in general, and can be solved only by means of a more careful designed loss function.

Interestingly, even though the waveforms returned from simulations look identical and feature similar FWHMs, the found amplitude ratios for both cases differ significantly. For example, the positive sawtooth (Fig. 2.15b,f) features FWHMs of $\tau_{exp} = 39.1\,\text{ps}$ and $\tau_{sim} = 46.4\,\text{ps}$ for the experiment and simulation, respectively. The respective amplitude ratios are shown in Table 2.5 and Fig. 2.16 shows the comparison between the experimental results and simulation with the same ratios. The ratios range from 0 (all light propagates through the short path) to 1 (all light propagates through the long path). The splitting ratios for the experiment were obtained from individual characterization curves (see Appendix A.2) as the experiment returns real-valued voltage values instead of ratios. Notably, in the comparison between simulations and experiments, no substantive pulse-shaping occurred with the numerically estimated experimental ratios (besides a small hump at -50 ps) clearly highlighting the limitation of *in silico* designs (vice versa, ratios found by means of numerical simulation might not lead to meaningful shaping in the experimental implementation).

Switch	1	2	3	4	5	6
Experiment	0.56	0.00	0.12	0.08	0.07	0.76
Simulation	0.71	0.00	0.67	0.00	0.48	0.69

Table 2.5: Overview of the amplitude ratios for the positive sawtooth waveform.

The main reason for this discrepancy is simplifications in the simulation model [153, 154]. This includes setting the phase offset per delay ($\Delta\Phi_m$ in eq. 2.7) to zero, whereas in the experiment this phase factor is arbitrary (but constant) for each delay. The exclusion of the MZI crosstalk and extinction ratio in the simulation is another simplification that might contribute to the discrepancy. These findings imply, that in the case of *in silico* approaches (such as previously demonstrated on-chip devices [145, 146]) extensive knowledge or calibration of the sample is required to achieve optimal performance [194]. Hence, the autonomous optimization approach might help circumvent this bottleneck in the design of functional circuits.

This approach does not only bypass potential bottlenecks in the design but also allows for versatility in the pulse-shaping application. While previous approaches were often limited to narrow

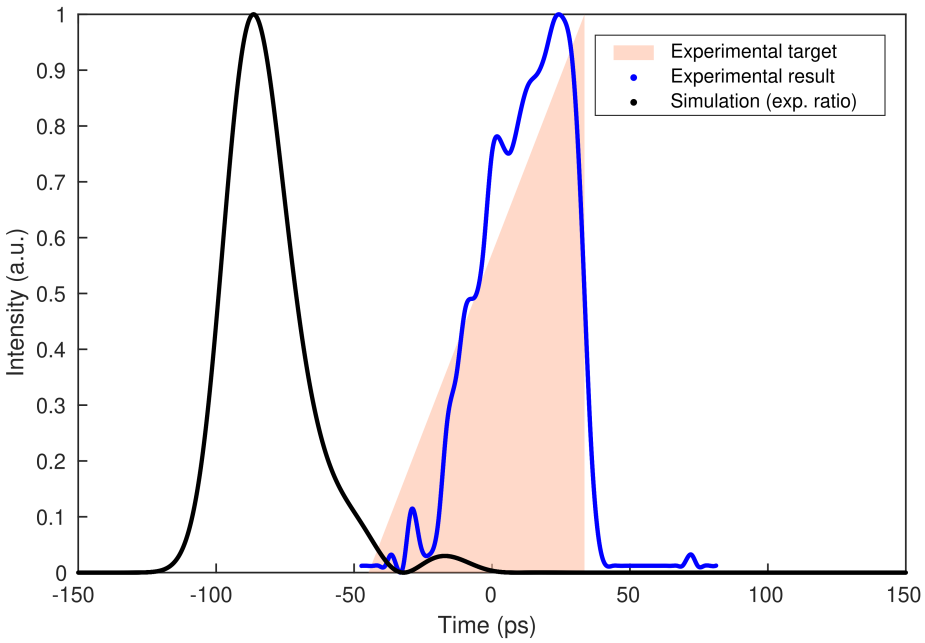


Figure 2.16: Waveform comparison between experiment and simulation for the same split settings. Experimental results (blue, pink from Fig. 2.15b) and simulation results (black) for the same experimentally obtained ratios (Table 2.5).

operational regimes, autonomous optimization approaches offer the ability to achieve larger operational spans. To elaborate on this advantage, waveform generation with a given target FWHM was investigated over a wide target range. The waveform FWHM was chosen randomly within a given range (as opposed to the previously unconstrained FWHM). Additionally, three different input pulse durations were considered: a shorter pulse duration of 3 ps, an intermediate pulse duration of 20 ps and a long pulse duration of 50 ps. The target pulse durations were between 4 ps and 150 ps, depending on the input duration. This allows to study the scalability in terms of input and output pulse duration, as well as the versatility in terms of different target waveforms. Figure 2.17 shows the simulated results of the versatility and scalability of the implemented pulse-shaper.

The overall best achieved output pulse durations (i.e., minimum loss value) for each input pulse duration and the positive sawtooth waveform target (Fig. 2.17a) are 7.7 ps, 55 ps, and 100 ps with targets of 7.3 ps, 56 ps, and 95 ps. For the flattop pulse (Fig. 2.17b), the output (and target) FWHMs are 15 ps, 97 ps, and 146 ps.

It can be seen, that in both cases a minimum can be achieved for the 3 ps input pulse, while the flattop waveform the 20 ps and 50 ps input did not show a minimum for the loss function indicating that the minimum appears at a target duration outside the chosen parameter space. The reason for

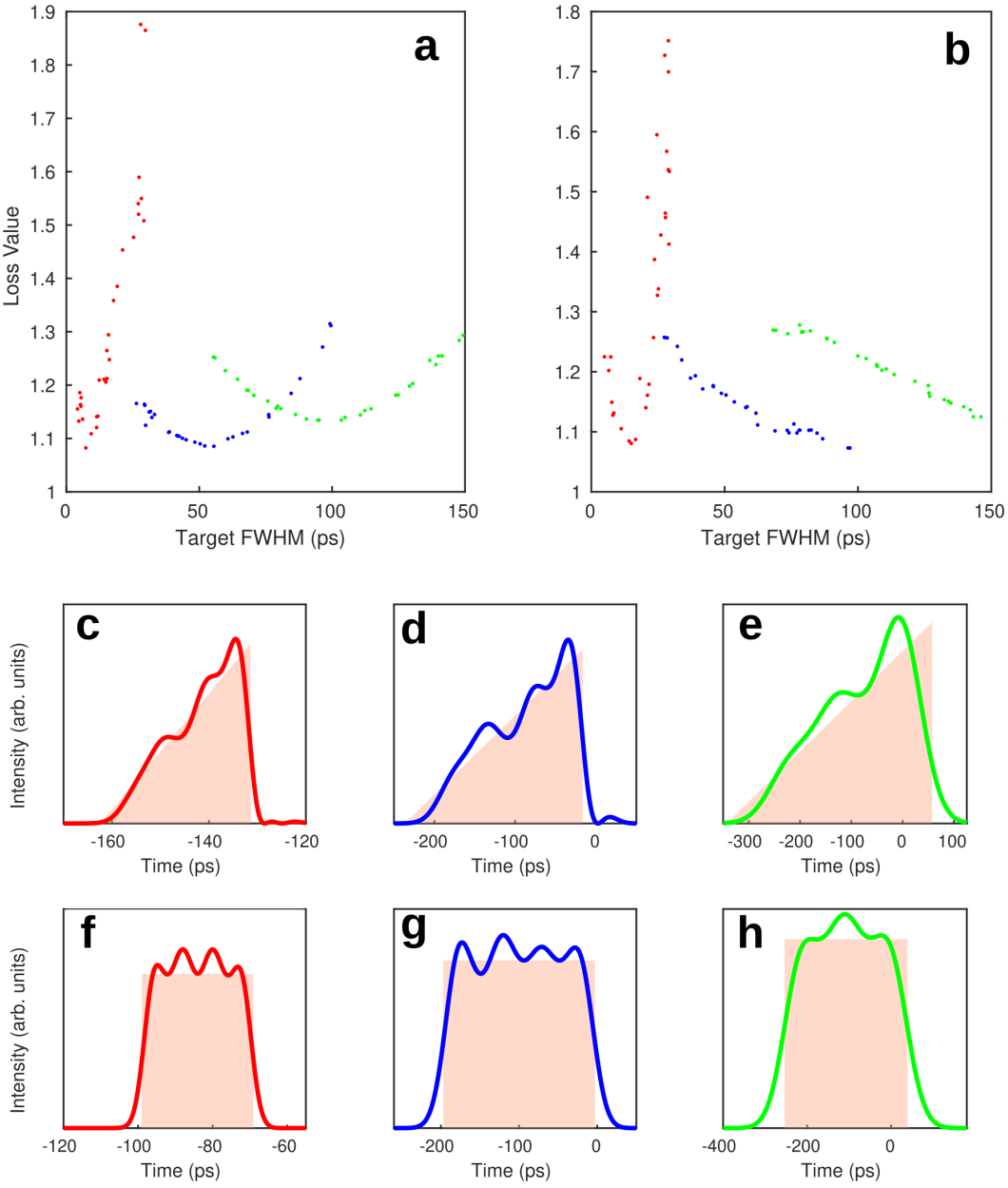


Figure 2.17: System scaling for two waveforms. a) Returned cost value for the positive sawtooth waveform, b) Returned cost value for the flattop waveform. c-e) Waveform for the best cost value of (a). f-h) Waveform for the best cost value of (b). The colors from (c-h) correspond to the colors in (a,b).

the larger deviation in the case of the sawtooth waveform is the appearance of modulations of the pulse envelope due to limitations in the chip design. This includes the resolution (minimum possible delay), the maximal achievable delay (in this case 62 ps), and the possible delay combinations (e.g., a 5 ps pulse requires two delays at the same time: 1 ps and 4 ps). Overall, good waveform fidelity

can be achieved for all tested target durations, as shown by the best performing waveforms in Figure 2.17c-h. The achieved accuracy of the shaping will be discussed in the following sections.

Comparison of Evolutionary Algorithms

As shown in the previous section, there exists a considerable difference between the results obtained from simulations and experiments. In such a case, the parameters found by the simulations are not suitable as starting parameters for the algorithm to accelerate the convergence in the experiment. The commonly used GA however, often depends on good initial system parameter guesses. As such, in case of unknown system parameters (or mixed integer problems) the PSO is expected to perform better than the GA [116, 195–197].
However, as the most common implementation in experimental studies relies on GAs, the performance between the PSO and GA are compared in this section for both cases, the simulations and experiment. To study the performance of both algorithms, the simulation model was used at first. The algorithm parameter used for the comparison are summarized in Table 2.6. For the waveform, a negative sawtooth was selected with an unconstrained FWHM, and as a general evaluation criterion the loss value from the results in Fig. 2.15g was used ($f_{loss} = 1.1351$, $Q = 8.81$) as a threshold criterion. Using this value, the algorithm performance was evaluated as the point where the threshold was surpassed in terms of runtime and required populations (i.e., function evaluations). Additionally, to investigate if the algorithms (both the GA and PSO) provide any benefit at all, the study was also repeated with a random search algorithm [118, 119]. A random search is an useful approach were a number M of random parameter sets are generated and tested over N iterations and the best result is kept. This allows to evaluate if by only choosing random parameters over and over again, the same results as an elaborate algorithm can be achieved. For the random search, M=60 and N=100 was used, thus providing a similar amount of tested settings as both optimization algorithms.

Figure 2.18 shows the results for the loss value for the two different population and swarm sizes. As it can be seen, both algorithm approaches significantly outperform the random search approach (Fig. 2.18 inset), thus validating the use of elaborate optimization approaches. The difference between both algorithms on the other hand are less significant. It can be seen, that in all cases the PSO converges faster (i.e., reaches lower loss values faster). This is especially of importance

	PSO		**GA**
swarmsize	20 / 100	initial population size	20 / 100
max. iterations	100	max. iterations	100
self adj. factor	1.49	crossover	single-point
social adj. factor	1.49	crossover fraction	0.7
tolerance	1e-6	tolerance	1e-6
max. stall iterations	20	max. stall iterations	20

Table 2.6: Overview of the PSO and GA parameters used for the algorithm comparison.

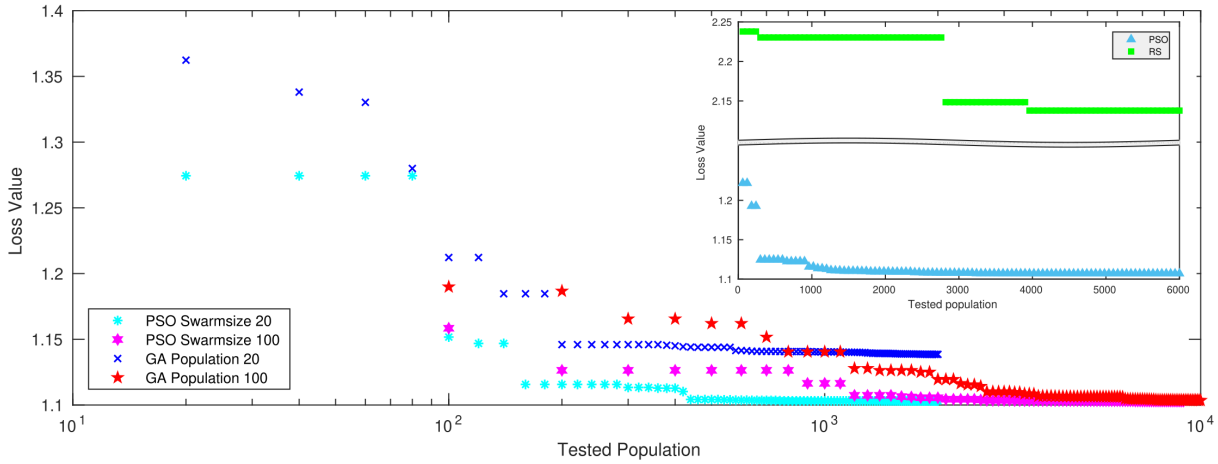


Figure 2.18: Comparison in convergence behavior for the PSO, GA and RS. The PSO shows for both population sizes a faster convergence compared to the GA. Moreover, for smaller population sizes the PSO achieves significantly better performance. Inset: Comparison between the PSO and a RS for the same amount of tested populations. The algorithm performs significantly better than repeated guessing.

in experimental implementations where the total measurement and detection time scales with the number of tested individuals. The performance outcomes of the algorithms are outlined in Table 2.7.

Algorithm	**Swarmsize / Initial Population**	**Q-Factor**	**Tested populations** until $f_{loss} = 1.1351$ ($Q = 8.81$) reached	**Converted Runtime**
PSO	20	9.07	160	3.4 s
PSO	100	9.07	200	3.8 s
GA	20	8.78	>2000	>39 s
GA	100	9.06	1200	21.6 s

Table 2.7: Overview of the simulation results for the PSO and GA algorithm.

Notably, the PSO performs better for smaller populations compared to the genetic algorithm in both the runtime and final loss value, while for large populations the final loss values are almost equal. However, in case of larger populations the PSO run reached the 20 stall iterations (i.e., the change of the loss value was less than the function tolerance), whereas the GA required the full 100 iterations without any stall iterations. Thus, the PSO effectively required significantly less

iterations compared to the GA (9000 tested populations incl. 2000 from the stall iterations vs. 10000 for the GA). This is of significant importance especially for experimental implementations in order to limit the overall required time to perform a given optimization task.

In order to investigate the performance of both algorithms experimentally, a similar study was conducted. For this, the positive sawtooth waveform was optimized with the PSO and GA. Here, the algorithm parameters were the same as in Table 2.6, but with small changes. For the GA, the population size and iterations were chosen as in a previous study [198] with only a few iterations (15) but a larger population size (500). This optimized parameter set was used to enhance the genetic diversity and thus provide a better exploration of the search space. For the PSO, the swarmsize and max. iterations were 100, and 20 stall iterations were allowed for both algorithms (function tolerance 1e-6). The shaping result is shown in Fig. 2.19.

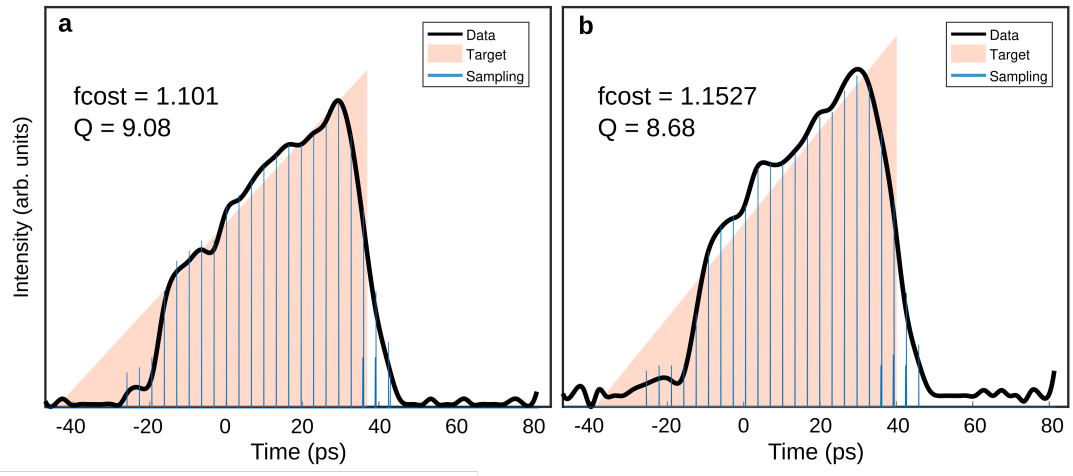


Figure 2.19: Experimental comparison of shaping fidelity between a) PSO and b) GA.

As a result, both algorithms are overall performing well, however the PSO achieves a slightly better waveform fidelity ($Q = 9.08$, $f_{loss} = 1.1010$) compared to the GA ($Q = 8.68$, $f_{loss} = 1.1527$). Similar to the simulations before, the PSO used all stall iterations (2000) while the GA required all iterations (15). Compared to the total number of tested populations (9800 for the PSO, 7500 for the GA), the PSO confirms the simulation findings, also converging faster in the experiments (better performances for the same number of iterations). Interestingly, when the runtime of both algorithms is compared, the PSO finished after ≈20 hours, whereas the GA required ≈30 hours. While the algorithm in itself should not have much difference for problems with low computational complexity (as it can be seen from the simulation in Table 2.7, both algorithms require approx.

20 ms per evaluation), the increased runtime might be caused by the post-processing, specifically the Hilbert transform used for the envelope retrieval. Indeed, from the experiments it was evident that non-uniform waveforms (i.e., more irregular) require more evaluation time compared to more uniform waveforms (e.g., closer to the target waveform). Since all particles of the PSO naturally covey towards the global minimum, a higher number of the population features a uniform waveform speeding up the convergence/evaluation. On the other hand, the GA does not naturally include such a mechanism[12] (i.e., populations might scatter more throughout the optimization progress) thus might explain the increase in experiment time as it features more non-uniform waveforms in the beginning due to the large population size (500). Thus, another important factor in choosing a suitable optimization algorithm might be the overall post-processing techniques.

Similar trends have been also reported for other integrated platforms [116] or in inverse component design [197]. A potential explanation for this is the fundamental principle behind the algorithms. The smaller swarmsize in case of the PSO leads to a faster exploration of the search space, while the small population size of the GA lacks genetic diversity, hence featuring a low exploration. It should be noted at this point, that this is usually highly problem dependent (exceptions have been found in the past [195]), however in case of unknown or unconstrained problems the PSO might be a better choice compared to the GA as likewise indicated by recent theoretical and experimental realizations [116, 199].

Evaluation of System Accuracy And Limitations

Finally, not only the scalability is an important aspect of the autonomous pulse-shaper, but also the system limitations. In particular, the achievable accuracy and waveform fidelity are of importance for any real-world application. In this section, the waveform accuracy of the system scaling (from Fig. 2.17) will be investigated, as well as potential improvements.
Figure 2.20 depicts the overall system performance of the scalability for the positive sawtooth and flattop target functions. For this, the returned waveforms were evaluated with an ordinary least square fit of the respective waveforms to retrieve the FWHM of the returned (shaped) pulse.

From this result, it is evident that the flattop pulse performs exceptionally well, while the sawtooth deviates more from the ideal case (i.e., a slope of 1). Nevertheless, the performance of

[12] a potential solution could be offered by advanced elitism options but was not used in this study

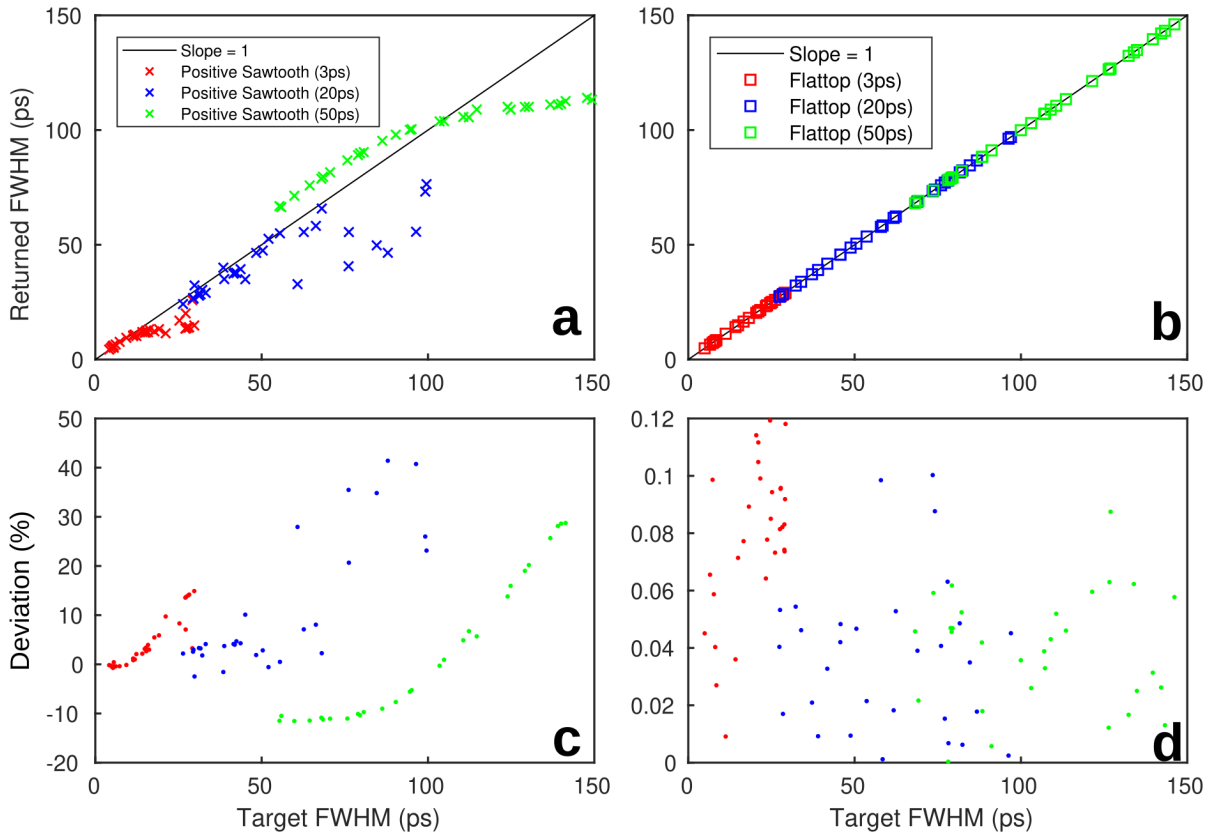


Figure 2.20: System accuracy for two waveforms. a,c) Positive sawtooth, b,d) Flattop. Notably, the two waveforms show vastly different deviations between the returned and target FWHM.

the positive sawtooth is still reasonable, as the majority of returned FWHMs are within a ±10% deviation range. The shaping accuracy might be improved by using a chip sample, which features individual phase control per delay or a better delay configuration (such as doubled delays).
Notably, as it can be seen in Fig. 2.17h the flattop can feature more Gaussian-like waveforms where the key feature (i.e., the flatness) is not well achieved, while the FWHM is still close to a perfect match (explaining the diminutive errors). To account for such specific waveform features or target requirements, additional constraints are obligatory for the chosen loss function, such as penalties for non-flat tops.

As highlighted in the previous Section (2.3.5) and in Fig. 2.15, the lack of phase control is likely to cause notable modulations in the returned waveforms. To investigate the effect of additional phase control, the simulations from Fig. 2.15 were repeated (same input duration and number of switches) but with two adaptions: Firstly, both, the amplitude ratios and the phase offset $\Delta\Phi_m$ were optimized by the PSO (effectively doubling the parameter space), and secondly, a fixed target

pulse duration was given for all cases ($\tau_{Target} = 35\,\text{ps}$) instead of finding the best performative settings (i.e., lowest loss value). The latter ensures to allow for a direct comparison of a given delay combination and avoid unwanted effects such as inaccessible delay combinations.

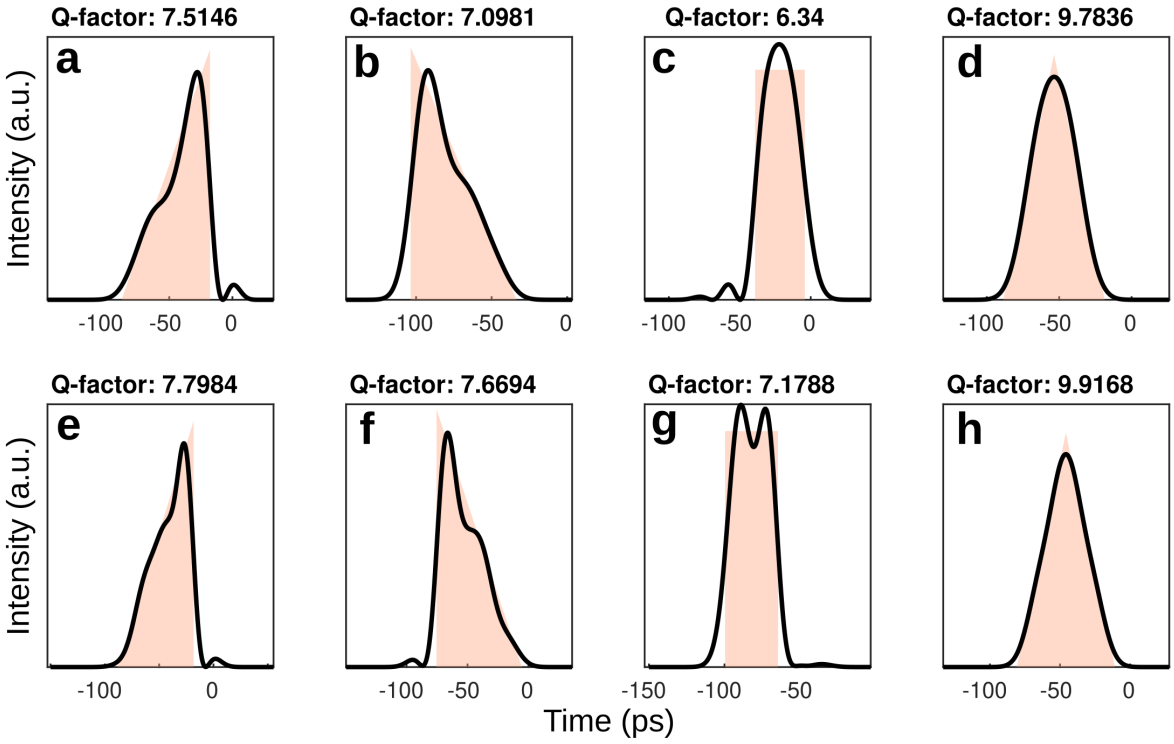


Figure 2.21: Shaping simulation with phase control for fixed FWHM. a-d) Simulation without phase control. e-h) Simulation including phase control. A clear improvement in waveform fidelity can be achieved for (e-h).

The improvement in shaping fidelity as a result of the included phase offset is evident from Fig. 2.21. Indeed, in all cases the waveform fidelity was improved from a small increase of 1.36% for the triangular waveform, up to more than 13% for the flattop, and detailed performances for each target waveform are summarized in Table 2.8. While there is still potential for improvement (e.g., better rise/fall time or flatness for the flattop waveform), the results demonstrate that devices featuring both, amplitude and phase control, have superior performance to devices with amplitude ratio only. However, the inclusion of additional phase control per delay also doubles the available parameter space, which might increase experimental time and system complexity (for both, the photonic chip and electrical control).

Another important factor for the shaping accuracy is the choice of loss function. As outlined in Section 2.3.4, there exist many different types of loss metrics ranging from straight-forward L^p

Target Waveform	Without Phase Control	With Phase Control	Improvement (%)
Positive Sawtooth	7.5146	7.7984	+3.7772
Negative Sawtooth	7.0981	8.0488	+8.0488
Flattop	6.3400	7.1788	+13.2297
Triangle	9.7836	9.9168	+1.3613

Table 2.8: Comparison of achieved Q-factor between ideal (with phase control) and non-ideal (without phase control) configuration for different waveforms.

norms comparing single parameters to highly customized and application specific metrics [200]. In light of the vast range of available loss metrics, a selection of popular loss metrics (see eq. 2.11) is compared in the following. In detail, four different cases were studied: L1 (Manhattan distance), L2 (Euclidian distance), cosine similarity, and mean square error (MSE). For the simulations, a positive sawtooth was chosen with a fixed target FWHM ($\tau_{Target} = 50\,\mathrm{ps}$) and six switches (no phase control). Furthermore, the maximum number of iterations was enlarged to 200 and the function tolerance was decreased to 1e-12 to account for smaller changes in the different loss functions. Fig. 2.22 shows the resulting waveforms for each loss metric.

For comparison, the Q-factor was determined by applying the cosine similarity to the returned waveform of each metric to allow for direct comparison of the waveform fidelity between different norms (as they return different loss values). As it can be clearly seen, the cosine metric is by far the most performing, with a Q-factor of twice as much as in the other cases. The reasons for the subpar performance of the distance metrics is twofold. Firstly, they might feature a limited sensitivity[13] compared to the used cosine similarity. This may result in a large ambiguity of the obtained loss values and as a result lead to a more noisy loss landscape, where ideal and non-ideal solutions are relatively close in switch settings and/or loss value (or ideal solutions feature worse loss values [88]). In such cases, more elaborate versions of the optimization algorithm (e.g., adaptive self- and social parameters) may increase the performance. Secondly, it has been shown in the context of text and speech analysis [39, 201] that cosine similarities are better suited for sparse vectors [39]. Indeed, in the case of the demonstrated on-chip pulse-shaper, the measured waveforms are represented by one-dimensional vectors for the target and output waveform and have a large data size (more than 5 million datapoints in the experiment, 2^{13} datapoints in the simulation), hence calculated differences between output and target might be very small, and thus might be responsible for the deteriorating performance of the distance metrics.

[13] the metrics were also tested (a) as power of ten and (b) with the GA, showing the same results

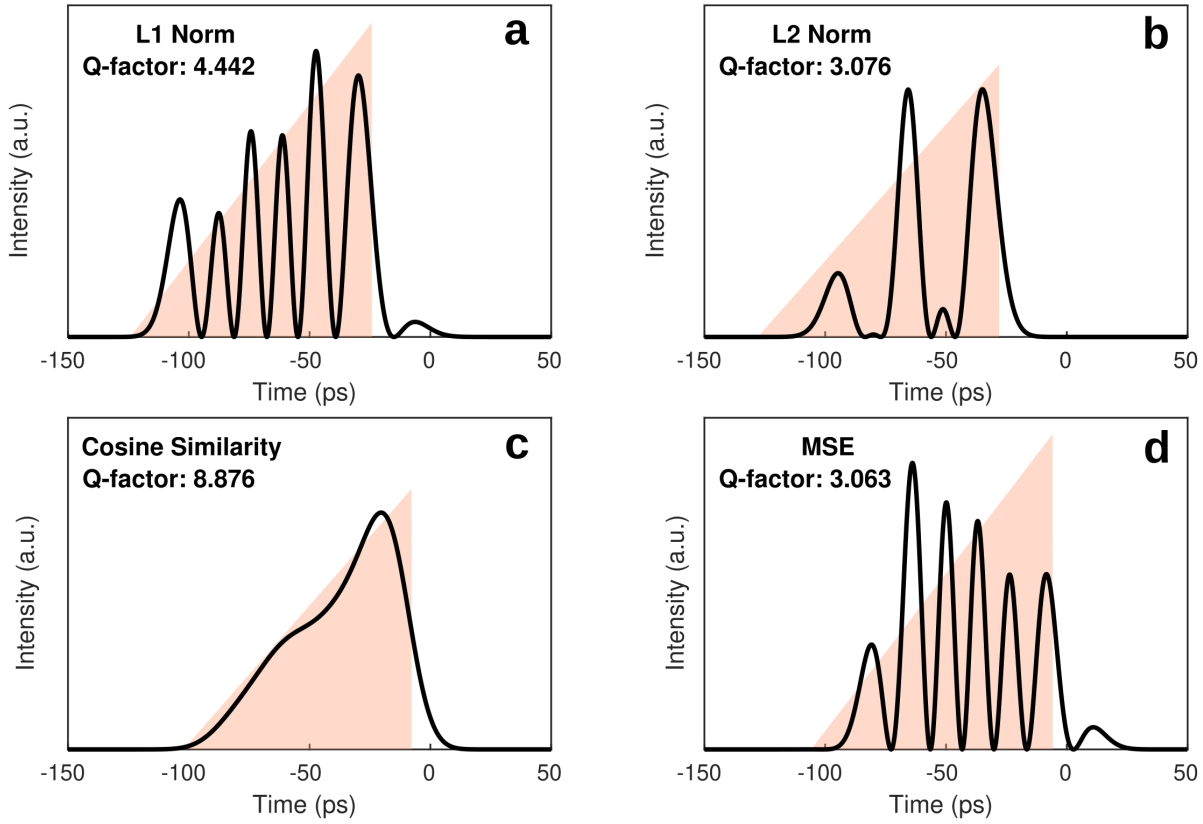


Figure 2.22: Comparison of different loss function for the pulse-shaping. a) L1 metric, b) L2 metric, c) cosine similarity metric, d) mean-square error (MSE) metric.

Lastly, the chip and system design in itself leads to several limitations. As mentioned before, chip limitations include the resolution (minimal delay), the span (maximal achievable delay), possible delay combinations, or access to the phase. As a result of unobtainable delays, severe waveform modulations are appearing. Especially the minimum delay dictates the minimum input pulse duration achievable with the on-chip TCS approach. For input pulses shorter than the shortest delay of the sample, significant pulse-shaping might not be viable (for continuous output waveforms). However, in such a case quasi-waveform [148, 199] or ultra-fast pulse train generation [42] are viable possibilities.

2.4 Conclusion

In conclusion, a novel approach for an integrated continuous picosecond pulse-shaper based on temporal coherence synthesis was investigated and for the first time successfully demonstrated.

The device is based on a concatenation of individually controllable Mach-Zehnder interferometers on a low-loss glass platform. Moreover, the implemented concept features a low energy consumption of below 2 W (excluding source and detection equipment). The majority of power consumption is caused by the voltage required to hold the amplitude ratio of each switch (approx. 300 mW electrical power for the highest voltage setting per switch), which can be further improved through elaborate system engineering.

In particular, the implementation of the cost-effective all-optical sampling scheme has been a key-enabler for fast and unambiguous data acquisition at low power consumption and equipment costs. As such, non-linear optics-based diagnostics can be a valuable tool for the emerging *smart*-photonics research field to enable out-of-the-lab applicability at reduced cost and complexity. Future improvements for such a scheme includes optimization of the nonlinear element. For example, the use of integrated, nonlinear waveguides [163] or the use of novel nonlinear fibers [202–204]. Indeed, a full integration of this system in a single material platform[14] might be possible in the near future, allowing for a compact, and stable picosecond pulse-shaping device. In particular, the sampling source can be implemented by means of efficient Kerr combs [208, 209], which emit soliton pulses with durations of <1 ps. For the readout, integrated waveguides or field reconstruction approaches (e.g., on-chip SPIDER) could be utilized as well [163, 210].

The obtained results indicate, that, despite using only amplitude-ratio optimization, good pulse-shaping results can be obtained for a large range of input and output pulse durations. Indeed, the simulations show that the platform can reach over 100 ps (FWHM) output pulse durations with only five delays. It was further shown, that the inclusion of additional phase control is able to further improve the waveform fidelity at the cost of higher system complexity (i.e., more electronics such as DACs required). The inclusion of more delays might open a path towards covering the regime between 100 ps and 1 ns, however at the expense of more involved device fabrication. Further improvements in the chip design might include different delay combinations such as double delays (instead of using the Faraday mirror at one output port of the device).

Remarkably, the results of this work give strong evidence that the utilization of optimization algorithms allows to circumvent fabrication imperfections, which often limit common *in silico* approaches and require, for example, elaborate characterization methods [194]. Moreover, as demon-

[14] recent progress in heterogeneous integration is also a viable option to leverage advantages of different material platforms [205–207], for example, the inclusion of optial amplifiers or modulators [147]

strated, the use of a straight-forward particle swarm optimization algorithm can outperform the commonly used genetic algorithm for an unconstrained problem (i.e., no good initial guess is available). These findings for this particular platform and application are consistent with recent studies on an integrated Mesh-type platform for signal processing [116]. The faster convergence of the PSO might be particularly attractive for ultra-fast optics (e.g., smart lasers), where fast system dynamics occur. Nevertheless, it should be highlighted that there is no generality in this. The use of one algorithm over another is often highly problem dependent and should be studied in detail for a specific problem [111, 195, 196]. Also, a combination of different approaches might be suitable for more complex problems [211, 212].

Finally, it is important to emphasize that guided optimization approaches are only suitable for finding a given set-point and are generally not capable to continuous adjustment, i.e., correcting for changes in the environment. The present study was carried out under laboratory conditions, where the influence of the environment is insignificant. If environmental changes are to be taken into account, the use of continuous techniques is required. For example, in the field of smart lasers, the combination of optimization techniques with local search algorithms [31] or the use of deep reinforcement learning strategies [213, 214] have been shown great promise and could also be a suitable next step for the presented technique.

Chapter 3

Photonic Machine Learning

Abstract: The implementation of efficient analog computing concepts is of high interest to circumvent the increasing energy consumption of current digital neural network approaches. As such, photonics is of particular interest due to the wide-spread use of optical data transmission in the telecommunication sector and its inherent advantages such as high transmission parallelism (i.e., different degrees of freedom), low transmission loss, and low latency. In the following, the use of highly nonlinear, complex pulse dynamics, in a single monomode optical fiber, is investigated towards the application in neuromorphic computing. First, the background for neural networks, and the state of the art in photonic neural networks is reviewed. Subsequently, broadband nonlinear light generation (i.e., supercontinuum generation) by means of nonlinear pulse-propagation is introduced. Following the analogy between the modeling of such pulse propagation, the physical similarity between the nonlinear differential equation describing the nonlinear propagation in optical waveguides and neural network functionality is elaborated. Then, different benchmark tasks from the machine-learning field will be implemented in the experimental system to assess the functionality towards tasks of different complexity. Additionally, the performance will be evaluated and compared to other optical platforms. Finally, an outlook for the implemented system is provided and future work is discussed.

The results presented in this chapter have been published in the following:

Journals:

B. Fischer, M. Chemnitz, Y. Zhu, N. Perron, P. Roztocki, B. MacLellan, L. di Lauro, A. Aadhi, C. Rimoldi, T. Falk and R. Morandotti, "Transient Optical Neural Emulation," *submitted, under review* (2023).

Patents:

B. Fischer, P. Roztocki, M. Kues, M. Chemnitz, C. Rimoldi, B. MacLellan, L. Romero Cortes, J. Azaña, Y. Jestin, and R. Morandotti, System and method for optical information processing with a reconfigurable nonlinear optical network, US 63/051,435 (patent pending).

3.1 Machine Learning Algorithms and Concepts

Machine learning concepts, particularly those based on neural networks (NNs), are the prime driver of a 'fourth industrial revolution' [1], already impacting information processing across a wide range of technologies [2, 215, 216]. Indeed, neural networks are becoming increasingly used in finding hidden patterns in vast data amounts, at interpolating from such data or in recognizing specific features. In particular, NNs enable both static and dynamic process automation and the development of new 'smart' applications, e.g. for natural language processing [217], medical diagnostics [218], or autonomous vehicles [219], to name a few.

In general, neural networks can be distinguished in terms of architecture and training strategy. The network architecture can be broadly categorized into unidirectional, recurrent, and multidirectional NNs (e.g. Hopfield networks [220]). Unidirectional networks, better known as feed-forward neural networks (FFNN) are characterized by a unidirectional flow of information within the network, i.e., from input to output. Prime examples for this type are deep and convolutional neural networks which are the state-of-the-art for applications such as image recognition [215]. In contrast, RNNs feature an intrinsic memory that can store past information [221], which makes them particularly suitable for processing of dynamic data such as for example text, speech or audio signals[215]. However, in principle, both architectures can and have been used for the opposite use cases albeit showing a certain inefficiency [222]. Interestingly, in recent years researchers started to combine different architectures types in order to reduce the training complexity or leverage specific features of different architectures [223–226].

The types of training strategy can be broadly categorized into supervised, unsupervised or reinforcement learning, sometimes with overlap between each other (e.g., semi-supervised). The general idea behind the first two concepts is the existence or absence of labeled training data. Obtaining useful labeled data is a tedious (as it is usually performed manually) and expensive

task especially for large data sets [227]. Unsupervised learning is often used for clustering (e.g., finding similarities and grouping of data) [228], or discovering patterns and structures in data. Reinforcement learning (RL, and variants such as deep RL) is an increasingly popular learning strategy. Analogously to the learning behavior of sentient beings, RL does not necessarily require labeled data but can learn from experience by rewarding a positive outcome (or vice versa penalizing a bad outcome): in a nutshell, by trial-and-error. This learning approach has been applied to several problem cases and was particular successful for computer games such as the famous AlphaGo [229] or more complex racing games, which necessitate real-time processing capability [230]. Therefore, it is of great interest to the optics community, especially for experimental implementations [214].

In the following, individual building blocks and core functionalities of neural networks are described as a precursor for optical neural networks.

3.1.1 The Single Neuron

The backbone of most neural networks are artificial neurons (ANs), which are aimed to mimic the behavior of neurons in a (human) brain. In a nutshell, biological neurons receive (weighted) input information, which is subsequently summed up and eventually nonlinearly processed before reaching the next neuron(s). An important distinction between a biological and artificial neuron is the type of nonlinear processing. While the biological neurons usually work as spiking neurons (i.e., they fire once they are triggered or not) [231], an artificial neuron can feature a larger variety of activation functions, ranging from 'hard' threshold such as step-like functions in the perceptron model to more continuous functions (i.e., fire a little) such as sigmoid [232] or ReLU (rectifier linear unit [233]). While spiking neurons[1] are an active field of investigation (also for photonic implementations), this topic is beyond the goal of this thesis [234–236].
Notably, also the way of learning differs between artificial and biological neural networks as outlined in the next sections.

On a more technical level, there exist several differences between biological neurons (BNs) and ANs. In particular, ANs and BNs feature different bit depths, which drastically affects computational (and training) time, model accuracy, and memory requirements. Computers nowadays generally operate on a 64-bit architecture with weights often stored as 32-bit floating point num-

[1] physical spiking neuron concepts work generally on discrete, i.e., pulses, rather than continuous signals [77]

bers, although models are sometimes restricted to lower bit depths (known as 'quantization') as a trade-off between computational resources and performance [237, 238]. Yet, a biological synapse (responsible for the 'weighting' [239]) is estimated to have only a bit depth of around 5 bit [240] while performing more efficiently and accurately (e.g., for image classification). Moreover, 'weights' and connections in BNs are able to regrow over time while adapting to new problems [241, 242], feature individual time-delays (i.e., memory) of different time-scales [243, 244] or even experiencing nonlinearities during the signal summation [245, 246]. Lastly, perhaps most importantly, the brain network is highly efficient compared to its digital analogue. It is estimated to consume only around 20 Watts of power [247] for $\approx 10^{11}$ neurons [248, 249] with up to 10^{15} synapses [250], while computer networks on the other hand require several orders of magnitude more power (e.g., brain simulations on a supercomputer require more than 10 MW[2]).
Indeed, it has been recently demonstrated that networks of *in vitro* neurons can be much more efficient at learning to play simple games, such as Pong, compared to computer approaches [252]. In general, brains (not only human) are exceptionally efficient, for example fruit flies and honeybees have only around 10^5 and 10^6 neurons, respectively, yet they can perform many different complex tasks [253] (for comparison, the IBM TrueNorth neuromorphic chip features 10^6 neurons on $4.3\,\mathrm{cm}^2$ [68]). Artificial neural networks on the other hand are usually tailored towards single tasks and require numerous trainable parameters. For example, one of the most performing neural network for the CIFAR image classification requires over 630 million trainable parameters [254]. It is difficult to apply the same neural network architecture to different tasks. A prime example is speech recognition (what is spoken) vs. speaker identification (who is speaking), known as the 'cocktail party'-problem [255].

In the past, single neurons have been essentially implemented in two ways. Either as a perceptron (i.e., using a step-like activation function [256, 257]) or as an artificial neuron with a different activation function. In fact, both concepts share significant similarities and mainly differ in choice of nonlinear function and training algorithms, and as such are often synonymously used. Figure 3.1 illustrates a single artificial neuron.

[2]estimate on the SUMMIT supercomputer [251]: `https://www.top500.org/system/179397/`(accessed Jan. 26th 2022)

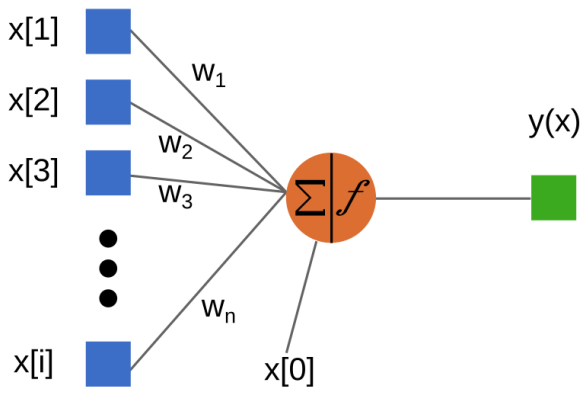


Figure 3.1: Illustration of an artificial neuron.

In detail, an input vector $\mathbf{X}$ containing $x[i]$ elements ($x[0]$ denotes a bias term) is weighted by a weight vector $\mathbf{w}$ and subsequently nonlinearly transformed to generate an output $\mathbf{y}(\mathbf{X})$. Mathematically, this can be described as [232]

$$\mathbf{y}(\mathbf{X}) = f(\mathbf{w}^T\mathbf{X}) = f\left(\sum_{i=1}^{n} w_i x_i + x_0\right). \tag{3.1}$$

with $\mathbf{w}^T$ as the transposed weight matrix and $f(\cdot)$ as a nonlinear activation function. The goal is to minimize a given loss function between the returned output $\mathbf{y}(\mathbf{X})$ and the true class label (i.e., supervised learning).
For the perceptron the activation function is defined as [232]

$$f(z) = \begin{cases} +1 & z \geq 0 \\ -1 & z < 0. \end{cases} \tag{3.2}$$

Indeed, this is an important distinction of a perceptron to an artificial neuron. Since the majority of training methods are based on gradient calculations (e.g., stochastic gradient descent), the activation functions are required to be differentiable. Other more popular choices for the activation function for artificial neurons (and more so for neural networks) are the hyperbolic tangent, sigmoid or ReLU[3] function [258]:

[3]The ReLU function also features variants such as leaky-ReLU, parametric ReLU, or nonlinear variants such as Softplus to account for certain drawbacks of the original ReLU

$$
\begin{aligned}
f(z) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} && \text{(hyperbolic tangent)} \\
f(z) &= \frac{1}{1 + e^{-z}} && \text{(Sigmoid)} \\
f(z) &= \max(0, z) && \text{(ReLU).}
\end{aligned} \tag{3.3}
$$

It is noteworthy, that essentially any (nonlinear) function could be used as an activation function, although within certain constrains [258]. More complex functions have been used in the past for specific problems [259], however require more computational power for evaluation and training. Hence, for practical implementations, often simple, yet sufficient, nonlinear functions are used (such as those in eq. 3.3). The choice of nonlinearity for the neurons is important and manifold, and will be discussed in the subsequent sections.

In summary, artificial neurons can be straight-forwardly implemented by a linear matrix multiplication (i.e. weighting and summing) and a nonlinear activation. This simplicity (and abstraction of a biological neuron) provides an excellent base for the implementation of more complex networks and subsequent physical implementations.

3.1.2 Multi-Layer Neural Networks

While the single neuron (and perceptron) has proven a simple, yet powerful tool in the machine-learning field, it reaches its limit at handling (significantly) nonlinear tasks. For such tasks (for example the n-Bit parity[4] problem) a single perceptron/artifical neuron is unable to provide a solution [260]. As a consequence, multi-layer networks have been developed [259, 261, 262]. It should be noted at this point, that multi-layer perceptrons (MLP) and feed-forward neural networks differ again only in the used activation function (discontinuous vs continuous activation, respectively [232]), although MLP is often used synonymous for multi-layer networks. Figure 3.2. illustrates a general multi-layer network architecture.

Similarly to the single neuron case, in a multi-layer network each neuron receives a weighted input signal, but instead of a single output they are connected to the neurons of the next layer. If all neurons are connected with each other, the network is referred to as a fully connected neural network, whereas a limited number of connections is termed sparse neural network. The optimization of

[4]for n=2 bits, the n-Bit parity test equals the exclusive OR (XOR) operation

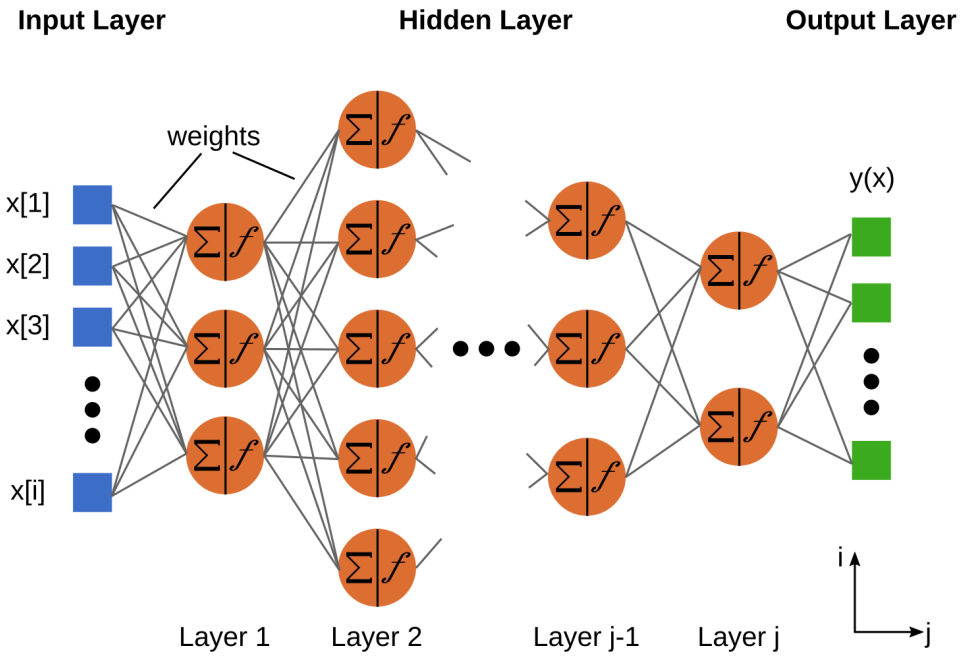


Figure 3.2: Illustration of a deep neural network. The illustration of a bias for each neuron was omitted for clarity.

connections is referred to as 'pruning' [238]. If the number of processing layers (not counting the input and output as a layer, blue and green, respectively in Fig. 3.2) exceed one[5], it is often referred to as a deep neural network, and subsequent layers as hidden layers (concurrent the neurons in a hidden layer are called *hidden units*) [232].

Expanding equation 3.1 to multiple layers provides the following output of layer j with i and n neurons for layer j and $j+1$, respectively:

$$\mathbf{x^{j+1}} = x_n^{j+1} = \sum_i \left(w_{in} x_i^j + b_i \right) = f\left(\mathbf{w}\mathbf{x}^j + \mathbf{b}^j \right). \tag{3.4}$$

3.1.3 Impact of Nonlinearity

Nonlinearity plays a pivotal role in neural network implementations. In the absence of any nonlinearity, any neural network would only perform a linear mapping of the input data, regardless of the architecture or network size (e.g., several linear layers can be condensed to a single linear layer). This is especially important as most 'real-world' problems are nonlinear in nature, thus

[5]this terminology is ambiguous throughout literature as some also count the input and/or output layers, or specify the amount of layers regardless, e.g., 2-layer network, 3-layer, etc.

linear networks are not able to perform meaningful operations on such data (e.g., for classification tasks)[6]. The choice of nonlinear function in itself is important [258], as the majority of training algorithms (e.g., stochastic gradient descent) require differentiable functions for optimum convergence[7]. Notably, very large (and thus complex) nonlinear networks can offer additional advantages, such as strong generalization capability even with limited data sizes or the ability to avoid the use of regularization methods [264].

The impact on the nonlinear data transformation can be visualized in different ways. A simple yet powerful example is the XOR task[8]. The XOR operation outputs a logical **0** for equal inputs (blue crosses in Fig. 3.3a), and a logical **1** for odd inputs (red dots in Fig. 3.3a). The raw data are illustrated in Fig. 3.3b, where it is evident that a single straight line (as provided by linear classifiers such as the perceptron) is unable to separate the data. By mapping the data into a higher-dimensional feature space via a nonlinear network or operation (Fig. 3.3a), the data can easily be linearly separated. The n-1 dimensional plane that separates the projected n-dimensional data is called an hyper plane.

To illustrate the behavior of the neural network for this task, a straight-forward multi-layer neural network is implemented with one hidden layer containing two hidden units. The activation for the hidden units is a hyperbolic tangent ($\tanh(\cdot)$) function and the output unit uses a sigmoid function. The training is performed for 100 epochs (i.e., complete processing of 100 randomized data batches). By evaluating the output and weights of the hidden nodes after training, it is possible to visualize the data transformation (Fig. 3.3c).

From this small experiment, it is evident that nonlinearity in neural networks has a drastic impact on data transformation and is thus a crucial feature for the successful application of neural networks towards processing real-world data.

3.1.4 Neural Network Training

The ultimate goal of network training is the adjustment of the weights $\mathbf{w}$ of the network towards minimizing the error between the system output $\mathbf{Y}_{Output}$ and the ground truth $\mathbf{Y}_{Target}$ towards

[6] it should be mentioned, that deep linear networks exist which feature aspects of nonlinear networks such as nonlinear information trajectories [263]

[7] exceptions exist, such as the popular ReLU function, which is not differentiable at 0

[8] example based on the code from `https://gist.github.com/CihanSoylu/6967249574192728a9fba367065e8949` (accessed Jan. 27th 2022)

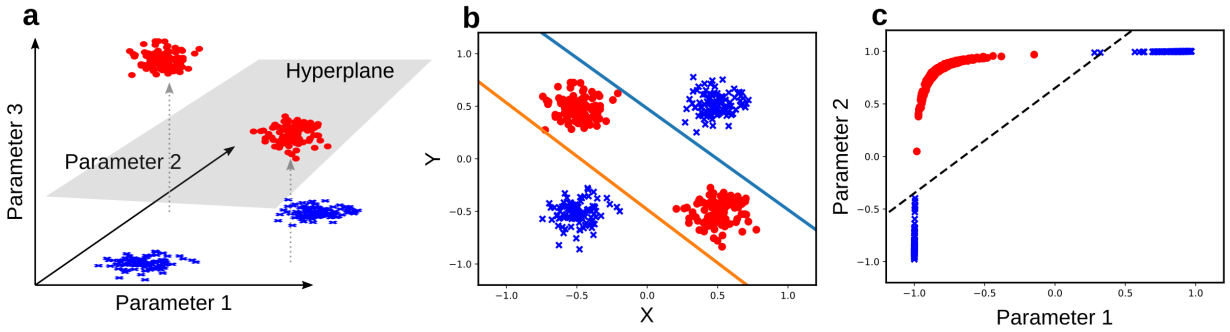


Figure 3.3: Visualization of the role played by the nonlinearity. a) Example on how data can be projected into a higher dimensional space (indicated by the dashed arrows), separable by a linear hyperplane. b) Rawdata X and Y (solid lines indicate the trained linear classifiers). c) Illustration of the hidden output node. The data are here transformed by the $\tanh(\cdot)$ **function and become separable through linear regression (the fit can be calculated from the hidden node output and training weights).**

achieving 'generalization' (i.e., successfully predicting unseen data after training). There are multiple factors to be considered for the training: The choice of the i) nonlinear activation function, ii) loss function (e.g., distance metrics, entropy metrics, etc.), and iii) training algorithm (e.g., error backpropagation, gradient descent, equilibrium propagation, etc.).

Other factors include the availability of training data (and labels[9]), and the specific machine-learning framework. The latter is especially important given the existence of simplified network architectures such as reservoir computing or extreme learning machines [266–268]. Such neural networks are of particular interest as they exploit random but fixed weights in a nonlinear network combined with a straight-forward training method such as linear regression, which can significantly reduce the training complexity and times.

Nowadays, most approaches utilize gradient-based methods such as stochastic gradient descent and error-backpropagation [269]. However, the use of such algorithms requires that i) the used nonlinear functions are differentiable, and ii) access to the individual networks nodes is possible. It is important to note that, while the backpropagation method is very popular, it is not a very efficient algorithm [270]. In today's neural network implementations, the training phase is considered the most inefficient and time-consuming part [73]. This is partially due to the fact that backpropagation is an 'artificial' concept (derived from calculus [239]) that has no biological equivalent [270]. As such, the training inefficiency of neural networks can be partially attributed to the chosen learning concepts and large efforts are undertaken into investigating more efficient training concepts (which are of special interest for physical neural networks) such as equilibrium learning [270, 271] or forward

[9] obtaining labels can be very time consuming and expensive [265]

error computing [272]. Moreover, an important consequence of using the backpropagation algorithm is the requirement for 'access' to nodes in order to calculate the gradient per node. While this is less of an issue for digital implementations, it can be very challenging for physical neural network implementations.

Lastly, an important aspect of training any neural network is the correct choice of training hyper-parameters, such as learning rates η, sample/network sizes, regularization or cross-validation methods [273, 274]. If the hyper-parameters are poorly (or not at all) chosen, the neural network features inadequate performance. Common training flaws include under- and overfitting as illustrated in Fig. 3.4 for a regression problem [275]. In the former case, the neural network is not able to perform meaningful regression of the data. This is often the result of ill-chosen network topologies, insufficient training duration or learning rates. In the latter case, the loss value decreases consistently, however it can be seen that after reaching optimum performance, the test error increases. The reason for this behavior are often too large parameter spaces (i.e., the network size is too big for the problem, thus featuring many free parameters) or the lack of regularization parameters. Consequently, the network fits even the noise in the provided training data (see Fig. 3.4c), which will be detrimental during inference tasks. Remarkably, recent studies show that really large nonlinear networks can avoid overfitting even without regularization [264], yet most practical neural networks operate in a parameter regime (e.g., amount of nodes/layers) where such issues are apparent.

3.2 Photonic Neural Networks: State-of-the-art

The development of photonic neural networks is not a recent innovation and has been of interest for several decades [276, 277]. The primary reason for investigating photonic platforms are promising advantages inherent to optics such as passive transmission, high degree of parallelism (e.g., exploiting the different degrees of freedom of light), intrinsic field operations (e.g., vector multiplications) and intrinsic material nonlinearities [70]. Thus, physical neural networks based on a photonic platform have the potential for highly efficient all-optical brain-inspired signal processing. The successful implementation of such networks (see Fig. 3.5) is of high interest for many application fields, ranging from the most obvious use of (photonic) machine-learning for optical telecommunication [11, 12, 23, 24, 49, 278], to the use towards machine-vision, smart-microscopy, -spectroscopy and -metrology [53, 279–285], next generation robotics [286] and quantum technology [287] or even

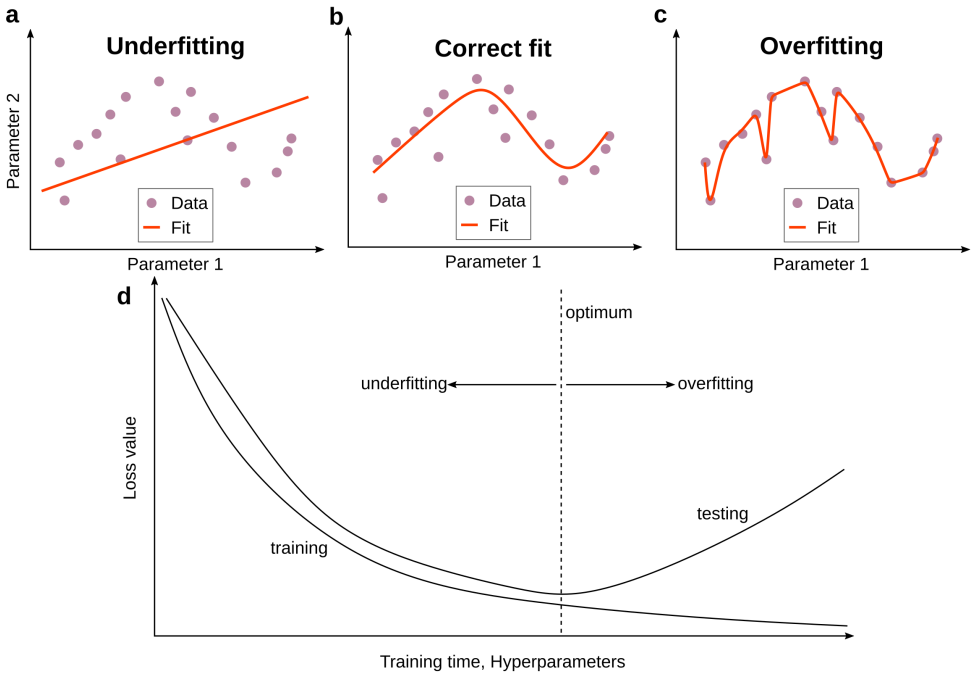


Figure 3.4: Illustration of different training behavior of neural networks.

for solving complex computations [288]. In light of this application potential, many different concepts have been developed over the years, and the main concepts will be briefly explained in the following.

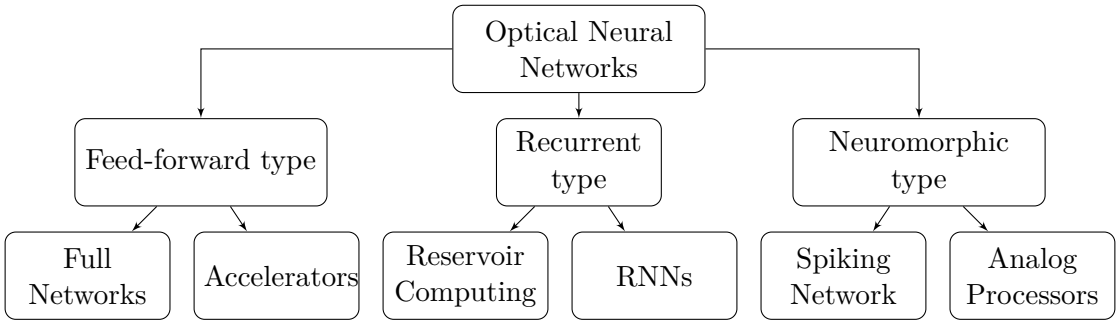


Figure 3.5: Overview over different optical neural network approaches.

Essentially, the implementation of photonic neural networks can be categorized into three main architectures based on their type of information processing: i) feed-forward type networks, ii) recurrent type networks, and iii) neuromorphic networks. The former two network architectures are well known from computer science, while neuromorphic implementations are inspired by biological concepts such as spiking neurons.

As outlined before, feed-forward type networks process information uni-directional (i.e., information passes each node/layer only once), while in recurrent neural networks the information can remain inside the network for a particular time (i.e., akin to memory) and subsequently pass previous nodes [289]. Especially in optics, the latter can be straight-forwardly implemented via cavities (e.g., free-space, fiber or integrated) and was therefore one of the main architectures investigated in the early years due to its simplicity for implementation [277, 290–292]. On the contrary, recent availability of large scale, reconfigurable photonic systems [63, 293–295], inexpensive SLMs, wide-spread availability of high-speed RF components, and rapid prototyping technology (such as photonic 3D printers [52, 296]) have redirected investigations toward the realization of large-scale feed-forward type networks.

Notably, each category features additional subcategories such as different architectures (e.g., deep neural networks [50, 52] or photonic accelerators [59, 62, 63]) and/or abstract concepts (e.g., physical vs. virtual nodes in reservoir computing [291, 297] - a popular sub-type of recurrent neural networks). In general, each of the implementations features unique advantages, for example, in terms of system complexity (incl. footprint), efficiency, scalability or achievable performance[10].

Although the ultimate goal is the implementation of an all-optical neural network, the realization of feed-forward architectures is especially challenging. While some all-optical neural networks have been demonstrated for reservoir computers [45, 47] and diffractive neural networks [52], most work predominantly focuses on the implementation of specific building blocks (either the nonlinear activation functions [298–300] or especially for efficient matrix multiplications [59, 62, 63, 293][11]). As a consequence, the majority of implementations are de-facto of opto-electronic nature [298], thus adding a bottleneck regarding efficiency, processing speed and scalability [301], even though they have shown remarkable results.

Physics Driven Neural Networks

One of the main problems of the previously proposed concepts is the implementation of nonlinearity and/or complex setups (i.e., multi-layer architectures). Most of the aforementioned implementations (especially physical feed-forward) are single-layer solutions which only perform the matrix

[10] in optics, efficiency and performance/speed are often a trade-off due to inefficient electro-optic conversion
[11] a comprehensive review of different implementations can be found in Refs. [16, 76–78]

multiplication optically while the nonlinear transformation is carried out digitally on a computer. This leads to a bottleneck in most of these approaches as OEO (optical-electrical-optical) conversions are inherently inefficient (known as the *"electro-optic bottleneck"*) posing stringent limits on system scalability [301]. Thus, recent research has been increasingly focusing on the implementation of concepts that feature intrinsic nonlinear system behavior. This includes for example the use of 'artificial' nonlinear materials [302, 303], as well as optical nonlinear effects [56, 304]. The advantages of such systems are i) in general more complex dynamics (required for more complex problems), ii) energy efficiency (data transformation occurs passively during propagation), and iii) less system complexity (most designs thus far are either passive and/or single component-based). As a consequence, all-optical networks promise more efficient processing and as such, physics inspired networks gained rapid interest [54, 302]. However, such physics-inspired implementations often lack specific node representations, therefore different training methods are required for either the system design and/or the actual system training of a platform [302, 303, 305].

A particular type of physics-inspired neural networks are wave-based neural networks (*Schrödinger networks* [56, 304, 306]), or other physical wave neural networks [54, 302, 307, 308]). These type of networks break with common conventions as they utilize complex physical wave dynamics for analog neuromorphic processing without 'artificially confining' optical hardware into one of the known network formats (feed-forward/recurrent neural networks). In general, these systems build on input or output training schemes (such as the extreme learning machine framework [54, 268]), which can be both highly time and energy efficient, while providing the computational power of deep neural networks to a certain extent.

Recent experimental realizations are, for example, based on spatial multi-mode approaches in fibers [56] utilizing nonlinear mode-coupling as a key mechanism. This work has shown remarkable benchmark results in a variety of inference tasks. Yet, multi-modes lack, in perspective, the ability of improving compactness, latency, or energy consumption. Other approaches include for example free-space systems [57] which feature only limited nonlinearity (provided by the photodiode saturation gain) or 'black-box' approaches [304]. However, while these early implementations have shown promising results, they are still limited in methodology and understanding of the underlying principles. It remains particularly unclear which role complex nonlinear dynamics in waveguides can play in the implementation of physics-based neuromorphic processors.

3.3 All-Optical Machine Learning with Waves

Disclaimer: The simulations shown in this section were performed by Dr. Mario Chemnitz, Postdoctoral fellow of Prof. Roberto Morandotti (INRS-EMT). The experimental realization of the optical system as shown in this section (setup implementation and characterization, data acquisition and analysis) was performed by the author (B. Fischer).

3.3.1 Soliton Dynamics for Transient Optical Neural Networks

Wave-computing seems to be a promising concept for the implementation of analog optical neural networks. However, it often comes with drawbacks in terms of reconfigurability [56, 57], energy efficiency [56] or experimental realization [302, 309]. Indeed, one of the 'working horses' of nonlinear optics, highly nonlinear fibers/waveguides, have just recently been considered for the specific purpose [304]. While these initial results seem very promising, these implementations work only in very limited regimes and/or treat the system as a 'black-box'. For example, they utilize only very limited spectral bands and/or rely on single nonlinear transformations [304]. Hence they do not utilize the fast range of nonlinear dynamics available to such platforms (e.g., nonlinear spectral broadening via soliton-fission), therefore lacking network scalability in the form of diverse spectral output dynamics. First implementations made use of amplitude encoding [304] in a limited operational regime (optical L-Band), however a deeper understanding of i) the underlying dynamics and adaptation towards more 'generalized' neural networks and ii) performance/influence of phase dependency is still lacking. Other implementations of such fiber- and waveguide-based concepts focus largely on spatio-temporal concepts exploiting multi-mode coupling and dynamics [56, 310], which are either bulky, or difficult to scale and energy inefficient. As such, exploiting the spectro-temporal space is of particular interest as it allows scalability in a compact footprint. Moreover, such processes can be often easily tailored to operate in different output regimes [42, 311]. For example, the system itself does not change if more spectral output bins are considered or changing the input pulse parameters, thus offering an easily scalable degree of freedom.

Soliton dynamics are an especially attractive framework as they offer highly complex nonlinear processes that can be implemented via easily accessible, off-the-shelf components [312, 313], and are also of high interest for telecommunication applications [314]. The strong phase dependency of these dynamics have also been recently investigated towards the implementation of all-optical logic

gates [315] or optical computing circuits [316]. Consequently, solitons have been also proposed as a fundamental tool for neuromorphic computing [54, 317].

In short, a (fundamental) soliton[12] is a light pulse that propagates inside an optical medium without experiencing dispersive broadening of the pulse shape. In general, during propagation in a fiber, any optical pulse would broaden in presence of anomalous dispersion (D>0) of the medium [191]. This occurs because of the finite spectral bandwidth of the pulse, leading to a negatively chirped pulse at the output of the optical medium (i.e., higher frequency components will arrive faster than lower ones). At sufficiently high peak power, the nonlinear Kerr effect causes a positive chirp (i.e., lower frequency components will arrive faster than lower ones) which is able to compensate for the negative chirp of the dispersion at conditions that are well defined (see introduction of soliton number in the next section). Interestingly, balancing dispersive and nonlinear effects (e.g., via optical power or dispersion design) will cancel each other out and lead to a temporally self-regulating pulse, i.e., a soliton, that propagates without any changes to its pulse form (within certain limitations) [313]. While fundamental solitons preserve their shape, higher-order solitons show periodic oscillations. Only after a certain period, i.e,. propagation distance, the initial pulseform is restored, the so-called soliton period [318] (the dynamics of such higher-order solitons will be briefly elaborated in the next section).

Therefore, solitons are a particular interesting choice of computing basis for optical neural networks, as they behave as step-wise attractors, i.e., the peak power and input pulse form decide about the soliton order and its location in the spectro-temporal space [191]. This means that in case of different input pulse parameters, specific spectral outputs can be generated repeatably if the experimental conditions are chosen correctly [319].

Therefore, the understanding and implementation of such soliton-based neuromorphic systems will permit a wide range of energy-efficient applications, particularly towards machine-vision, sensing and spectroscopy tasks where ultra-fast laser sources and HNLFs are already wide-spread.

[12] in the following, soliton refers to a temporal soliton

Background & Theory

The generation of a supercontinuum (SC) through the use of highly nonlinear waveguides (e.g. integrated waveguides or fibers) is among the most complex nonlinear phenomena observed to date in fiber optics. Since the discovery of SC generation in bulk glass [320] in 1970, substantial effort has been devoted to fully understand and model the dynamics of such complex nonlinear phenomenon. Recent advances in fiber optics, especially the invention of photonic crystal fibers(PCFs), have allowed to precisely engineer the optical properties in order to reach optimal performance in terms of, for example, mode-guiding, spectral broadening, or output powers in specific wavelength bands. In detail, the broadening is achieved through SC generation, mediated by effects such as Kerr nonlinearity, Raman scattering, cross- and self-phase modulation (XPM and SPM, respectively), or modulation instabilities (MIs). Due to SPM, (higher order) solitons [313] can be generated, leading to broadening process via soliton fission [312] and ultimately to coherent broadband supercontinuum generation [312]. Essentially, perturbations during propagation, such as dispersion mismatch, lead to the break-up of higher-order solitons which results in the ejection of fundamental (lower-energy) solitons. Higher-order dispersion terms provide phase-matching for these solitons, which results in the generation of non-solitonic radiation (i.e., dispersive waves) on the lower frequency side (blue shifted). During propagation, solitons shift more towards lower frequencies (so-called self-frequency shift, see Fig. 3.6) and as a consequence the generated dispersive waves also appear at shifted frequencies leading to very broadband output spectra. Importantly, the blue shift of the dispersive waves occurs only as long as soliton and dispersive waves overlap in time [321, 322].

Consequently, the generated SC can span over multiple octaves and are crucial to a variety of fields in metrology (e.g. in the stabilization of frequency combs [323]), microscopy [292], telecommunications [65], and ultra-fast sciences [324] (e.g. in the context of pulse compression). Nevertheless, many parameters, both of the material (nonlinear coefficient and second-order dispersion) and the optical pulse (pulse duration and peak power), determine (and ultimately limit) the generation efficiency and properties of SC. In particular, the efficient generation of broad spectral bandwidths is usually dependent on the optical pulse peak power and duration (usually limited to the sub 100 fs regime).

Coherence is an important property for the goal of this specific implementation, as the coherence of supercontinuum describes the pulse-to-pulse stability and is thus a measure of the spectral reproducibility. Since a high coherence implies that the output for a given input remains the same

even under the influence of noise, this provides an important stability requirement for machine learning applications (i.e., the same input will repeatably lead to the same output). The coherence of two waves can be described in form of a normalized cross correlation of the waves with average intensities $\langle I_{1,2} \rangle = \langle |E_{1,2}|^2 \rangle$ as [159]

$$|g_{1,2}| = \left| \frac{\langle E_1 * E_2 \rangle}{\sqrt{I_1 I_2}} \right|. \tag{3.5}$$

In the case of $I_1 = I_2$, the cross correlation function $|g_{1,2}|$ equals the visibility which can be measured through interferometers [159, 325]. Fig. 3.6 illustrates the role and impact of coherence on supercontinuum generation.

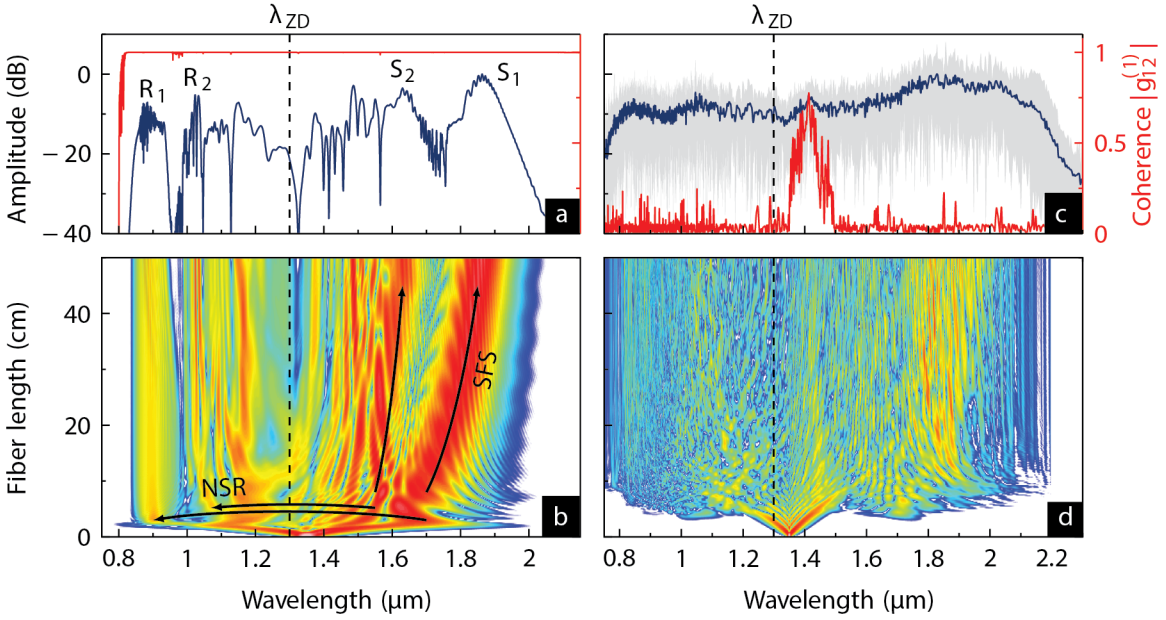


Figure 3.6: Impact of pulse parameters on the coherence of supercontinuum generation. Simulation results showing different effects within SC generation. a) Output spectrum of a coherent SC after 50 cm of propagation in a silica fiber, generated by 50 fs pulses (sech2) with 250 kW peak power (pump wavelength = 1350 nm) and b) spatial evolution of the SC. The initial short pulse leads to soliton formation and soliton self-frequency shift due to stimulated Raman scattering. The fission is accompanied by energy transfer from the soliton to shorter wavelengths via non-solitonic radiation (NSR) responsible for e.g. dispersive waves, indicated as R_1 and R_2 in plot (a). c-d) Simulated output spectrum of the same fiber with input pulses of 300 fs (and other parameters as in (a,b)). Modulation instabilities arising from random phase noise leads to strong incoherent output spectra. NSR = Non-Solitonic Radiation (i.e. dispersive wave), SFS = Self-Frequency Shift i.e. soliton amplified through Raman gain), $R_{1/2}$ = Radiant wave, $S_{1/2}$ = Soliton, λ_{ZD} = Zero-dispersion wavelength. Figure adapted from [326]

In particular, Fig. 3.6(a-b) shows the case of a coherent output spectrum generated through soliton fission (i.e. the splitting of higher order solitons into multiple, fundamental solitons), and

Fig. 3.6(c-d) shows the case of an incoherent spectrum arising from MIs. It can be seen that in the coherent case, the output spectrum shows very few fluctuations between individual single shots and deep fringes in the average spectrum arising from interference effects. Further, the coherent case shows distinct soliton shifts as well as non-soliton radiation such as dispersive waves. On the contrary, the incoherent spectrum is much noisier with the absence of distinct fringes in the average spectrum and furthermore presents a high variance in output power. It is important to note, that instead of measuring the visibility through an interferometer, it can also be estimated from experimental parameters. Indeed, by calculating the dispersive length $L_{dispersive}$, and nonlinear dispersion length $L_{nonlinear}$, the coherence can be estimated through the soliton number [319]. The respective lengths are calculated as follows

$$\begin{aligned} L_{dispersive} &= T_0^2 |\beta_2|^{-1} \\ L_{nonlinear} &= \gamma^{-1} {P_0}^{-1}. \end{aligned} \tag{3.6}$$

Here, T_0 is the $1/e$ pulsewidth, which can be calculated for Gaussian pulses as $\Delta\tau = 2 \cdot T_0 \sqrt{\ln(2)}$ with $\Delta\tau$ being the FWHM width of the input pulse. β_2 is the second-order dispersion parameter. The coherence can thus be estimated through the numbers of generated solitons N as

$$N^2 = \frac{L_{dispersive}}{L_{nonlinear}}. \tag{3.7}$$

Here, a number of solitons smaller than $N = 16$ has been proven an indicator of good coherence [319].

Neuromorphic resemblance of nonlinear pulse propagation in waveguides

The fundamental tool for modeling supercontinuum generation is nonlinear wave propagation based on the generalized nonlinear Schrödinger equation (GNLSE), which allows to accurately model the nonlinear dynamics throughout the propagation in an optical medium [191]:

$$\frac{\partial A}{\partial z} + \frac{1}{2}\left(\alpha(\omega_0) + i\,\frac{\partial \alpha}{\partial \omega}\Big|_{\omega=\omega_0}\frac{\partial}{\partial t}\right)A - i\sum_{n=1}^{\infty}\frac{i^n\beta_n}{n!}\frac{\partial^n A}{\partial t^n} = \\ i\left(\gamma(\omega_0) + i\,\frac{\partial \gamma}{\partial \omega}\Big|_{\omega=\omega_0}\frac{\partial}{\partial t}\right)\left(A(z,t)\int_0^b R(t')\left|A(z,t-t')\right|^2 dt'\right). \tag{3.8}$$

Here, $A(z,t)$ describes the pulse envelope of the electric field under the slowly varying envelope approximation, α and ω the coefficients of attenuation and nonlinearity, respectively, β is the dispersion coefficient, ω_0 the central angular frequency, $i^2 = -1$ the imaginary unit, and $R(t)$ the nonlinear response function describing the contribution of Raman nonlinearity.
This nonlinear partial differential equation has no analytical solution and can only be solved through numerical approaches. A reliable and widely used approach for solving nonlinear propagation of intense optical pulses is the split-step Fourier algorithm [189, 319], as illustrated in Fig. 3.7. In a nutshell, the evolution of a pulse along the waveguide is split into a alternating sequence of linear and nonlinear steps[13]:

$$\frac{\partial}{\partial z}A(z,t) = \{\mathcal{D} + \mathcal{N}(z,t)\}A(z,t), \tag{3.9}$$

with the propagation coordinate z, the optical field envelope $A(z,t)$, the linear dispersion operator $\mathcal{D}$, and the time and space dependent nonlinear operator $\mathcal{N}(z,t)$ [191].

For a significant small step-size Δz, the linear and nonlinear propagation steps can be assumed independent of each other and as such allows to solve each step separately. The linear term (i.e., dispersive step) features an exact solution in the Fourier-domain, while the nonlinear step requires, in general, numerical integration in the time domain in order to obtain a solution [191]. The accuracy of the split-step Fourier method depends on the chosen step-size Δz and can be further improved by other methods such as symmetrized SSFM (i.e., half dispersive step $\Delta z/2$, followed by the nonlinear step and another half dispersive step) [189, 191] or the inclusion of adaptive (i.e., non-uniform) step-sizes [327].

Strikingly, the well established split-step Fourier approach for modeling pulse propagation in optical fibers [191] resembles the basic functionalities of a digital neural network, i.e., a sequence of linear and nonlinear transformations as illustrated in Fig. 3.8!

[13] here, $\{\dots\}A(z,t)$ denotes the application of operator $\{\dots\}$ on $A(z,t)$

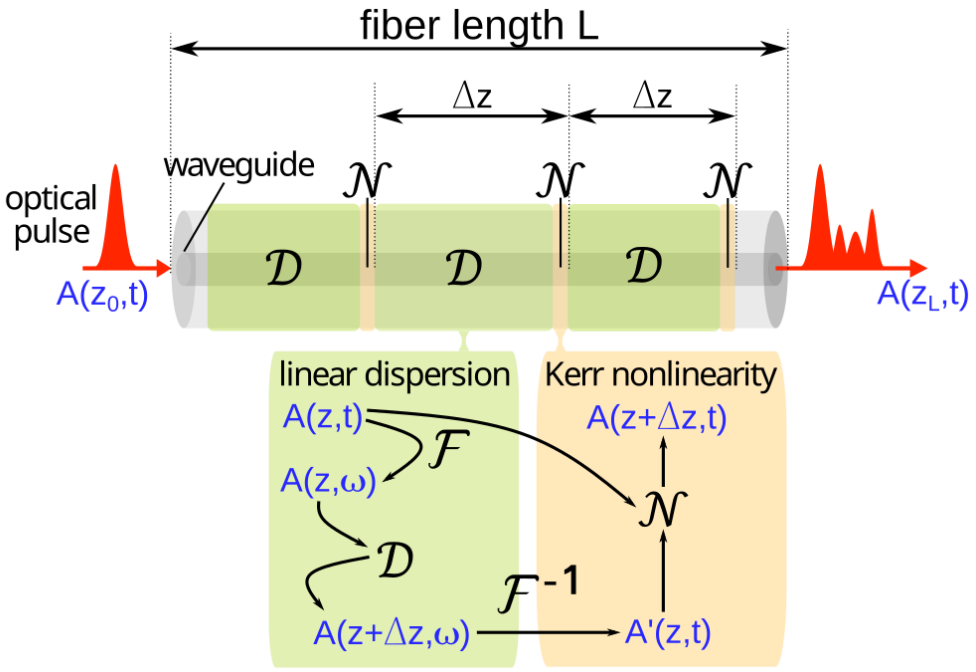


Figure 3.7: Illustration of the split-step Fourier method. A fiber of length L is divided into steps of size Δz, and the linear (dispersive) and nonlinear (Kerr) part of the pulse propagation are solved individually in the frequency (ω) and time-domain (t), respectively. $\mathcal{D}$, $\mathcal{N}$, $\mathcal{F}$, and $\mathcal{F}^{-1}$ represent dispersive operation, nonlinear operation, Fourier and inverse Fourier transform, respectively. A(z_0,t), A(z_L,t) denotes the input and output pulse, A(z,t) and A(z+Δz,t) the pulse at the beginning and end of one step Δz.

Indeed, recent studies have explored this analogy in large-scale telecommunication frameworks [307], as well as in numerical studies towards the implementation of optical neural networks [309], yet a experimental demonstration and insights into potential operational regimes is still lacking.

In the GNLSE framework, the linear steps can be described by the linear dispersion and Fourier operator (corresponding to a summation in the frequency domain) and the nonlinear step is described by the nonlinear operator (a cubic power of the field in the time-domain). Evidently, the Fourier transformed (discretized with N grid points) linear solution ($\mathcal{N} = 0$) shares visible analogy to a sum and weight operation:

$$\begin{aligned} A(z+\Delta z,t) &= \mathcal{FT}\{A(z+\Delta z,\Delta\omega)e^{-\alpha/2-i\sum\beta/n!}\} \\ &= \frac{1}{\sqrt{N}}\sum(A(z+\Delta z,t)e^{-\alpha/2-i\sum\beta/n!})e^{i\omega t}, \end{aligned} \tag{3.10}$$

where the evolution of the spectral field evolution (determined through the local refractive index) takes the role of network weights [309]. On the other hand, the nonlinear solution ($\mathcal{D} = 0$)

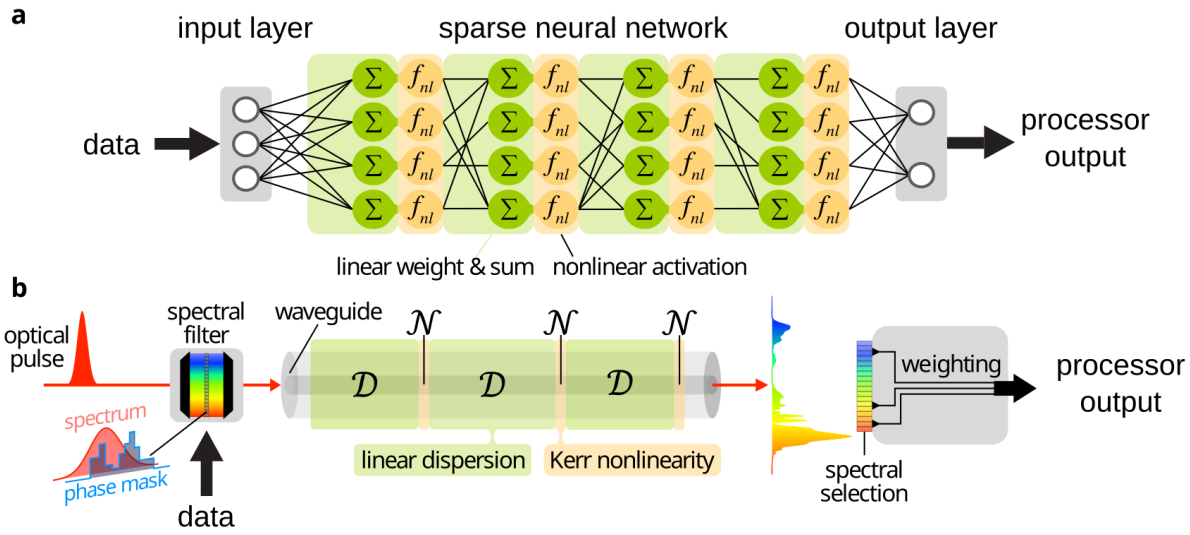


Figure 3.8: Illustration of a) a sparse digital neural network, and b) a analog neural network based on nonlinear waveguide propagation. Information can be encoded onto the spectral phase of an ultrashort optical pulse and subsequently undergoes a series of linear and nonlinear transformations during propagation. At the fiber output, the initial pulse is transformed into a broadband spectrum via nonlinear effects and consequently regions of interest can be identified and evaluated.

$$A(z+\Delta z,t) = A'(z,t)e^{-i\gamma|A(z,t)|^2\Delta z}, \tag{3.11}$$

can be seen as cubic nonlinear self-activation with temporal field weights A, thus representing the instantaneous nonlinear activation after the summation. In this picture, the network weights correspond to the dispersive propagation constant beta, the nonlinear activation threshold to the nonlinear coefficient, and the bias to the power distribution of the incoming field (before being masked or encoded with data).

Contrary to classical neural networks, a fiber-based optical neural network active updating of the network weights is challenging yet not impossible to implement. For example, the refractive index distribution of a fiber can be designed via simulations to act as weights to perform specific neural network tasks [309]. However, even without such *in silico* training the parameters of off-the-shelf fibers (dispersion, nonlinear coefficient, etc.) are imposing an intrinsic random weighting for the dispersive (i.e. linear) step, which enables the use of training paradigms such as extreme learning machines [268] (see following section for details).

Within this picture, a virtual neural network may be constructed based on transient network nodes projected in a two-dimensional spectro-temporal space. An optical field input to the net-

work can be represented as a selection of nodes in time and frequency carrying non-zero node weights, while waveguide dispersion and nonlinearity will define the optical field evolution in time and frequency, respectively.

3.3.2 Experimental Implementation

The concept and experimental setup to verify the neuromorphic computing capabilities of soliton fission is illustrated in Figure 3.9.

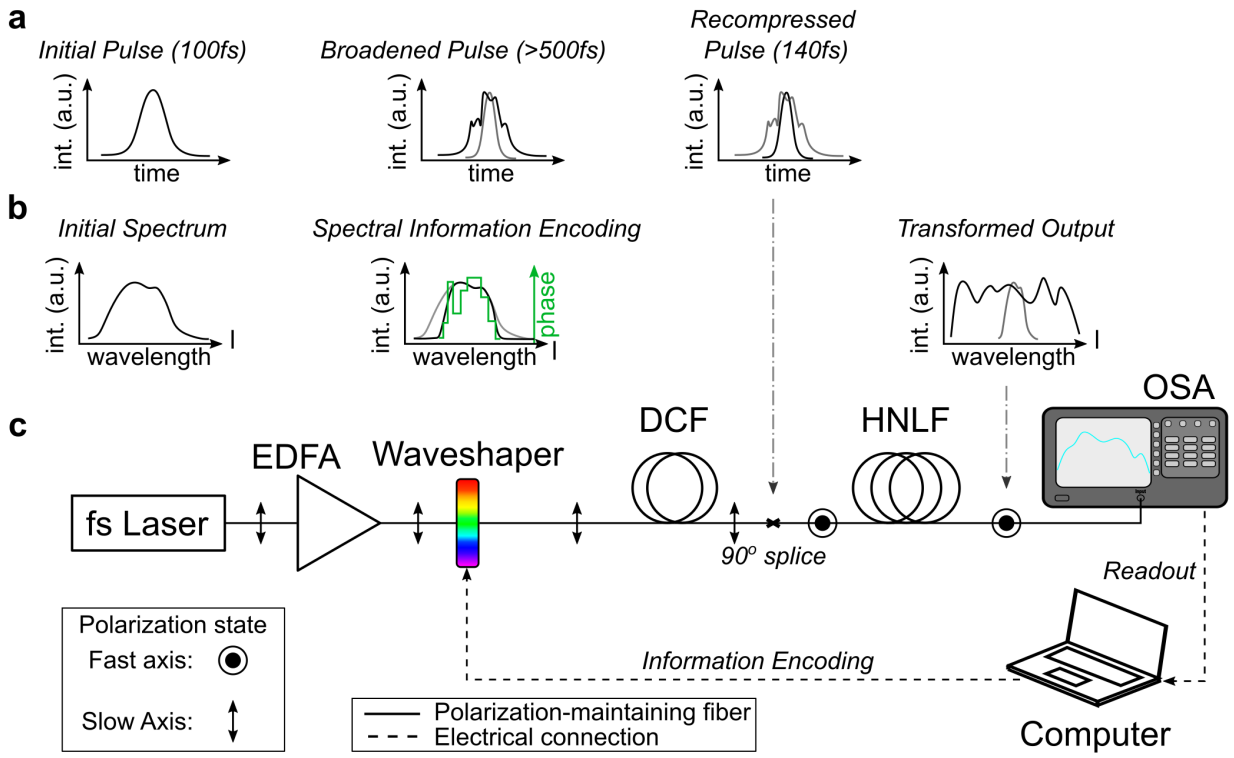


Figure 3.9: Concept and experimental setup of the supercontinuum-based neural network. a,b) Temporal and spectral representation of the pulse at different stages (grey pulse forms represent the signal of the previous stage). c) Experimental setup. EDFA = erbium-doped fiber amplifier, DCF = dispersion compensating fiber, HNLF = highly nonlinear fiber, OSA = optical spectrum analyzer.

The setup consists of

1. A femtosecond laser source (Menlo FC1500-250-WG)

2. An EDFA for loss compensation (Pritel PM-SPFA-23)

3. A programmable filter that allows spectral phase and amplitude manipulation (Finisar Waveshaper 1000X)

4. A dispersion compensating fiber (DCF) for pulse compression (Thorlabs PMDCF, Dispersion $D = -100\,\mathrm{ps}/(\mathrm{nm}\cdot\mathrm{km})$)

5. The HNLF as the *optical processor* (OFS Fitel PM-HNLF, see Table 3.1 for fiber parameters)

All components in the setup are realized as polarization-maintaining fibers. The data is collected using a standard optical spectrum analyzer (Ando AQ6317B) and readout via a computer, which is also used for updating the training data on the Waveshaper. Notably, compared to spatial light modulator (SLM) approaches, the amounts of encoding bins are limited when a Waveshaper is used. However, the latter are fiber-coupled, off-the-shelf components, which have proven long-term reliability and stability in the telecommunications sector. For the spectral readout, different configurations (spectral resolution, sampling points) were used depending on the tasks.

The 90° splice was implemented after the Waveshaper in order to switch from slow axis (the laser emits on the slow axis) to fast axis (the Waveshaper blocks the fast axis). The reason for this is the better overlap to simulations (see Appendix B.1), which allows for additional comparison to theoretical investigations of the system (not shown). The 90° splice was fabricated using a commercial fusion splicer (Fujikura PM100), to align the slow axis after the DCF with the fast axis of the HNLF[14]. Since the connector keys of all pigtails in the setup are aligned on the slow axis, this permits to switch the polarization axes.

Parameter	**Slow Axis**	**Fast Axis**	**Unit**
Fiber length	100		[m]
Effective area	12.57		[μm^2]
Total Loss	0.69		[dB]
Nonlinear Coefficient γ	11.3		[$W^{-1}\cdot km^{-1}$]
Dispersion D	2.35237	1.4158	[ps/nm/km]
Dispersion Slope S	0.0232	0.0211	[$ps/nm^2/km$]
Zero Dispersion Wavelength	1437.0	1478.2	[nm]
Dispersion Coefficient β_2	-2.9983	-1.8045	[ps^2/km]
Dispersion Coefficient β_3	4.266e-8	3.7326e-8	[ps^3/km]

Table 3.1: Fiber parameters for the PM-HNLF. All values for 1550 nm.

[14] the use of polarization beam combiners was discarded as they generally feature bandwidth constraints

Smart System Optimization

The soliton fission regime features stringent requirements on the pulse duration in order to operate in a coherent (i.e., repeatable) regime. Hence the dispersion of the processing system has to be carefully engineered in order to maintain a pulse width below ≈150 fs, which is a challenge on its own in a meter-long fiber system. For this reason, DCFs were employed in the setup so as to reduce the pulse duration down to the transform-limit given by the Waveshaper bandwidth. However, since the setup involves PM fibers from different manufacturers (leading to different higher-order dispersion coefficients) and unknown internal fiber lengths (e.g., inside the laser or EDFA) an exact compensation can not be achieved by DCFs alone. As illustrated in Fig. 3.10a, the compensation with a DCF leads only to ≈375 fs pulse length (FWHM). Although this is already a good compression after more than 20 m of fibers and an initial pulse width > 2 ps, it is still not sufficient for coherent soliton fission dynamics [312, 319, 328]. Therefore, in order to compress the pulse further (to <150 fs) an active pulse compensation scheme was implemented. The active phase control was realized by means of a Waveshaper [329] which allows to imprint an arbitrary phase profile to compensate for the higher-order dispersion terms.

The custom phasemask was implemented using a 5^{th} order polynomial, and the polynomial coefficients were optimized using a particle swarm optimization, akin to the one used in Chapter 2, to reach shortest pulse duration on an autocorrelator, similar to earlier approaches with GAs [36, 97]. The insights gained from using a PSO for optical pulse shaping in Chapter 2, where reused here in combination with the benefit of employing a commercial pulse shaper (i.e., the Waveshaper). The used hyper-parameters for the PSO are summarized in Table 3.2.

Parameter	swarmsize	max. iterations	self adjust. factor	social adjust. factor	max. stall iterations	function tolerance
Value	30	20	1.49	1.1	5	1e-6

Table 3.2: Overview of the PSO parameters used for the smart pulse compression.

The readout of the pulse duration was performed by means of a commercial autocorrelator (Femtochrome FR-103XL) and oscilloscope (Tektronix TDS5052B). The measured pulse duration was evaluated using a two-term Gaussian fit. Equation 3.12 shows the used loss function:

$$Loss = \left(\frac{\mathrm{FWHM} - 130}{130}\right)^4 \cdot \sqrt{\mathrm{FWHM}} + \left(\frac{\mathrm{FWHM} - 130}{130}\right) \cdot \mathrm{Width}_{1/e^2}. \tag{3.12}$$

Here, FWHM is the returned full-width at half maximum of the fit, and Width_{1/e^2} is the $1/e^2$ width of the fit. The FWHM has a larger weight in the loss function ($(\dots)^4$) and a penalty is added for the background in form of the $1/e^2$ pulse width.

As a result, a pulse compression from 375 fs to ≈140 fs was achieved, as shown in Fig. 3.10b. The phase profile obtained via the smart optimization routine is shown in Fig. 3.10c.

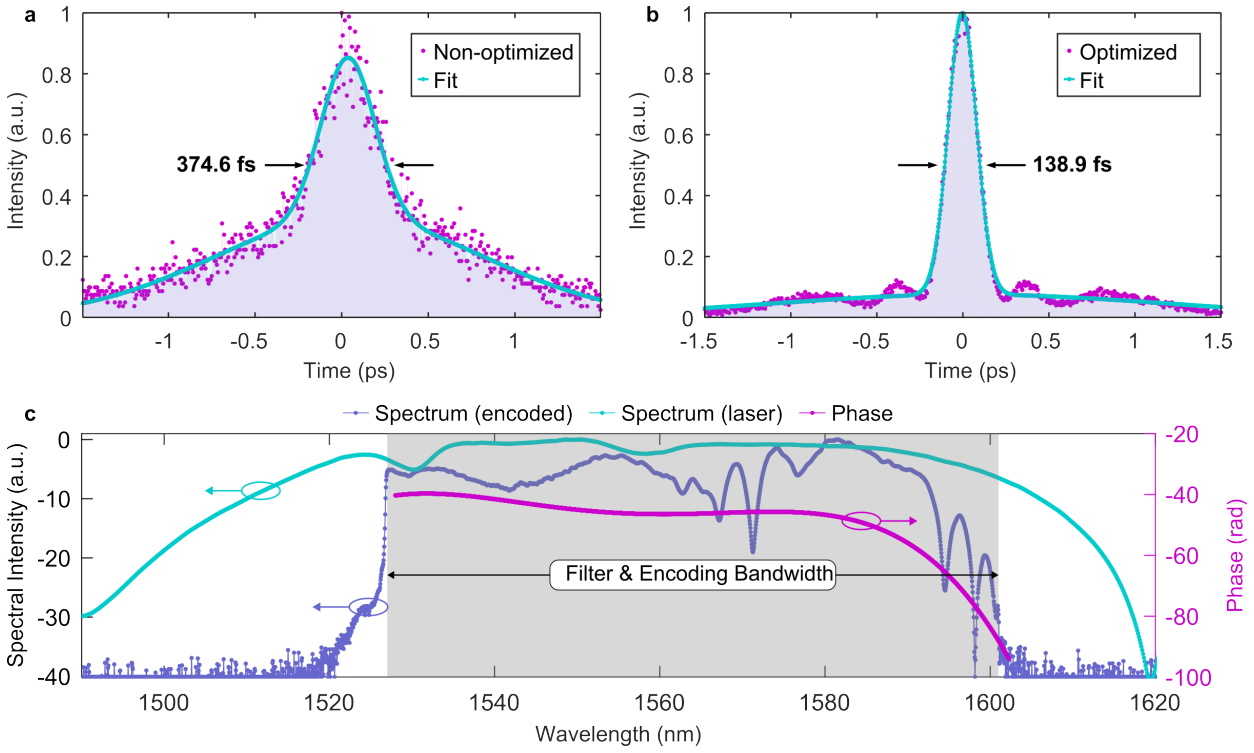


Figure 3.10: Spectro-temporal system characterization. a) Initial pulse duration (before optimization), b) Final pulse duration (after optimization). In both cases, an autocorrelation correction factor of 0.707 was applied for retrieving the full-width at half maximum. c) Spectrum of the frequency comb (cyan), truncated by the programmable filter outside of the grey bandwidth (blue) and the optimized phase profile for dispersion compensation (magenta). The available encoding bandwidth is highlighted by the grey area. For all tasks, the encoding bandwidth was limited to the optical C-band.

In the present implementation, the spectral position for the information encoding can be arbitrarily chosen within the bandwidth of the Waveshaper (≈1528 nm-1601 nm) with up to 400 individual bins. Additional amplitude encoding would be possible in a range of $[0, 35]$ dB, while the employed phase encoding can occur with phases values between 0 and 2π.

System Sensitivity

An important aspect of the concept is the systems' sensitivity to the optical phase of the input pulse (i.e., the information carrier of the developed platform). In order to investigate this, a straightforward "bit-shift" experiment was implemented. For this, a single bit encoding with a 1.0 nm bandwidth (centered at 1553 nm, approx. the center of the pump) was used. Subsequently, the amplitude of the single bit was linearly varied in the range $[0.1\pi, 1\pi]$, and the corresponding spectra were measured. Fig. 3.11 illustrates the phase sensitivity of the system for increasing phase values.

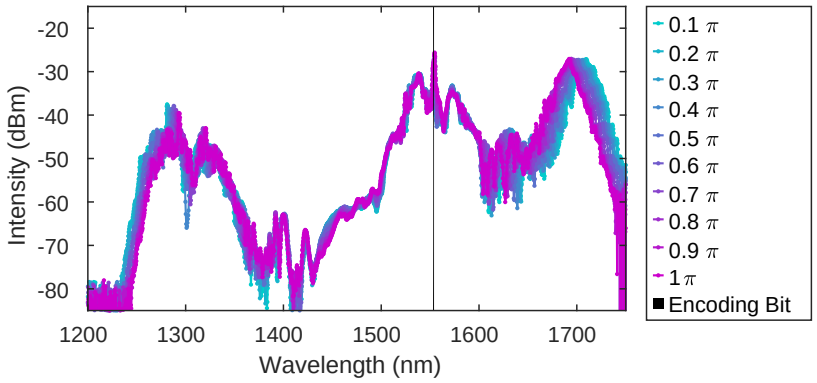


Figure 3.11: System dynamics for a single, static encoding bit. The phase value of a single bit (centered at 1553 nm, indicated by the black bar) was varied from 0.1π to 1π, the color plots show the corresponding spectra.

Interestingly, there is only limited effects from the phase encoding observable in the pump region (between 1500 nm and 1600 nm), which recent concepts put as the operation region [304]. On the contrary, the solitonic ($>$ 1600 nm) and non-solitonic radiation regime ($<$ 1400 nm) feature significant spectral variation for different input phase values. This strong phase dependency of the spectral output is a fundamental aspect of this implementation. Especially, the nonlinear 'mapping' (i.e., phase input at the pump affects the spectrum far outside the pump region) is crucial for the discrimination of highly non-linear separable problems.

To further investigate the phase impact, an additional wavelength sweep of the single bit was performed. For this, the bit was iteratively moved through the available bandwidth of the Waveshaper ($\approx$1528 nm to $\approx$1600 nm) with a step-size of 2.0 nm. For each individual bit position, the output spectrum was recorded. The sweep was repeated for different phase values in the range $[0.1\pi, 1\pi]$, and the results for four different phase values are shown in Fig. 3.12.

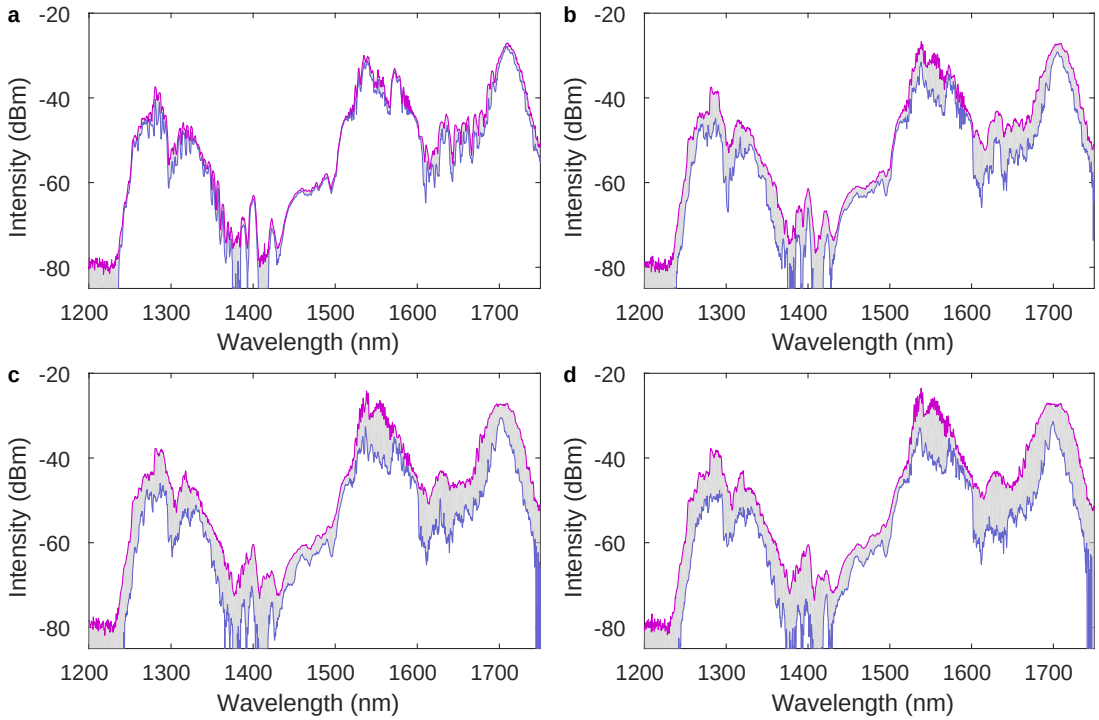


Figure 3.12: System dynamics based on the input phase value and bit positions. a) 0.1π phase. b) 0.4π phase. c) 0.7π phase. d) 1.0π phase. The gray area represents the individual spectra for all bit positions, and the colored spectra (blue and magenta) illustrate the minimum and maximum of each phase configuration.

From these results, it is evident that for relatively small phase values ($\approx 0.1\pi$), no significant spectral variation occurs, regardless of the input bit position. Once the phase value increases, a higher variance of the spectral output can be observed indicating a better operational regime that potentially allows for spectral distinction of different inputs. However, it is important to note that the optical pulse experiences more disruptions (in the temporal domain) as the phase increases, which in turn would ultimately decrease again the systems' coherence and hence accuracy. Fig. 3.13 shows autocorrelation traces for encoded input pulses with different phase values. For this measurement, a random digit from the MNIST data set (down-sampled to 1×100, see following section for details) was encoded with two different phase values (π and $\pi/8$) and compared to the reference case of no active encoding.

As the autocorrelation traces indicate, for relatively small phase values (here $\pi/8$) only small changes in the optical pulse are observable (FWHM broadens slightly from $\approx 140\,\text{fs}$ to $\approx 155\,\text{fs}$). On the other hand, for larger phase values (here π) the optical pulse is strongly affected, leading to a significantly increased FWHM (from $\approx 155\,\text{fs}$ to $\approx 250\,\text{fs}$). Moreover, a pulse splitting of

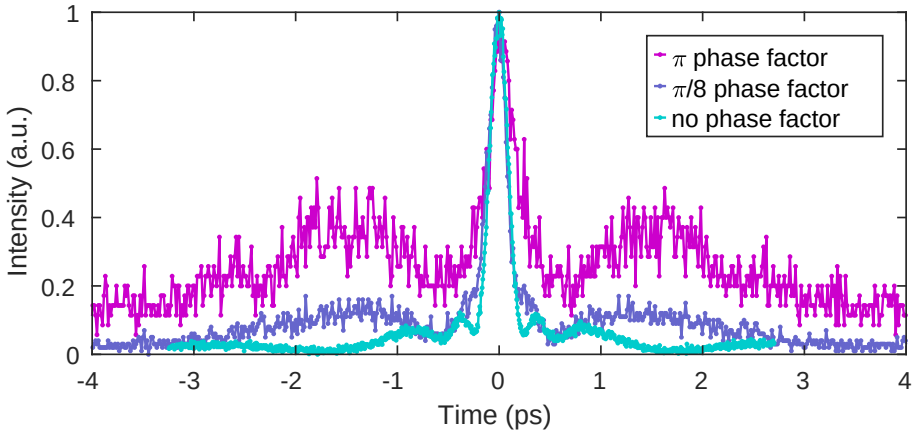


Figure 3.13: Autocorrelation measurement for different encoding phase values. Blue and magenta depict $\pi/8$ and π phase factors, respectively. The cyan curve shows the pulse without encoding for reference.

$\approx$1.5 ps is observable (a double pulse results in 'side' peaks symmetric to the main pulse in the autocorrelation).

Network Training

As outlined at the beginning of this chapter, the training of neural networks is commonly performed by error backpropagation algorithms [215, 269], and consequently many of the physical counterparts follow a similar strategy [16, 77]. However, implementing error backpropagation methods in physical platforms is very challenging as the gradients can not be straightforwardly measured [330–332]. As such, these networks are usually pre-trained through digital models of the specific chip on a regular computer (where a gradient can be easily evaluated) and subsequently transferred to the physical platform. While this method has been widely adopted for different platforms [52, 293, 294], it comes with various drawbacks. Chapter 2 already discussed, how the transfer of digital models to physical realizations generally demonstrate inferior performance. As such, digital training of a network and subsequent transfer to a physical platform imposes stringent requirements on the fabrication accuracy and parameter knowledge to avoid any deviance between theory and experiment, which would require additional counter-measures adding even more complexity [153–155]. Moreover, the network training via simulation adds significant electronic overhead to the platform negating potential energy efficiency aspects during the training phase, which is usually considered to be the most power-hungry [73]. Thus, only the inference phase after training fully benefits from the advantages of a physical neural network implementation. Proposed solutions to

overcome the *in silico* training issue include for example adjoint methods [330], optical cameras monitoring backpropagating light [332], hybrid approaches considering full system simulations [308, 333] or evolutionary algorithms [95, 334, 335] for on-chip mesh structures, as well as adaptive optics for diffractive networks [305].

Interestingly, concepts such as extreme learning machines (ELMs) do not require any elaborate digital simulations for training but rather make use of straight-forward linear regression methods [268]. ELMs follow an analogous principle as reservoir computing methods [336, 337], where a highly complex reservoir (or network) of randomly interconnected nodes is exploited for the data transformation, while the training is solely performed by linear regression at the output layer. The main distinction of ELM implementations to RC is the absence of any intrinsic network memory. Both, RC and ELM, commonly utilize straight-forward algorithms such as linear regression [338] or its regularized version - the so-called Ridge regression (also known as Tikhonov regularization) [273]. The latter introduces a regularization parameter λ in order to avoid over-fitting[15] (i.e. fitting noise behavior rather than information),

$$\mathbf{W}_{out} = \mathbf{Y}_{Target}\mathbf{X}(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I})^{-1}. \tag{3.13}$$

Here, $\mathbf{Y}_{Target}$ is a $k \times n$ matrix containing the targeted output for each time step, $\mathbf{X}$ is a $m \times n$ matrix containing the system output (i.e., the spectral components of the selected readout bins), and $\mathbf{I}$ is the identity matrix.

Therefore, the ELM framework is adopted for training (i.e., linear regression with $\lambda = 0$ to determine the output layer weight matrix) in the proposed implementation of an optical wave-computer. The linear regression is combined with a straight-forward readout-bin search over the recorded spectra (dubbed *equal-search* what follows) to determine the best-performing bin positions of the broadband output. Fig. 3.14 illustrates the equal-search algorithm.

[15] for $\lambda = 0$, the equation represents ordinary least squares

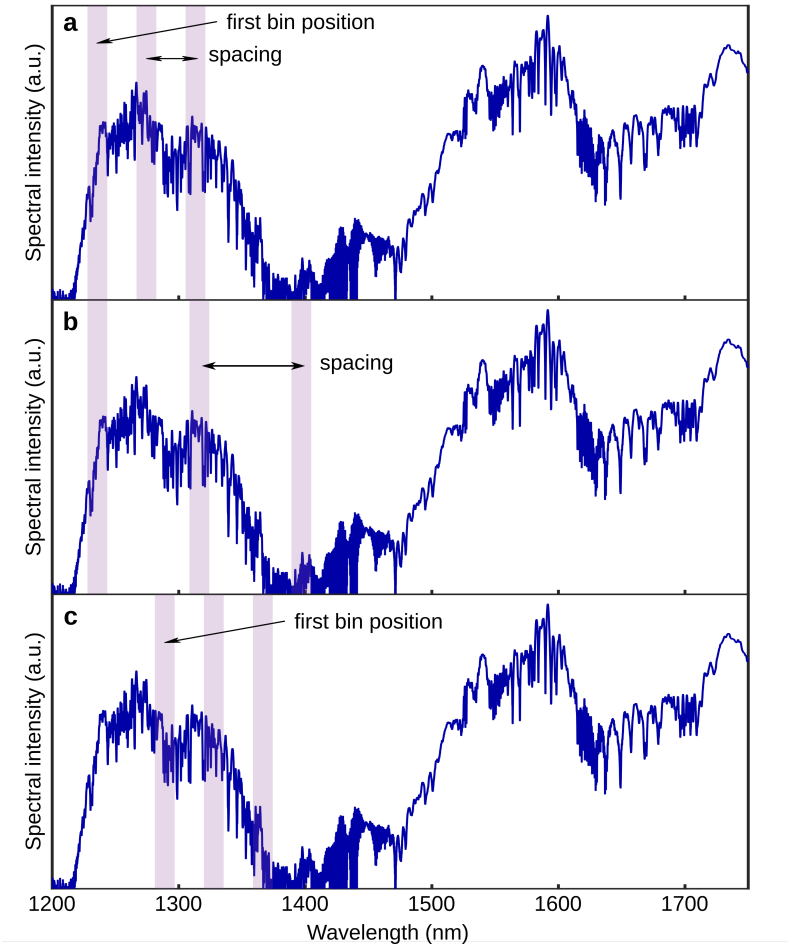


Figure 3.14: Illustration of the equal search algorithm for three bins. a) Illustration of equally distributed bins with a given first bin position and a given spacing. b) Variation of the bin spacing (fixed first bin position), c) Variation of the first bin position (fixed bin spacing).

Output Data Selection

For a given amount of readout bins (determined by e.g., the amount of classes), equidistant bins are selected based on a defined starting wavelength (first bin position) and a defined bin-spacing. Both parameters are iteratively swept over the recorded training spectra (see Fig. 3.14bc). For each combination, the implemented loss function (here mean square error - MSE) is evaluated using an n-fold cross-validation [274]. Such algorithm divides the training set into n equal batches from which n-1 batches are used for training and the remaining one is used for testing. Subsequently, the batches are iterated and re-evaluated for all combinations, as illustrated in Fig. 3.15.

Finally, the best performing (in terms of MSE) bin combination is selected. The unseen test set is then applied to this readout configuration, i.e., $\mathbf{Y}_{Predicted} = \mathbf{W}_{out}\mathbf{X}_{Test}$. For the performed

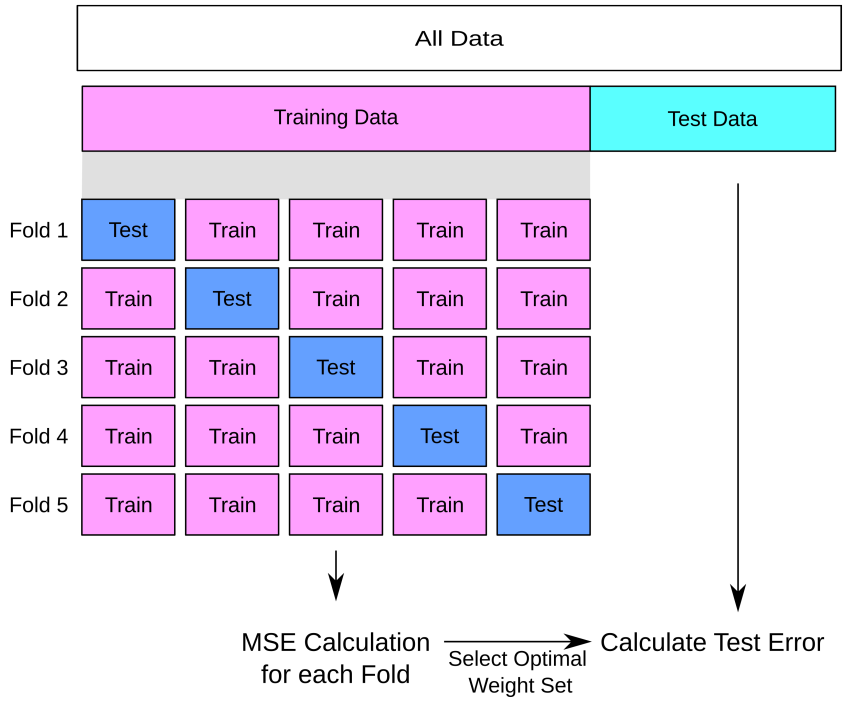


Figure 3.15: Illustration of the n-fold cross-validation for $n = 5$ folds. All measured data are first divided into a training and test set. Subsequently, the training set is split into n equal batches, and n-1 folds are used for training, while the remaining fold is used for testing. Finally, based on the obtained MSE, the best performing parameters (such as bin combination, weights, etc) can be selected.

classification tasks, a one-hot (or 1-of-K [232]) encoding was used as illustrated in Figure 3.16. The individual columns correspond to each class and for the retrieval of a single prediction the '*winner takes it all*' criterion (i.e., $argmax(\mathbf{Y}_{Predicted})$) was used to determine the column index featuring the highest value and thus the winning class.

Index 1	1
Index 2	6
Index 3	2
Index 4	5
Index 5	4
Index 6	7
Index 7	8
Index 8	5

(a) Original data set.

Index 1	1	0	0	0	0	0	0	0
Index 2	0	0	0	0	0	1	0	0
Index 3	0	1	0	0	0	0	0	0
Index 4	0	0	0	0	1	0	0	0
Index 5	0	0	0	1	0	0	0	0
Index 6	0	0	0	0	0	0	1	0
Index 7	0	0	0	0	0	0	0	1
Index 8	0	0	0	0	1	0	0	0

(b) One-hot converted set.

Figure 3.16: Illustration of one-hot encoding: a) Original data in a one-dimensional array of length n containing k individual elements. b) One-hot encoded data in a $n \times k$ matrix.

It is important to note that the output bins found are highly dependent on the sample used. Since fabricated fibers can have significant variations in their parameters along the fiber length [339, 340] (such as geometry, dispersion or nonlinearity), the propagation dynamics vary between different

samples, thus requiring individual training. However, to reduce training times for more complex tasks, approaches such as transfer-learning and online training may be indispensable [91, 341, 342].

3.3.3 Experimental Computing Results

Benchmarking Tasks

Over the years, a plethora of benchmark tests for the evaluation of neural networks have been made publicly available. In order to evaluate the developed system, and allow a comparison to other optical implementations, a selected number of sets were used. In detail, the benchmark is split into two different kind of sets. Firstly, there are more fundamental tests ('task-independent benchmarks') which can be performed without any prepared data library, such as the n-Bit parity test [343, 344] or the universal function approximation [345]. Both of these tests allow to investigate fundamental aspects and potential scalability of the developed platform. Subsequently, more application-forward tests ('task-dependent benchmarks') are used to evaluate the platform for specific type of tasks such as categorization based on numerical inputs or image classification. As such, three different tasks commonly used in optical machine-learning were tested[16]: the IRIS flower recognition [346], the WINE dataset [347], the Abalone dataset [348], as well as the popular MNIST handwritten digit recognition [349]. Notably, these sets are of varying complexity and structure, therefore allow also a more rigorous investigation into the specific performances.
The theoretical inference speed of the system is given by the fiber length, see Table 3.1, and corresponds to approx 230 ns. However, due to experimental constraints, i.e., the update rate of the spectral encoder (approx. 0.5 s) and the data acquisition (tens of seconds), the experimental inference speed is significantly reduced at this point (see Chapter 4 for potential solutions).

Data Encoding and Training Hyper-parameters

Contrary to recent implementations [304], the information encoding of the demonstrated platform is based on spectral-phase encoding and operation in the soliton fission regime. Notably, this configuration avoids any loss in power due to the encoding scheme. The information (i.e., features)

[16] while original references are provided, the datasets were be obtained from the University of California Irving Machine Learning Repository (IRIS, WINE, Abalone) at `https://archive.ics.uci.edu/ml/index.php`, and from kaggle (MNIST) at `https://www.kaggle.com/datasets/hojjatk/mnist-dataset`

are converted into spectral phase values and subsequently imprinted onto the pulse spectrum using a programmable filter (Finisar Waveshaper 1000X). The programmable filter features a transmission bandwidth covering the optical C- and L-Band, which avoids spectral truncation, and hence pulse broadening.

For all results presented in the sections that follow, the entirety of the available transmission bandwidth of the Waveshaper was used (1528 nm-1601 nm), while the phase encoding bandwidth was limited to the C-band (approx. 1528 nm-1568 nm). For this, the input features were encoded into equally sized wavelength bins. In all cases, the individual features were first normalized (division by the ensemble maximum for each feature), multiplied with a fixed phase factor of $\pi/8$ (determined through the bit-shift and preliminary measurements) and finally all data samples were measured. Afterwards for the offline training, the data were first randomized and finally partitioned into training and test sets.

General training parameters for the results presented in this section are summarized in Table 3.3, and the following subsections show the training results for the different tasks in more detail. All experiments were performed for a soliton number of 5, corresponding to a pulse energy of 90pJ, unless otherwise stated.

Dataset	**Training Samples**	**Test Samples**	**Trainings Ratio**	**Cross-Vals.**	**Input Features**	**Readout Bins**
Sinc	500	500	0.50	5	100	100
n-Bit Parity	200	300	0.40	5	varying	varying
Iris	105	45	0.70	5	4	4
Wine	124	54	0.70	4	13	18
Abalone	1500	500	0.75	5	8	31
MNIST	1200	300	0.80	5	100	94
COVID-19	642	214	0.75	0	30	119

Table 3.3: Training hyper-parameters for the different benchmark tasks.

Task-independent Benchmarks

n-Bit Parity Problem

The n-bit parity problem, i.e. the generalization of the XOR operation (see equation 3.14), becomes increasingly hard to solve in neural networks for increasing bit lengths n [343, 344]. Notably, the XOR task is not solvable with a single perceptron [232]. Generally, the accuracy of solving the

n-bit parity test using neural networks correlates directly with the amount of hidden-layer neurons in the network [343, 344]. It has been found that the number or neurons can be reduced using more sophisticated nonlinearities than the common sigmoid or *tanh(·)* [259]. *Thus, the n-bit parity problem can be seen as a measure of the nonlinear separation capabilities of a neuromorphic processor and provides a straight-forward way of scaling the problem complexity.*
The n-Bit parity problem for a vector $\vec{B}$ with elements b_k containing n uniformly distributed pseudorandom integers from the sample interval [0, 1] can be defined as [350]

$$\vec{B} = \begin{cases} 0 & \text{if } \sum_{k=1}^{n} b_k \text{ is even} \\ 1 & \text{if } \sum_{k=1}^{n} b_k \text{ is odd.} \end{cases} \tag{3.14}$$

Several paramteric iterations of the n-bit parity problem were performed. The performance was investigated as a function of the bit length n, the optical input power (in terms of soliton number N, see eq. 3.7), and number of readout bins m. The results are shown in Figs. 3.17 and 3.18. In detail, for the results in Fig. 3.17, the system was trained for different input power levels (expressed via the soliton number N), and a readout configuration corresponding to the input bit length (i.e., *bit length n = n readout bins*).

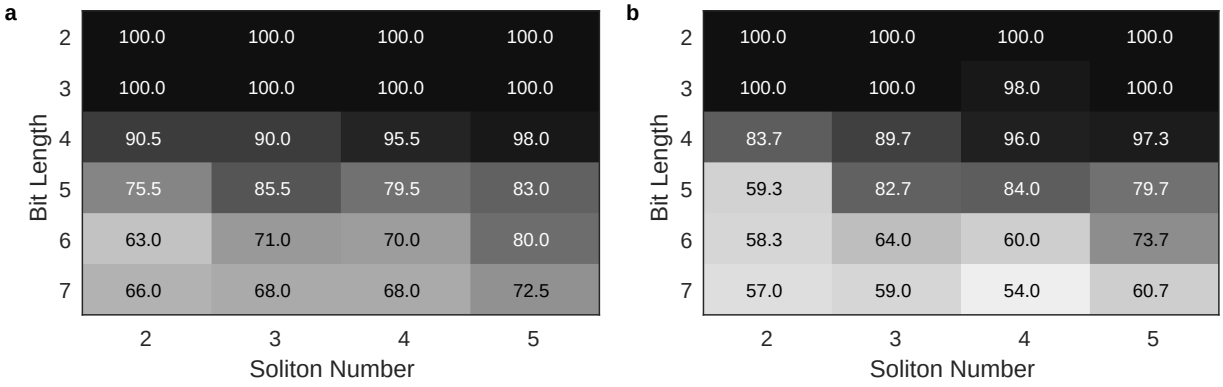


Figure 3.17: Results for the n-bit parity test for different input power levels. a) Training results. b) Test results.

It is evident from Fig. 3.17, that for a given, constant system setting (i.e., input power and readout configuration), the performance deteriorates for increasing bit lengths as expected from prior work [343]. Interestingly, by only increasing the input power of the system, which is equivalent to an increase in system nonlinearity, the performance can be improved.

This behavior is further investigated by increasing the number of readout bins from the initial *readout bin = bit length* up to 125 readout bins for two different input power configurations (low nonlinearity N=2 and high nonlinearity N=5). The results for this parametric study are shown in Fig. 3.18.

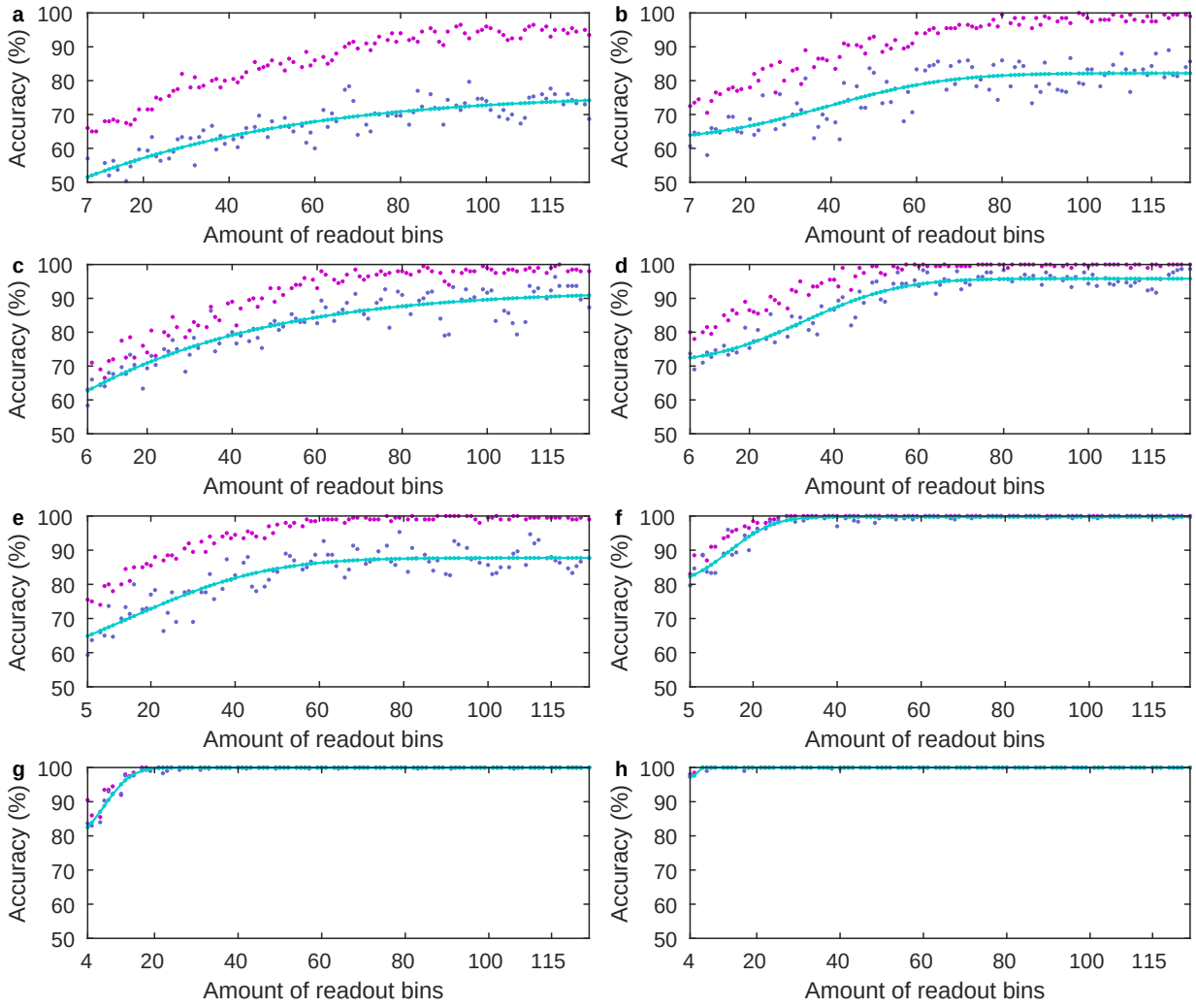


Figure 3.18: Experimental results for the n-bit parity test for increasing number of readout bins. Left column: Soliton number N=2. Right column: Soliton number N=5. For both cases, bottom to top represents 4 to 7 bit. Magenta = Training, Blue = Test, Cyan line = indication of mean value.

In all cases, an increase of readout bins beyond the input feature size improves the system performance significantly. Therefore, the number of readout bins used for training provide an additional hyper-parameter. Moreover, Fig. 3.17 confirms the trend, that an increase in system nonlinearity leads to an increase in performance for highly nonlinear tasks. From a practical viewpoint, a reduction of readout bins can be achieved by increasing system nonlinearity (i.e., less bins are required

for the same accuracy), which is especially important if programmable filters are considered for training [351], ultimately avoiding the long acquisitions times of an OSA and hence speeding up the processing rate of the system. However, it can be seen that after a certain number of bins, the performance does not increase any further. This indicates a practical system limitation in terms of nonlinearity for this task and can also be seen as an early indication of overfitting (see Fig. 3.22 for a further study).

Universal Function Approximation

A popular choice in neuromorphic computing and machine-learning for the investigation of the universal function approximation theorem [345, 352], i.e., the ability of the neural network to represent any given complex function. For ELMs, the normalized sinc regression is often used to investigate this [54, 56, 353]:

$$sinc(x) = \frac{\sin(\pi x)}{\pi x}. \tag{3.15}$$

Here, the task was performed by using a random encoding (i.e., a random digit from the MNIST dataset was used as a 'base' signal) with random phase values drawn from a uniform distribution in the range $[-\pi, \pi]$. In total, 1000 phase values (=samples) were measured for this task, as depicted in Fig. 3.19. An extraordinarily low (for physical systems) root mean square error (RMSE) of 0.0656 was achieved, making the soliton-fission based approach perform similar to recent multi-modal fiber ELM approaches [56] (RMSE of 0.0671). Yet, the implemented soliton-fission regime operates at two-orders of magnitude lower pulse energy (i.e. $\approx 90\,\mathrm{pJ}$ compared to $\approx 35\,\mathrm{nJ}$). For comparison, digital ELMs, in absence of noise[17], can achieve a test accuracy of RMSE = 0.0097 with just 20 nodes [353].

The results of the two task-independent benchmarks suggest that the implemented soliton fission concept can be interpreted as a complex neural network with at least one hidden layer. Solving the n-Bit parity problem requires either a single-hidden layer network [343] or multi-layer network [354] (if no direct input-output connections exist). Similarly, universal function approximation capability requires in general multi-layer networks [345, 352]. Notably, the propagation through the nonlinear fiber cannot be neither described through a single transfer function [191, 319] nor by a single

[17] for training, uniform noise was added resulting in RMSE = 0.1148 [353]

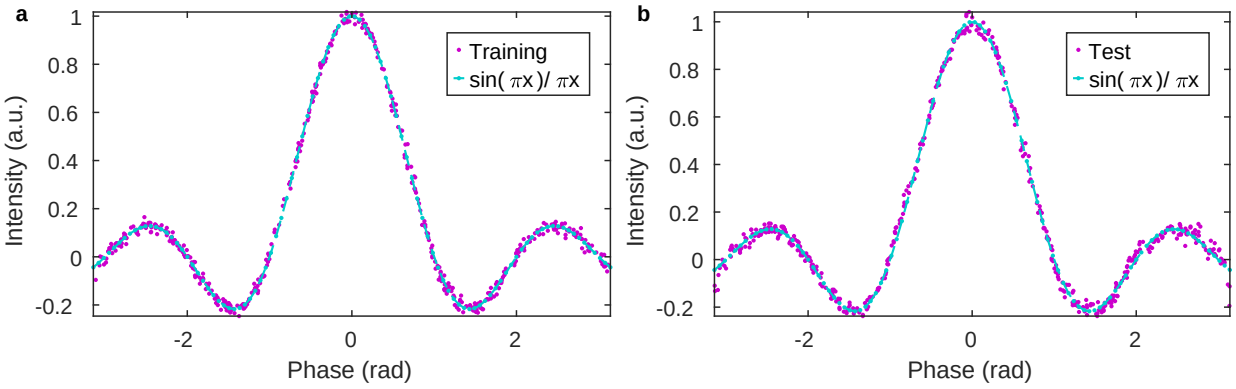


Figure 3.19: Experimental results for the sinc regression task. a) Training results (RMSE = 6.433e-7), b) Test results (RMSE = 0.0656).

hidden layer [80, 355]. *Hence, solving the n-Bit parity test for $n > 2$ and demonstrating nonlinear function approximation capability trough nonlinear propagation indicates the presence of a multi-layer network topology, in contrast to earlier works that treated fiber-based systems as nonlinear 'black-boxes' or transformers [56, 304].*

Task-dependent Benchmarks

Abalone Regression

The abalone task is a medium-sized, multi-parameter regression task [348] containing 4177 samples with eight input features. The goal of this task is to predict the age of different species of sea snails. Initially, the age of these snails is determined by cutting the shell in half, and counting the number of rings (similar to a tree). While this is a very time-consuming and invasive process, the age determination based on other factors such as size or weight pose a viable alternative. As such, the dataset features eight input features such as sex (encoding in a binary fashion), length/diameter/height, and several weight measurements with the target to determine the age. In fact, this can be a hard task for digital neural networks using computational less demanding metrics such as the L2 loss (see Appendix B and Appendix Fig. B.2).

To allow for comparison to previous implementations [56, 57], the data labels were normalized in the range [0,1]. Moreover, to allow reasonable measurement times and accuracy (see discussion around Fig. 3.22), the data set size was reduced to 2000 samples (incl. the minimum and maximum

age). Finally, the RMSE was calculated for performance evaluation. Fig. 3.20 shows the achieved performance for the Abalone age prediction task.

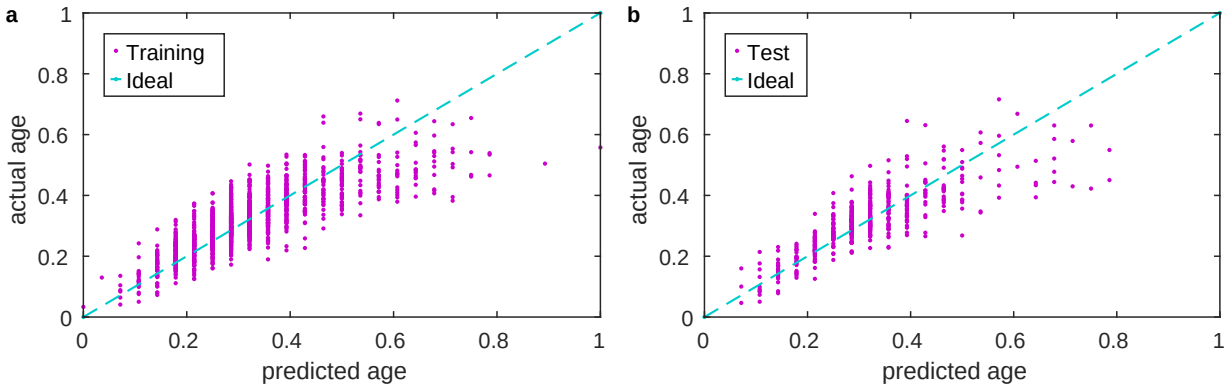


Figure 3.20: Experimental results for the Abalone regression task. a) Training results (RMSE = 0.0027). b) Test results (RMSE = 0.0766).

An RMSE of 0.0766 could be achieved for the test set, which is significantly better than previous approaches (RMSE of ≈0.12 in [56, 57]). Moreover, compared to spatial multi-mode systems [56] the required pulse energy is significantly lower (90 pJ vs. 35 nJ), highlighting the energy-efficiency of the soliton fission approach. However, while the data approximation works well for younger ages it can be seen that after a normalized age of ≈ 0.6, there is a larger deviation. Such deviation was also seen in previous studies [56] and might be either caused by the data itself (i.e. lack of samples of high age) or a lack of system complexity. Indeed, using a lower number of readout bins with the presented system *deforms* the data more significantly, i.e., the deviation occurs already at lower ages. The presented results were obtained for an optimized readout bin configuration.

IRIS Classification

The IRIS flower recognition is a highly relevant benchmark for testing out new classifiers [346]. It consists of measurements of petal/sepal lengths and widths of three different species of the Iris flower: Iris Setosa, Iris Versicolor, Iris Virginica. *While the dataset is relatively small, featuring only 150 samples and 3 classes, it is a non-linearly separable problem (i.e., two classes can be linearly separated, while the third one is not linearly separable).* Therefore, it is a common benchmark used in physical implementations [95, 294, 356–358].

The individual features were normalized per ensemble versus the maximum per feature (i.e., feature/max (feature)). Fig. 3.21(a,b) shows the results for the IRIS classification, where the di-

agonal refers to the correctly identified classes, while the off-diagonal elements correspond to false predictions. Remarkably, the system achieves a test accuracy for the unseen data of 100 %. This strong data separability is likely originating from the highly nonlinear transformation following the complex pulse propagation dynamics.

WINE Classification

The Wine task [347] is a standard benchmark for new classifiers. It features 178 samples with 13 input features (such as alcohol level or phenol levels) and the goal is to classify three different types of wine based on the provided input features. *The WINE task is a popular type of classification problem for new classifiers as it is regarded as a well-posed problem with low complexity and good class distribution.*

For the WINE data set, the same normalization as for the IRIS data set is applied. Fig. 3.21(c,d) shows the results for the WINE classification. Similarly to the IRIS test, a 100 % test accuracy for the unseen data could be achieved, although the training data contained one miss-classification in this case. Nevertheless, the implemented system shows state-of-the-art performance for these tasks and notably, no optical system to date has achieved such high accuracy for either of these data sets (see Appendix B.2).

MNIST Classification

A more complex task (in terms of input and output dimensionality) is the MNIST handwritten digit recognition. The handwritten digits are based on the modified set from the National Institute of Standards and Technology (That is why this set is commonly referred to as MNIST). The set contains a total of 70,000 handwritten digits (60,000 training, and 10,000 test images) from the numbers 0 to 9 [349]. While this task is not very challenging for digital network approaches (which reach up to 100% accuracy), it is an established benchmark for new physical systems. *Notably, it is not a very nonlinear data set and linear classifiers can already reach a high accuracy [359].*

Since the original MNIST images are 28×28 pixels (784 data points), two additional operations are required for information encoding through our node-limited programmable filter. Firstly, down-scaling is necessary as the used Waveshaper only features ≈400 wavelength bins. This was achieved by means of a biqubic interpolation method, were one output pixel is the result of a weighted average of the nearest 4×4 neighborhood. Secondly, since the Waveshaper features only a one-

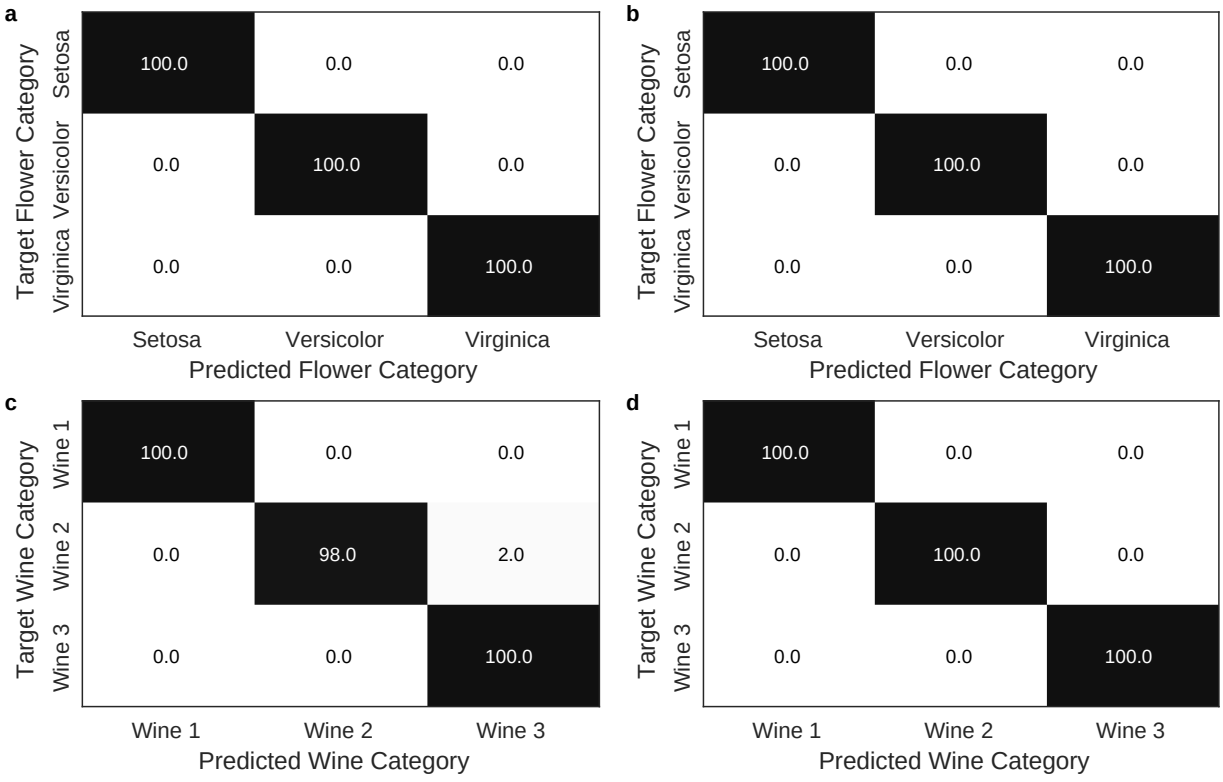


Figure 3.21: Experimental results for the IRIS and WINE tasks. a) Training result for the Iris task (100% accuracy), b) Test result for the Iris task (100% accuracy), c) Training result for the Wine task (99.3% accuracy), d) Test result for the Wine task (100% accuracy).

dimensional input array, the two-dimensional image data has to be flattened in order to generate a one-dimensional vector. This flattening was achieved by a straight-forward clock-wise spiral unwrapping starting from the top-left corner of the images. After these operations, the initial image ($28 \times 28 = 784$) was downsized to 10×10 and finally to a one-dimensional vector of size 1×100.

In order to study the effect of different system parameters, an iterative parameter search was performed first. For this, a limited data set size of 3000 random images was selected, while also ensuring an equal distribution among the classes. Fig. 3.22 summarizes the results. From this multi-parameter study, it can be seen that the system performance depends on several parameters. Notably, in the current system implementation, an additional trade-off exists between the system performance and data acquisition time. Therefore, not the best resolution (0.125 nm, ≈2 min data acquisition time for a 550 nm span) was selected for the benchmark tests, but a lower one (0.5 nm, 1 min data acquisition time) to allow for a reasonable measurement time.

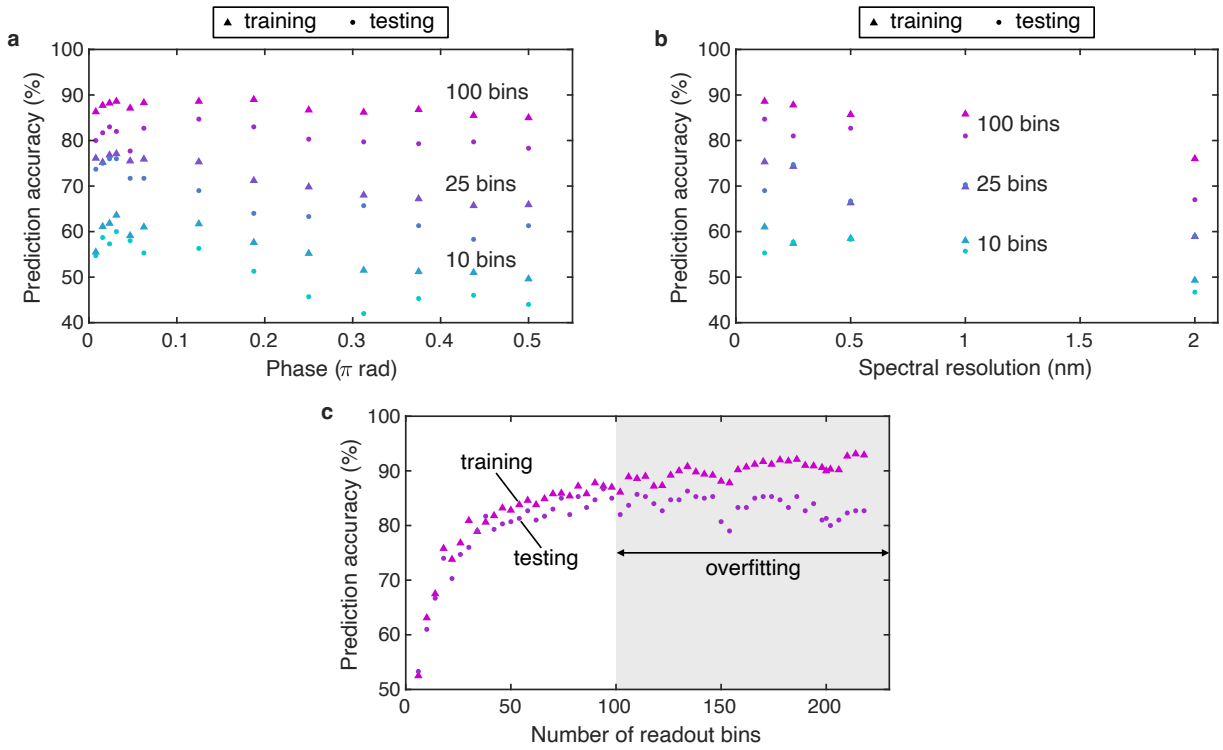


Figure 3.22: Performance for the MNIST task in dependence of different system parameters: a) Phase dependency for selected readout bin configurations (0.125 nm spectral resolution), b) Spectral resolution for selected readout configurations (phase value fixed to $\pi/8$), c) Dependency of readout bins (phase value fixed to $\pi/8$, 0.5 nm spectral resolution).

Finally, Fig. 3.23 shows the results for the MNIST task for the optimized parameter set ($\pi/8$ phase amplitude, 94 readout bins, 0.5 nm resolution).

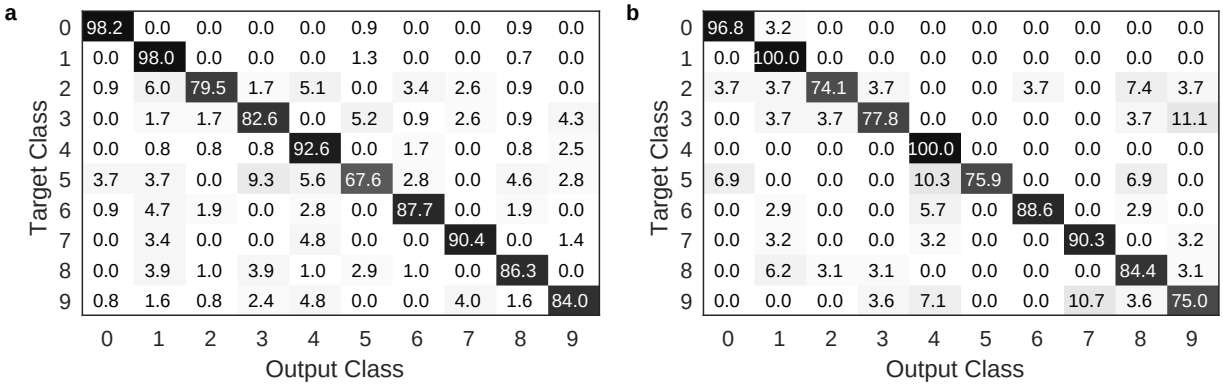


Figure 3.23: Experimental results for the MNIST task. a) Training result (87.2% accuracy), b) Test result (86.7% accuracy).

The system achieved a performance of ≈87 % for the unseen test data, which is on average for optical systems. A potential explanation for the lower performance in this task might be the strong

nonlinear data transformation due to the soliton fission process. Indeed, the MNIST is already solvable to a high degree with a linear digital classifier [359]. Remarkably, the highest accuracies reported for optical systems feature only weak nonlinear transformations such as those performed by photodiodes [57, 360]. This suggests, that i) the MNIST task of lower nonlinearity and ii) the strong nonlinearity mediated by the soliton fission process might be detrimental for problems of lower nonlinearity. Indeed, a lack of linearity in the system can be one of the main reasons for the detrimental performance in the MNIST task as previous studies highlighted for systems based on second harmonic generation [308]. Additionally, the 'equal-search' algorithm might also impose restrictions regarding potential solutions thus limiting the achievable performance. In this case, a more complex non-equidistant search algorithm might help improving the results.

Another factor for the decreased performance for the MNIST task might be related to the relatively simple down-sampling of the input data. More elaborate pre- and post-processing techniques can be implemented in order to increase performance. For example, a combination of input dimensionality reduction, such as PCA or histograms of oriented gradients (HOD) [360, 361] combined with linear classifiers such as SVMs [304] (instead of a linear regression at output) might lead to improved performance. Indeed, the use of dimensionality reduction techniques can have a big impact on the achievable accuracy in physical systems [360] as well.

Audio Classification for Health Monitoring

Disclaimer: The presented results in this section were part of an internal INRS collaboration. In particular, the data pre-processing of the initial COVID-19 speech data and results from digital neural network and support vector machines comparisons were performed by Yi Zhu, PhD student of Prof. Tiago Falk (INRS-EMT). The experimental realization with the optical system as shown in this section (encoding, collection, analysis) was performed by the author (B. Fischer).
The author (B. Fischer) would like to acknowledge the University of Cambridge for sharing the COVID-19 speech database and the INTERSPEECH 2021 ComParE Challenge organizers. The University of Cambridge does not bear any responsibility for the analysis and results presented in this thesis. All results and interpretations presented here represent only the view of the author (B. Fischer).

To showcase the performance of this approach for a real-world problem, the system is applied to a audio classification task in health monitoring. Indeed, utilizing audio recordings for health assess-

ment is a highly desired application [362], offering an inexpensive, continuous method for remote and decentralized health monitoring. Autonomous recognition of audio samples is of particularly interest for identifying respiratory illnesses like COVID-19 infections [363–365], where the use of such methods can speed-up testing, while also reducing costly analyses such as X-ray or computer tomography scans [365]. Thus, the use of machine-learning methods, such as straight-forward support vector machines or neural networks, have become of increasing interest for audio recognition and classification. However, identifying and classifying audio signals can be very challenging as they commonly feature a broad range of dynamics and generally signals are often distorted by environmental background noises. Usually, this directly affects the signal (and thus feature) quality and hence, ultimately, the performance of the implemented machine-learning framework. In particular, support vector machines (SVMs) shown remarkable performance for this task [363, 366, 367] and are also of high interest for remote applications (e.g., for use on mobile phones [368]) given their efficiency and reduced complexity [232, 369].

For the particular demonstration of the soliton fission-based system for audio classification, the data set provided by the *INTERSPEECH 2021* challenge is used and the performance is compared to a recent SVM implementation [367]. The data set consists of 856 (labeled) audio samples from individuals with and without a COVID-19 infection. The goal of this task is to identify infected versus healthy individuals solely based on the provided audio samples. Usually, natural language processing (NLP) such as speech recognition tasks requires extensive pre-processing often resulting in thousands of features (such as pitch or pace) [363, 370]. Recently, spectrograms (i.e., 2-dimensional time-frequency maps) have been investigated as input for neural networks. Using such maps, a speech recognition problem becomes essentially an image recognition/classification task. This is a common strategy for speech recognition in both software [371] and hardware approaches [372] as it simplifies the pre-processing significantly. For example, for the data from the *INTERSPEECH 2021* challenge, traditional feature extraction (e.g., using the openSMILE toolbox [373]) has been utilized in most demonstrations. However, this approach generates over 6000 features thus requiring large networks for processing [363]. Therefore, recent implementations have explored the use of spectrogram approaches with remarkable accuracy yet significantly reduced complexity[367].

In what follows, the feature extraction method is briefly described (see Fig. 3.24): First, the raw audio signal was converted to a spectrogram by means of a short-time Fourier transform (STFT) generating a two-dimensional map (or image). The spectrogram represents the different frequency

components at different times during speaking. Subsequently, another Fourier transform was implemented to generate so-called modulation spectrograms, which are particularly suitable for real-world data (i.e., data affected by environmental noise) [374]. The modulation spectrogram represents the rate of change of the individual frequency components during speaking. To provide usable features to the system (i.e., feature sizes <400), the modulation spectrogram is binned using ten bins (1 Hz size) for the modulation frequency (f_{mod}) and twenty bins (400 Hz size) for the conventional frequency (f), generating the final 20×20 map. Lastly, a two-step feature extraction is applied to obtain 30 features from the maps. In a first step, spectral low-level descriptors (LLDs) such as entropy, centroid and kurtosis were isolated from the generated maps [367], resulting in 320 LLDs per audio signal ($2 \times (8 \times 20)$). In a second step, 30 features were extracted from the 720 preselected LLDs (320 LLDs plus 400 energies from the binned modulation spectrogram) by means of a principal component analysis [375, 376].

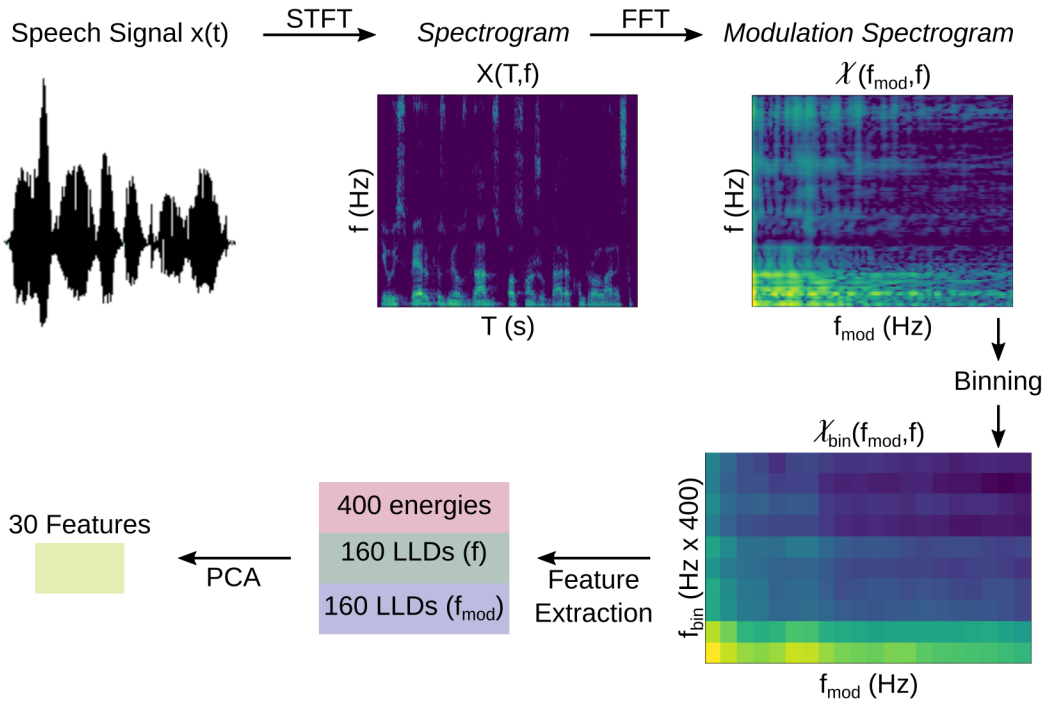


Figure 3.24: Illustration of the feature extraction for the COVID-19 task. STFT = short-time Fourier transform (256-point fast Fourier transform), FFT = fast Fourier transform, LLDs = low-level descriptors.

The extracted features were then normalized (min-max feature scaling) and encoded similarly to the previous tasks over the optical C-band and subsequently the data (i.e. output spectra for each COVID sample) were measured. Fig. 3.25 shows the achieved results. Since the provided data set shows a large imbalance (i.e., negative samples $\ll$ positive samples: 271 negative vs.

585 positive), the precision (total accuracy) is not a sufficient evaluation criterion. Therefore, the challenge organizers employed the unweighted average recall (UAR) as the evaluation metric for comparisons. The UAR can be calculated as

$$UAR = \frac{1}{\#\text{of classes}} \cdot Sensitivity + \frac{1}{\#\text{of classes}} \cdot Specificity. \tag{3.16}$$

The *sensitivity* can be calculated by dividing the number of correct positive classifications (true positives) by the total amount of positive classifications (false positives plus true positives). The *specificity* is calculated by dividing the number of correct negative classifications (true negatives) by the total amount of negative classifications (false negatives plus true negatives). A value for the UAR greater than 0.5 indicates a result above random guessing. The benchmark value for the COVID-19 challenge is 0.709 for the speech data [366].

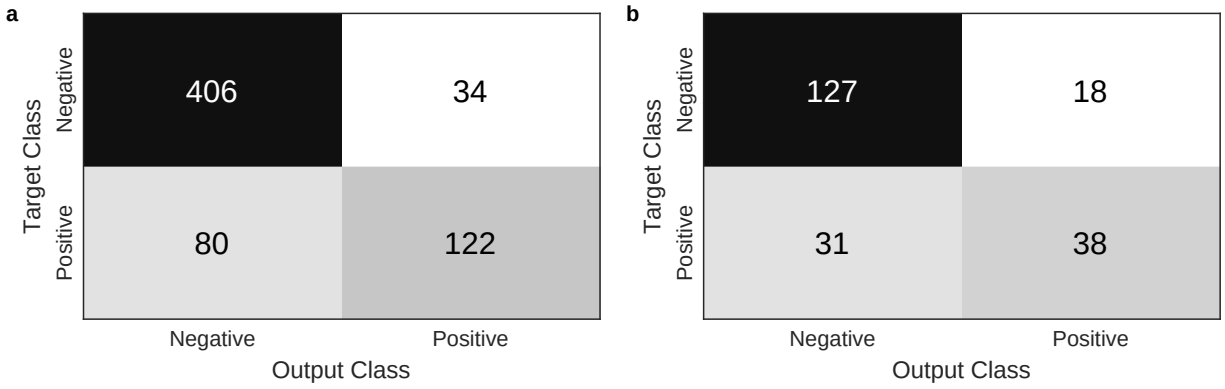


Figure 3.25: Experimental results for the COVID-19 data set (absolute numbers). a) Training results (UAR = 0.7633), b) Test results (UAR = 0.7133).

Since digital SVM approaches are state-of-the-art for this challenge [363, 366], the achieved result is compared to a digital SVM following a recent approach [367]. The digital SVM could achieve a training and test UAR of 0.8334 and 0.6309 (see Appendix B.3) using the same data partition as in the optical neural network experiment. Remarkably, the optical system (0.7133) outperforms the implemented SVM (0.6309), and also exceeds the benchmark value (0.709) for this task.

3.3.4 On-the-fly Network Training

Although good performance was achieved for several benchmark tasks (as summarized in Table 3.4), the developed platform has still room for improvement. In the current system implementation, the maximum number of solitons is 16, which limits scalability accordingly[18]. Therefore, a direct training of the supercontinuum network topology may be indispensable for more complex tasks. Such training can be either achieved by means of *in silico* waveguide design[309] or on-the-fly by manipulating the propagation dynamics. Indeed, the control over supercontinuum dynamics have been recently gained significant attraction for customized sources using for example SLMs [44, 81] or integrated devices such as the SDL from the previous Chapter in conjunction with evolutionary algorithms [42]. Thus, these techniques allow to alter the propagation dynamics (and thus the transient network topology) for different tasks.

In particular, a unique possibility that arises from the short time-scales achievable with the SDL is the generation of customizable supercontinua as recently demonstrated [42]. In this work, it was demonstrated that by employing a SDL, together with learning techniques such as genetic algorithms [377], the supercontinuum spectra can be optimized for certain parameters (e.g. high optical power in one or two selected wavelength bands). In particular, SC customizability arises from complex inter- and intra-pulse dynamics during propagation within the HNLF as illustrated in Fig. 3.26.

Fig. 3.26a) shows the simulated (based on equation 3.8) spectro-temporal dynamics of a single short pulse launched into an HNLF (fiber properties similar to the experiment in Chapter 2, fiber length 30 m) which mediates soliton fission. After propagating within 30 m of fiber, small nonlinear intra-pulse interaction between the soliton and the dispersive wave can be observed [321]. However, in the case of multiple input pulses (after 100 m propagation, see Fig. 3.26b), more complex intra- and inter-pulse dynamics arise. On top of the intra-pulse interaction between the soliton and the dispersive wave (compare the small circle in the bottom left of Fig. 3.26b), the interplay of multiple successive pulses shows complex inter-pulse dynamics arising from a rich interplay between different nonlinear processes, such as four-wave mixing cascades and cross-phase modulation between consecutive dispersive waves, solitons, and residual pump pulses. Consequently, the output spectrum of the supercontinuum significantly changes due to redistribution of energy caused by the complex

[18] representing the coherence limit [319]

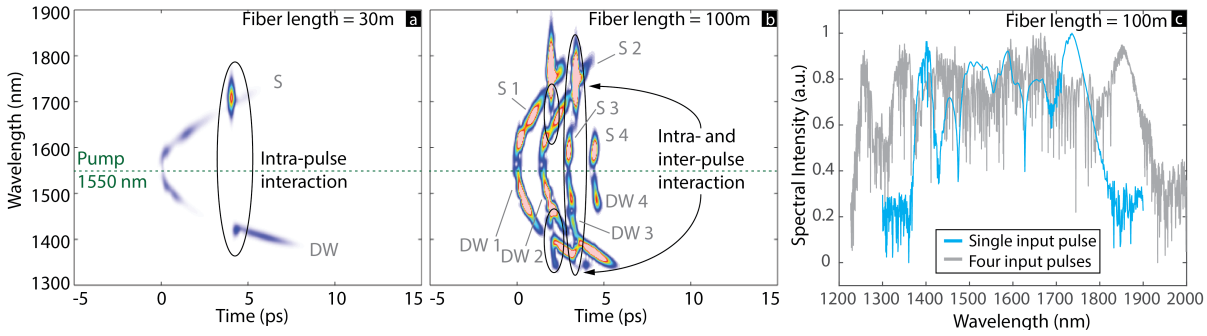


Figure 3.26: Spectrogram of simulated supercontinua: a) Single optical input pump pulse centered at 1550 nm (fiber parameters from Table 2.2, fiber length 30 m) with 1 kW peak power and 120 fs input pulse duration. In this case, distinct soliton fission (S) and dispersive wave (DW) generation can be observed as expected from theory [319]. Additionally, only intra-pulse interactions such as cross-phase modulation and four-wave mixing occur, leading to limited short-term nonlinear dynamics. b) Multiple optical input pulses, spaced by 1 ps, with slightly different amplitudes (while the first two pulses have the same amplitude, the second two pulses have 1/3 of the first two amplitudes, other parameters are the same as in (a)), are launched into 100 m HNLF resulting in multiple solitons and dispersive waves. In addition to the typical intra-pulse interactions (e.g. DW1-S1, left small circles), multiple pulses featuring separations in the order of the temporal walk-off also allow for inter-pulse interactions. This leads to multiple effects such as interaction of DW 2-DW 3 (S1-S2) or more complex S2-S3-DW3-DW4 (right big circle) resulting in significantly different output spectra. c) Output spectra for both simulated cases after 100 m propagation for a single input pulse (blue) and four input pulses (grey).

nonlinear interactions of multiple pulses, as illustrated in Fig. 3.26c.

In particular, this unique control mechanism can not only be used to generate customizable output spectra from a single temporally split input pulse [42], but moreover might be used to implement a training procedure for the presented system (compare the increase in 'density' of the spectrograms, which correspond to the virtual network nodes in Fig. 3.8). Indeed such an increase in 'density' might be comparable to changes in the network topology of a classical neural network. *Thus, active training of such devices might enable the realization of re-dense or pruning strategies for increased performance in such optical neural networks [238, 378, 379].* By monitoring selected, but in this case fixed, spectral readout bins, the training of the system can be mediated by supervised learning or guided optimization techniques akin to recent developments in digital approaches [380–383], especially for optimizing ELMs [384], as well as in optical neural networks [76, 95, 285, 334, 335]. The key factors for implementing learning algorithms on this scheme are similarly to the implementation presented before (i.e., requirements on pulse width and peak power for good coherence) but additionally it requires a fast optical encoding and readout for the active training.

3.4 Conclusion

In summary, a novel type of optical processor based on highly complex soliton fission dynamics has been implemented for the first time and characterized using different benchmark tasks. For the majority of tasks, the system shows above state-of-the-art performance (see Appendix Table B.1 for a comparison). In particular, the system seems well suited for highly nonlinear or complex tasks such as the performed IRIS or COVID-19 classification, while it appears to be less optimal for more linear-type tasks such as the popular MNIST task. The achieved results for all performed tasks are summarized in Table 3.4.

		Achieved Accuracy		**Readout Bin**		
Task	**Type**	**Training**	**Test**	**Amount**	**Position (nm)**	**Comment**
Sinc	regression	3.05e-4	0.0175	100	1500-1347	RMSE error
Abalone	regression	0.0684	0.0686	31	1485-1527	RMSE error
IRIS	classification	100%	100%	4	1237-1523	
WINE	classification	99.3%	100%	18	1537-1565	
MNIST	classification	87.2%	86.7%	94	1511-1562	
COVID-19	classification	0.7633	0.7133	119	1290-1408	UAR score

Table 3.4: Performance overview of the developed system for the different tasks, rmse = root mean square error, UAR = unweighted average recall.

A particular interesting result of these benchmark tasks is the spectral position of the readout bins. Indeed, the positions seem to indicate that more complex tasks (e.g., IRIS or COVID-19) result in readout bins far away from the pump as a result of a strong nonlinear transformation (i.e., dispersive wave generation), while for more linear tasks (e.g., WINE or MNIST) the readout bins are closer to the pump (i.e., a more linear regime). This finding suggests that such a photonic system might be even used to assess the degree of nonlinearity for a given task, which is a very useful information for the digital machine learning community as well, as it might help to estimate the number of hidden layers (i.e. dynamical range of a NN) for particular applications (or data sets) and hence potentially lower network complexities and power consumption.

Interestingly, optical neural networks are inevitably affected by environmental noise, hence providing a unique opportunity for the realization of more noise tolerant neural networks. In fact, noise can be detrimental for digital neural networks (e.g., adversarial attacks [385], where a single pixel can already significantly reduce the prediction performance [386]), as digital neural networks are

generally trained on noiseless data. Optical neural networks are therefore currently investigated as a potential solution to this problem as they feature intrinsic noise during training and operation [387].

Chapter 4

Conclusion & Outlook

In conclusion, this work has explored the possibility of using nonlinear photonics, machine-learning and guided optimization techniques for the implementation of two *smart* photonic devices based on very similar off-the-shelf components and repurposed devices.

In particular, the use of optimization techniques not only permits the generation of customized device outputs (e.g., the autonomous pulse-shaper of Chapter 2), but moreover enables the re-use of existing components towards new applications (e.g., the same chip was utilized towards nonlinear light generation [42]). Interestingly, the combination of such optimization methods with integrated optics could enable novel tools for improved device performances by reducing the impact of imperfect device fabrication. Indeed, strategies to reduce the impact of fabrication tolerances are of high interest to boost performance of optical neural networks and quantum optics applications [153–155]. As such, the presented methods can provide an essential tool in the *smart* photonic field, where success might rely on the implementation of inexpensive monitoring solutions with a low complexity. Such methods are crucial for the successful deployment of such techniques in real-world scenarios, where affordable, compact or efficient device operation is required. However, there are still open questions to be solved in these kind of applications. In detail, while the optimization algorithms are well suited to find an optimal operation point, they are usually not well suited for a long-term implementation (e.g., system stabilization). Systems often feature deteriorating performance if operated for longer times due to, for example, drift caused by changes in the environment such as temperature or vibrations. In such cases, researchers have explored the use of recovery algorithms in combination with local search algorithms [31] or reinforcement learning strategies that enable a long-

term operation by constantly monitoring the output state [102, 105, 214]. Thus, hybrid approaches combining different methodology, such as optimization and search algorithms, or machine-learning and evolutionary algorithms[388] might be pivotal for successful deployment.

Similarly, a standard off-the-shelf fiber, similar to the one used for the optical sampling in Chapter 2, has been demonstrated to be capable of performing machine-learning tasks in Chapter 3. Exploiting the rich dynamics of existing components has already been exploited in recent years for the implementation of optical machine-learning approaches in the context of reservoir computing [389]. Hence, the utilization of such off-the-shelf components provides straight-forward implementation of optical machine learning concepts compared to highly complex photonic circuits [63, 293].
The implemented system presents also several possibilities for future improvements, besides the previously discussed online training approach: Firstly, utilizing multi-core fibers can allow a high degree of parallelism where information can be processed in different cores without increasing the system footprint [390]. Such fibers might allow to reach TBit data rates comparable to recent integrated platforms [62, 63]. Additionally, multi-core fibers can provide further complexity to the systems. In particular, recent process in coupled multi-core fibers might allow adding additional dimensionality at just a fraction of power compared to multi-mode approaches [391–394]. Interestingly, also hybrid material fibers such as liquid-core fibers might be suitable options for such nonlinear processing schemes given the unique properties such as easily tunable dispersion through the use of different liquids, adaptive control over the dynamics through temperature gradients, or low-power operation [203, 204, 326, 395, 396]. The use of such hybrid fibers also allows a higher soliton number beyond the coherence limit of standard fibers for coherent soliton fission [395]. Notably, the implemented concept can also be combined with existing multi-mode approaches [56, 310, 397] as the combination of such system with multi-mode soliton dynamics [398] might offer even richer nonlinear dynamics compared to the single-mode regime. However, such an increase in nonlinear dynamics might come at the expense of decreased device efficiency as higher pulse energies are required.

Secondly, the system implementation towards single-shot operation can allow for highly efficient computing at ultra-high inference speeds (<1 ns for integrated waveguides), where the energy usage for data processing is only determined by the used pulse energy[1] (≈fJ-pJ if integrated [399–401]). This single-shot operation might be employed using time-stretch techniques in combination with

[1] in the demonstrated system the spectrum analyzer averages over thousands of pulses

electro-optic phase modulators [402] or dispersive imaging [403, 404]. Combining such methods with real-time spectral measurements such as dispersive Fourier transforms[177, 178, 404, 405] can enable single-shot operation. Given the high-speeds of these systems, training likely requires field-programmable gate arrays instead of standard micro-controller/computers [58, 91]. Realizing such an ultra-fast photonic computing system is of high interest for several applications.
For example, the timescales and achievable complexity can be particularly useful to compensate for fiber non-linearity in telecommunications networks [49, 406], where digital signal processing is becoming increasingly challenging. Another potential application can be direct analysis in optical imaging systems, which already use femto-second pulses as information carriers [18, 404].

Lastly, transferring the whole system to a single chip will allow for increased environmental stability, potential mass producability, and provide more flexibility for waveguide design (especially in terms of dispersion profiles). Each of the used elements can be for example realized in a silicon and silicon-nitride platform, from efficient femtosecond sources [209, 407], over encoding and imaging schemes [150, 408], to amplification [409], spectral measurement schemes [410–412] and optical weighting [356] for the readout layer [150, 413, 414]. But also hybrid integration of different material platforms can provide an excellent strategy to leverage specific advantages of certain materials [415]. Finally, this work demonstrates the breath of applicability that nonlinear optics has towards solving key tasks in the implementation of *smart* photonic devices.

Outlook - What's next?

The use of machine-learning and guided optimization approaches together with nonlinear optical effects provides a suitable basis to implement various *smart* photonic applications based on both, existing photonic infrastructure and highly specific integrated circuits. Recent progress in the field of photonic machine learning also increasingly investigates the use of different nonlinear effects for highly efficient optical computing. This includes for example, $\chi^{(2)}$ effects such as second-harmonic generation [308, 416, 417] or $\chi^{(3)}$ effects such as four-wave mixing [418]. In addition, the approaches outlined in Chapters 2 and 3 could be used in a complementary manner with novel fabrication methods such as photonic 3D printing. The resulting devices have gained significant interest over the last years for the implementation of waveguides, on-chip components, diffractive elements or coupler structures towards machine-learning applications [52, 205, 296, 419]. As such, 3D printing might provide a pivotal tool in combination with existing off-the-shelf infrastructure and components

to provide rapid prototyping and accelerated technology deployment, hence enabling complementary methodologies for complex on-chip circuits in a timely manner [64].
The combination of such approaches to leverage different aspects of each approach might seem challenging at first, yet is not out of reach. Nevertheless, some questions still remain:

- What kind of algorithm or loss function is best suitable for a given task, as discussed in Chapter 2. A recent example is broadband light generation, which has been demonstrated using evolutionary, as well as deep learning strategies [42, 43]. Yet, a comparison in i) performance and ii) efficiency (e.g., in terms of convergence speed or used computational resources) is lacking. Choosing an approach solely based on existing literature (e.g., "*this has always be done like this, that's why we use it*") might not be beneficial if no comparison between different implementations exist.

- Physical implementation offer unique platforms that allow to break with traditional computer science conventions, enabling the realization of novel and transformative approaches. But what kind of complexity is required for photonic neuromorphic systems as described in Chapter 3? How would one implement a system that can be very linear or very nonlinear depending on the different tasks to be solved. How would one efficiently train such a complex system?

Solving these questions will undoubtedly be a crucial milestone for the *smart* photonics community and will definitely spark a plethora of new approaches and implementations.

Bibliography

[1] Maynard AD (2015). Navigating the fourth industrial revolution. *Nature Nanotechnology*, 10(12):1005–1006. DOI:10.1038/nnano.2015.286.

[2] Jordan MI & Mitchell TM (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260. DOI:10.1126/science.aaa8415.

[3] Nehme E, Weiss LE, Michaeli T & Shechtman Y (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458. DOI:10.1364/OPTICA.5.000458.

[4] Zahavy T, Dikopoltsev A, Moss D, Haham GI, Cohen O, Mannor S & Segev M (2018). Deep Learning Reconstruction of Ultra-Short Pulses. *Optica*, 5(5):666. DOI:10.1364/OPTICA.5.000666.

[5] Butler KT, Davies DW, Cartwright H, Isayev O & Walsh A (2018). Machine learning for molecular and materials science. *Nature*, 559(7715):547–555. DOI:10.1038/s41586-018-0337-2.

[6] Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A & Kim C (2017). Machine learning in materials informatics: recent applications and prospects. *npj Computational Materials*, 3(1):54. DOI:10.1038/s41524-017-0056-5.

[7] Yang KK, Wu Z & Arnold FH (2019). Machine-learning-guided directed evolution for protein engineering. *Nature Methods*, 16(8):687–694. DOI:10.1038/s41592-019-0496-6.

[8] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P & Hassabis D (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589. DOI:10.1038/s41586-021-03819-2.

[9] Goldenberg SL, Nir G & Salcudean SE (2019). A new era: artificial intelligence and machine learning in prostate cancer. *Nature Reviews Urology*, 16(7):391–403. DOI:10.1038/s41585-019-0193-3.

[10] McIntosh C, Conroy L, Tjong MC, Craig T, Bayley A, Catton C, Gospodarowicz M, Helou J, Isfahanian N, Kong V, Lam T, Raman S, Warde P, Chung P, Berlin A & Purdie TG (2021). Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nature Medicine*, 27(6):999–1005. DOI:10.1038/s41591-021-01359-w.

[11] Khan FN, Fan Q, Lu C & Lau APT (2019). An Optical Communication's Perspective on Machine Learning and Its Applications. *Journal of Lightwave Technology*, 37(2):493–516. DOI:10.1109/JLT.2019.2897313.

[12] Nevin JW, Nallaperuma S, Shevchenko NA, Li X, Faruk MS & Savory SJ (2021). Machine learning for optical fiber communication systems: An introduction and overview. *APL Photonics*, 6(12):121101. DOI:10.1063/5.0070838.

[13] Genty G, Salmela L, Dudley JM, Brunner D, Kokhanovskiy A, Kobtsev S & Turitsyn SK (2021). Machine learning and applications in ultrafast photonics. *Nature Photonics*, 15(2):91–101. DOI:10.1038/s41566-020-00716-4.

[14] Willner AE, editor (2020). *Optical Fiber Telecommunications VII*. Willner AE, editor. Academic Press.

[15] Argyris A (2022). Photonic neuromorphic technologies in optical communications. *Nanophotonics*, 11(5):897–916. DOI:10.1515/nanoph-2021-0578.

[16] Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Soljačić M, Denz C, Miller DAB & Psaltis D (2020). Inference in artificial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47. DOI:10.1038/s41586-020-2973-6.

[17] Shastri BJ, Tait AN, Ferreira de Lima T, Pernice WHP, Bhaskaran H, Wright CD & Prucnal PR (2021). Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics*, 15(2):102–114. DOI:10.1038/s41566-020-00754-y.

[18] Chen CL, Mahjoubfar A, Tai LC, Blaby IK, Huang A, Niazi KR & Jalali B (2016). Deep learning in label-free cell classification. *Scientific Reports*, 6(1):21471. DOI:10.1038/srep21471.

[19] Saguy A, Jünger F, Peleg A, Ferdman B, Nehme E, Rohrbach A & Shechtman Y (2021). DeepROCS: from speckle patterns to superior-resolved images by deep learning in rotating coherent scattering microscopy. *Optics Express*, 29(15):23877. DOI:10.1364/OE.424730.

[20] Zhang Y, Liu T, Singh M, Çetintaş E, Luo Y, Rivenson Y, Larin KV & Ozcan A (2021). Neural network-based image reconstruction in swept-source optical coherence tomography using undersampled spectral data. *Light: Science & Applications*, 10(1):155. DOI:10.1038/s41377-021-00594-7.

[21] Närhi M, Salmela L, Toivonen J, Billet C, Dudley JM & Genty G (2018). Machine learning analysis of extreme events in optical fibre modulation instability. *Nature Communications*, 9(1):4923. DOI:10.1038/s41467-018-07355-y.

[22] Salmela L, Tsipinakis N, Foi A, Billet C, Dudley JM & Genty G (2021). Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network. *Nature Machine Intelligence*, 3(4):344–354. DOI:10.1038/s42256-021-00297-z.

[23] Rafique D & Velasco L (2018). Machine Learning for Network Automation: Overview, Architecture, and Applications [Invited Tutorial]. *Journal of Optical Communications and Networking*, 10(10):D126. DOI:10.1364/JOCN.10.00D126.

[24] Lun H, Liu X, Cai M, Zhang Y, Gao R, Hu W, Yi L & Zhuge Q (2021). Machine-learning-based telemetry for monitoring long-haul optical transmission impairments: methodologies and challenges [Invited]. *Journal of Optical Communications and Networking*, 13(10):E94. DOI:10.1364/JOCN.426826.

[25] Matzner R, Semrau D, Luo R, Zervas G & Bayvel P (2021). Making intelligent topology design choices: understanding structural and physical property performance implications in optical networks [Invited]. *Journal of Optical Communications and Networking*, 13(8):D53. DOI:10.1364/JOCN.423490.

[26] Ballard Z, Brown C, Madni AM & Ozcan A (2021). Machine learning and computation-enabled intelligent sensor design. *Nature Machine Intelligence*. DOI:10.1038/s42256-021-00360-9.

[27] Krenn M, Malik M, Fickler R, Lapkiewicz R & Zeilinger A (2016). Automated Search for new Quantum Experiments. *Physical Review Letters*, 116(9):090405. DOI:10.1103/PhysRevLett.116.090405.

[28] Melnikov AA, Poulsen Nautrup H, Krenn M, Dunjko V, Tiersch M, Zeilinger A & Briegel HJ (2018). Active learning machine learns to create new quantum experiments. *Proceedings of the National Academy of Sciences*, 115(6):1221–1226. DOI:10.1073/pnas.1714936115.

[29] Krenn M, Erhard M & Zeilinger A (2020). Computer-inspired quantum experiments. *Nature Reviews Physics*, 2(11):649–661. DOI:10.1038/s42254-020-0230-4.

[30] Woodward RI & Kelleher EJ (2016). Towards 'smart lasers': Self-optimisation of an ultrafast pulse source using a genetic algorithm. *Scientific Reports*, 6(October):1–9. DOI:10.1038/srep37616.

[31] Pu G, Yi L, Zhang L & Hu W (2019). Intelligent programmable mode-locked fiber laser with a human-like algorithm. *Optica*, 6(3):362. DOI:10.1364/OPTICA.6.000362.

[32] Kokhanovskiy A, Ivanenko A, Kobtsev S, Smirnov S & Turitsyn S (2019). Machine Learning Methods for Control of Fibre Lasers with Double Gain Nonlinear Loop Mirror. *Scientific Reports*, 9(1):2916. DOI:10.1038/s41598-019-39759-1.

[33] Pu G, Yi L, Zhang L & Hu W (2020a). Genetic Algorithm-Based Fast Real-Time Automatic Mode-Locked Fiber Laser. *IEEE Photonics Technology Letters*, 32(1):7–10. DOI:10.1109/LPT.2019.2954806.

[34] Pu G, Yi L, Zhang L, Luo C, Li Z & Hu W (2020b). Intelligent control of mode-locked femtosecond pulses by time-stretch-assisted real-time spectral analysis. *Light: Science and Applications*, 9(1). DOI:10.1038/s41377-020-0251-x.

[35] Girardot J, Billard F, Coillet A, Hertz É & Grelu P (2020). Autosetting Mode-Locked Laser Using an Evolutionary Algorithm and Time-Stretch Spectral Characterization. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(5). DOI:10.1109/JSTQE.2020.2985297.

[36] Baumert T, Brixner T, Seyfried V, Strehle M & Gerber G (1997). Femtosecond pulse shaping by an evolutionary algorithm with feedback. *Applied Physics B: Lasers and Optics*, 65(6):779–782. DOI:10.1007/s003400050346.

[37] Yelin D, Meshulach D & Silberberg Y (1997). Adaptive femtosecond pulse compression. *Optics Letters*, 22(23):1793. DOI:10.1364/OL.22.001793.

[38] Meshulach D, Yelin D & Silberberg Y (1998). Adaptive real-time femtosecond pulse shaping. *Journal of the Optical Society of America B*, 15(5):1615. DOI:10.1364/JOSAB.15.001615.

[39] Ohno K, Tanabe T & Kannari F (2002). Adaptive pulse shaping of phase and amplitude of an amplified femtosecond pulse laser by direct reference to frequency-resolved optical gating traces. *Journal of the Optical Society of America B*, 19(11):2781. DOI:10.1364/JOSAB.19.002781.

[40] Fischer B, Chemnitz M, MacLellan B, Roztocki P, Helsten R, Wetzel B, Little BE, Chu ST, Moss DJ, Azaña J & Morandotti R (2021). Autonomous on-chip interferometry for reconfigurable optical waveform generation. *Optica*, 8(10):1268. DOI:10.1364/OPTICA.435435.

[41] Michaeli L & Bahabad A (2018). Genetic algorithm driven spectral shaping of supercontinuum radiation in a photonic crystal fiber. *Journal of Optics*, 20(5):055501. DOI:10.1088/2040-8986/aab59c.

[42] Wetzel B, Kues M, Roztocki P, Reimer C, Godin PlL, Rowley M, Little BE, Chu ST, Viktorov EA, Moss DJ, Pasquazi A, Peccianti M & Morandotti R (2018). Customizing supercontinuum generation via on-chip adaptive temporal pulse-splitting. *Nature Communications*, 9(1):4884. DOI:10.1038/s41467-018-07141-w.

[43] Valensise CM, Giuseppi A, Cerullo G & Polli D (2021). Deep reinforcement learning control of white-light continuum generation. *Optica*, 8(2):239. DOI:10.1364/OPTICA.414634.

[44] Chang J, Li D, Xi T, Zhang L & Hao Z (2021). Spectral Hump Formation in Visible Region of Supercontinuum from Shaped Femtosecond Laser Filamentation in Fused Silica. *Photonics*, 8(8):339. DOI:10.3390/photonics8080339.

[45] Duport F, Schneider B, Smerieri A, Haelterman M & Massar S (2012). All-optical reservoir computing. *Optics Express*, 20(20):22783. DOI:10.1364/oe.20.022783.

[46] Dejonckheere A, Duport F, Smerieri A, Fang L, Oudar JL, Haelterman M & Massar S (2014). All-optical reservoir computer based on saturation of absorption. *Optics Express*, 22(9):10868. DOI:10.1364/OE.22.010868.

[47] Vinckier Q, Duport F, Smerieri A, Vandoorne K, Bienstman P, Haelterman M & Massar S (2015). High-performance photonic reservoir computer based on a coherently driven passive cavity. *Optica*, 2(5):438. DOI:10.1364/optica.2.000438.

[48] Tait AN, de Lima TF, Zhou E, Wu AX, Nahmias MA, Shastri BJ & Prucnal PR (2017). Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports*, 7(1): 7430. DOI:10.1038/s41598-017-07754-z.

[49] Huang C, Fujisawa S, de Lima TF, Tait AN, Blow EC, Tian Y, Bilodeau S, Jha A, Yaman F, Peng HT, Batshon HG, Shastri BJ, Inada Y, Wang T & Prucnal PR (2021). A silicon photonic–electronic neural network for fibre nonlinearity compensation. *Nature Electronics*, 4(11): 837–844. DOI:10.1038/s41928-021-00661-2.

[50] Ashtiani F, Geers AJ & Aflatouni F (2022). An on-chip photonic deep neural network for image classification. *Nature*, 606(7914):501–506. DOI:10.1038/s41586-022-04714-0.

[51] Chang J, Sitzmann V, Dun X, Heidrich W & Wetzstein G (2018). Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports*, 8(1):12324. DOI:10.1038/s41598-018-30619-y.

[52] Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M & Ozcan A (2018). All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008. DOI:10.1126/science.aat8084.

[53] Li J, Mengu D, Yardimci NT, Luo Y, Li X, Veli M, Rivenson Y, Jarrahi M & Ozcan A (2021). Spectrally encoded single-pixel machine vision using diffractive networks. *Science Advances*, 7(13):eabd7690. DOI:10.1126/sciadv.abd7690.

[54] Marcucci G, Pierangeli D & Conti C (2020). Theory of Neuromorphic Computing by Waves: Machine Learning by Rogue Waves, Dispersive Shocks, and Solitons. *Phys. Rev. Lett.*, 125:093901. DOI:10.1103/PhysRevLett.125.093901.

[55] Momeni A & Fleury R (2022). Electromagnetic wave-based extreme deep learning with nonlinear time-Floquet entanglement. *Nature Communications*, 13(1):2651. DOI:10.1038/s41467-022-30297-5.

[56] Tegin U, Yildirim M, Oguz I, Moser C & Psaltis D (2021). Scalable optical learning operator. *Nature Computational Science*, 1(8):542–549. DOI:10.1038/s43588-021-00112-0.

[57] Pierangeli D, Marcucci G & Conti C (2021). Photonic extreme learning machine by free-space optical propagation. *Photonics Research*, 9(8):1446. DOI:10.1364/PRJ.423531.

[58] Miscuglio M, Hu Z, Li S, George JK, Capanna R, Dalir H, Bardet PM, Gupta P & Sorger VJ (2020). Massively parallel amplitude-only Fourier neural network. *Optica*, 7(12):1812. DOI:10.1364/OPTICA.408659.

[59] Huang Y, Zhang W, Yang F, Du J & He Z (2019). Programmable matrix operation with reconfigurable time-wavelength plane manipulation and dispersed time delay. *Optics Express*, 27(15):20456. DOI:10.1364/OE.27.020456.

[60] Ahmed M, Al-Hadeethi Y, Bakry A, Dalir H & Sorger VJ (2020). Integrated photonic FFT for photonic tensor operations towards efficient and high-speed neural networks. *Nanophotonics*, 9(13):4097–4108. DOI:10.1515/nanoph-2020-0055.

[61] Miscuglio M & Sorger VJ (2020). Photonic tensor cores for machine learning. *Applied Physics Reviews*, 7(3):031404. DOI:10.1063/5.0001942.

[62] Xu X, Tan M, Corcoran B, Wu J, Boes A, Nguyen TG, Chu ST, Little BE, Hicks DG, Morandotti R, Mitchell A & Moss DJ (2021). 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature*, 589(7840):44–51. DOI:10.1038/s41586-020-03063-0.

[63] Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M, Le Gallo M, Fu X, Lukashchuk A, Raja AS, Liu J, Wright CD, Sebastian A, Kippenberg TJ, Pernice WHP & Bhaskaran H (2021). Parallel convolutional processing using an integrated photonic tensor core. *Nature*, 589(7840):52–58. DOI:10.1038/s41586-020-03070-1.

[64] Arthurs EG (2019). Perspectives on the future of photonics: the best is yet to come. *Advanced Photonics*, 1(01):1. DOI:10.1117/1.AP.1.1.010501.

[65] Willner AE, Khaleghi S, Chitgarha MR & Yilmaz OF (2014). All-Optical Signal Processing. *Journal of Lightwave Technology*, 32(4):660–680. DOI:10.1109/JLT.2013.2287219.

[66] Skinner S, Steck J & Behrman E (1994). Optical neural network using Kerr-type nonlinear materials. *Proceedings of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems*, IEEE Comput. Soc. Press, pages 12–15, Turin, Italy.

[67] Piccinotti D, MacDonald KF, Gregory S, Youngs I & Zheludev NI (2020). Artificial Intelligence for Photonics and Photonic Materials. *Reports on Progress in Physics*, 51(40):404001. DOI:10.1088/1361-6633/abb4c7.

[68] Furber S (2016). Large-scale neuromorphic computing systems. *Journal of Neural Engineering*, 13(5):051001. DOI:10.1088/1741-2560/13/5/051001.

[69] Markovic D, Mizrahi A, Querlioz D & Grollier J (2020). Physics for neuromorphic computing. *Nature Reviews Physics*, 2(9):499–510. DOI:10.1038/s42254-020-0208-2.

[70] Miller DAB (2017). Attojoule Optoelectronics for Low-Energy Information Processing and Communications. *Journal of Lightwave Technology*, 35(3):346–396. DOI:10.1109/JLT.2017.2647779.

[71] Garduño-Mejía J, Greenaway AH & Reid DT (2004). Programmable spectral phase control of femtosecond pulses by use of adaptive optics and real-time pulse measurement. *Journal of the Optical Society of America B*, 21(4):833. DOI:10.1364/JOSAB.21.000833.

[72] Andral U, Si Fodil R, Amrani F, Billard F, Hertz E & Grelu P (2015). Fiber laser mode locked through an evolutionary algorithm. *Optica*, 2(4):275. DOI:10.1364/OPTICA.2.000275.

[73] Strubell E, Ganesh A & McCallum A (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Stroudsburg, PA, USA.

[74] Miller DAB (2010). Are optical transistors the logical next step? *Nature Photonics*, 4(1):3–5. DOI:10.1038/nphoton.2009.240.

[75] Burr GW (2019). A role for optics in AI hardware. *Nature*, 569(7755):199–200. DOI:10.1038/d41586-019-01406-0.

[76] Wu J, Lin X, Guo Y, Liu J, Fang L, Jiao S & Dai Q (2022). Analog Optical Computing for Artificial Intelligence. *Engineering*, 10:133–145. DOI:10.1016/j.eng.2021.06.021.

[77] Xu R, Lv P, Xu F & Shi Y (2021). A survey of approaches for implementing optical neural networks. *Optics & Laser Technology*, 136:106787. DOI:10.1016/j.optlastec.2020.106787.

[78] Zhou H, Dong J, Cheng J, Dong W, Huang C, Shen Y, Zhang Q, Gu M, Qian C, Chen H, Ruan Z & Zhang X (2022). Photonic matrix multiplication lights up photonic accelerator and beyond. *Light: Science & Applications*, 11(1):30. DOI:10.1038/s41377-022-00717-8.

[79] Adams GS, Converse BA, Hales AH & Klotz LE (2021). People systematically overlook subtractive changes. *Nature*, 592(7853):258–261. DOI:10.1038/s41586-021-03380-y.

[80] Salmela L, Hary M, Mabed M, Foi A, Dudley JM & Genty G (2022). Feed-forward neural network as nonlinear dynamics integrator for supercontinuum generation. *Optics Letters*, 47(4): 802. DOI:10.1364/OL.448571.

[81] Michaeli L & Bahabad A (2018). Genetic algorithm driven spectral shaping of supercontinuum radiation in a photonic crystal fiber. *Journal of Optics*, 20(5):055501. DOI:10.1088/2040-8986/aab59c.

[82] Jiang HM, Xie K & Wang YF (2004). Design of multi-pumped Raman fiber amplifier by particle swarm optimization. *Guangdianzi Jiguang/Journal of Optoelectronics Laser*, 15(10):1190–1193.

[83] Zhang WQ, Afshar V. S & Monro TM (2009). A genetic algorithm based approach to fiber design for high coherence and large bandwidth supercontinuum generation. *Optics Express*, 17(21):19311. DOI:10.1364/OE.17.019311.

[84] Wang Q & Ho ST (2010). Ultracompact Multimode Interference Coupler Designed by Parallel Particle Swarm Optimization With Parallel Finite-Difference Time-Domain. *Journal of Lightwave Technology*, 28(9):1298–1304. DOI:10.1109/JLT.2010.2040582.

[85] Han Kz, Huang Y, Liu Ff, Pang X, Hu P, Liu Gw, Qin H, Zhang F, Ge Xl, Liu Xj & Geng X (2018). An intelligent method to design laser resonator with particle swarm optimization algorithm. *Optoelectronics Letters*, 14(6):425–428. DOI:10.1007/s11801-018-8073-2.

[86] Guo D, Yin L & Yuan G (2019). New automatic optical design method based on combination of particle swarm optimization and least squares. *Optics Express*, 27(12):17027. DOI:10.1364/oe.27.017027.

[87] Kokhanovskiy A, Kuprikov E, Bednyakova A, Popkov I, Smirnov S & Turitsyn S (2021). Inverse design of mode-locked fiber laser by particle swarm optimization algorithm. *Scientific Reports*, 11(1):13555. DOI:10.1038/s41598-021-92996-1.

[88] Simon D (2013). *Evolutionary Optimization Algorithms*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 784 pages.

[89] Oppenheim A, Willsky A & Nawab H (2015). *Signals and Systems*. Pearson, 2 edition.

[90] Ziegler J & Nichols N (1942). Optimum Settings for Automatic Controllers. *Transactions of the ASME*, 64:759–768.

[91] Antonik P, Haelterman M & Massar S (2017). Online Training for High-Performance Analogue Readout Layers in Photonic Reservoir Computers. *Cognitive Computation*, 9(3):297–306. DOI:10.1007/s12559-017-9459-3.

[92] Iegorov R, Teamir T, Makey G & Ilday FÖ (2016). Direct control of mode-locking states of a fiber laser. *Optica*, 3(12):1312. DOI:10.1364/OPTICA.3.001312.

[93] Zhang Y, Wang S, She M & Zhang W (2022). Artificial Intelligence Enabled Spectral Reconfigurable Fiber Laser. *arXiv*, 2204.05146 pages.

[94] Lee KF, Zhou G, Yun Y, Jiang J, Winful HG & Fermann ME (2021). Surpassing soliton compression limits in anomalous dispersion high-power erbium fiber comb. *Optica*, 8(6):846. DOI:10.1364/OPTICA.427977.

[95] Zhang H, Thompson J, Gu M, Jiang XD, Cai H, Liu PY, Shi Y, Zhang Y, Karim MF, Lo GQ, Luo X, Dong B, Kwek LC & Liu AQ (2021). Efficient On-Chip Training of Optical Neural Networks Using Genetic Algorithm. *ACS Photonics*, 8(6):1662–1672. DOI:10.1021/acsphotonics.1c00035.

[96] Winters DG, Kirchner MS, Backus SJ & Kapteyn HC (2017). Electronic initiation and optimization of nonlinear polarization evolution mode-locking in a fiber laser. *Optics Express*, 25(26):33216. DOI:10.1364/OE.25.033216.

[97] Omenetto FG, Taylor AJ, Moores MD & Reitze DH (2001). Adaptive control of femtosecond pulse propagation in optical fibers. *Optics Letters*, 26(12):938. DOI:10.1364/OL.26.000938.

[98] Hellwig T, Walbaum T, Groß P & Fallnich C (2010). Automated characterization and alignment of passively mode-locked fiber lasers based on nonlinear polarization rotation. *Applied Physics B*, 101(3):565–570. DOI:10.1007/s00340-010-4224-2.

[99] Feng Q, Zhang B, Liu Z, Lin C & Ding Y (2017). Research on intelligent algorithms for amplitude optimization of wavefront shaping. *Applied Optics*, 56(12):3240. DOI:10.1364/ao.56.003240.

[100] Wei X, Jing JC, Shen Y & Wang LV (2020). Harnessing a multi-dimensional fibre laser using genetic wavefront shaping. *Light: Science and Applications*, 9(1). DOI:10.1038/s41377-020-00383-8.

[101] Meng F & Dudley JM (2020). Toward a self-driving ultrafast fiber laser. *Light: Science & Applications*, 9(1):7–10. DOI:10.1038/s41377-020-0270-7.

[102] Yan Q, Deng Q, Zhang J, Zhu Y, Yin K, Li T, Wu D & Jiang T (2021). Low-latency deep-reinforcement learning algorithm for ultrafast fiber lasers. *Photonics Research*, 9(8):1493. DOI:10.1364/PRJ.428117.

[103] Wu X, Peng J, Boscolo S, Zhang Y, Finot C & Zeng H (2022a). Intelligent Breathing Soliton Generation in Ultrafast Fiber Lasers. *Laser & Photonics Reviews*, 16(2):2100191. DOI:10.1002/lpor.202100191.

[104] Wu X, Zhang Y, Peng J, Boscolo S, Finot C & Zeng H (2022b). Farey tree and devil's staircase of frequency-locked breathers in ultrafast lasers. *Nature Communications*, 13(1):5784. DOI:10.1038/s41467-022-33525-0.

[105] Sun C, Kaiser E, Brunton SL & Nathan Kutz J (2020). Deep reinforcement learning for optical systems: A case study of mode-locked lasers. *Machine Learning: Science and Technology*, 1(4):045013. DOI:10.1088/2632-2153/abb6d6.

[106] Shampine LF (1982). Implementation of Rosenbrock Methods. *ACM Transactions on Mathematical Software*, 8(2):93–113. DOI:10.1145/355993.355994.

[107] Bagley JD (1967). *The behavior of adaptive systems which employ genetic and correlation algorithms*. Ph.d. dissertation, University of Michigan.

[108] Holland JH (1975). *Adaption in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

[109] Kennedy J & Eberhart RC (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, IEEE, pages 1942–1948.

[110] Liepins GE & Hilliard MR (1989). Genetic algorithms: Foundations and applications. *Annals of Operations Research*, 21(1):31–57. DOI:10.1007/BF02022092.

[111] Clerc M & Kennedy J (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73. DOI:10.1109/4235.985692.

[112] Zhang W, Liu Y & Clerc M (2003). An adaptive PSO algorithm for reactive power optimization. *6th International Conference on Advances in Power System Control, Operation and Management. Proceedings. APSCOM 2003*, IEE, volume 2003, pages 302–307.

[113] Nickabadi A, Ebadzadeh MM & Safabakhsh R (2011). A novel particle swarm optimization algorithm with adaptive inertia weight. *Applied Soft Computing*, 11(4):3658–3670. DOI:10.1016/j.asoc.2011.01.037.

[114] Chander A, Chatterjee A & Siarry P (2011). A new social and momentum component adaptive PSO algorithm for image segmentation. *Expert Systems with Applications*, 38(5):4998–5004. DOI:10.1016/j.eswa.2010.09.151.

[115] Mancinelli M, Bettotti P & Pavesi L (2020). A silicon photonics feed-forward neural network for nonlinear distortion mitigation in an optical link. *arXiv*, 2012.07981 pages.

[116] Pérez-López D, López A, DasMahapatra P & Capmany J (2020). Multipurpose self-configuration of programmable photonic circuits. *Nature Communications*, 11(1):6359. DOI:10.1038/s41467-020-19608-w.

[117] Grefenstette J (1986). Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(1):122–128. DOI:10.1109/TSMC.1986.289288.

[118] Schmidhuber J, Hochreiter S & Bengio Y (2009). Evaluating Benchmark Problems by Random Guessing. *A Field Guide to Dynamical Recurrent Networks*. Kolen JF & Kremer SC (editors), IEEE, pages 1–12.

[119] Hendrix EM & Lančinskas A (2015). On Benchmarking Stochastic Global Optimization Algorithms. *Informatica (Netherlands)*, 26(4):649–662. DOI:10.15388/Informatica.2015.69.

[120] Chen S & Montgomery J (2013). Particle swarm optimization with thresheld convergence. *2013 IEEE Congress on Evolutionary Computation*, IEEE, pages 510–516.

[121] Weiner AM (2009). *Ultrafast Optics*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[122] Weiner AM (2011). Ultrafast optical pulse shaping: A tutorial review. *Optics Communications*, 284(15):3669–3692. DOI:10.1016/j.optcom.2011.03.084.

[123] Boyd RW (2008). *Nonlinear Optics*. Elsevier.

[124] Fülöp J, Major Z, Horváth B, Tavella F, Baltuška A & Krausz F (2007). Shaping of picosecond pulses for pumping optical parametric amplification. *Applied Physics B*, 87(1):79–84. DOI:10.1007/s00340-006-2488-3.

[125] Asghari MH & Azaña J (2008). Proposal and analysis of a reconfigurable pulse shaping technique based on multi-arm optical differentiators. *Optics Communications*, 281(18):4581–4588. DOI:10.1016/j.optcom.2008.05.037.

[126] Huh J & Azaña J (2015). Generation of high-quality parabolic pulses with optimized duration and energy by use of dispersive frequency-to-time mapping. *Optics Express*, 23(21):27751. DOI:10.1364/OE.23.027751.

[127] Ma Y, Huang X, Wang X, Ji L, He Y, Qiu L, Zhao J, Wang Y & Wu S (2020). Precise pulse shaping for quantum control of strong optical transitions. *Optics Express*, 28(12):17171. DOI:10.1364/OE.389700.

[128] Yongqian Liu, Sang-Gyu Park & Weiner A (1996). Terahertz waveform synthesis via optical pulse shaping. *IEEE Journal of Selected Topics in Quantum Electronics*, 2(3):709–719. DOI:10.1109/2944.571771.

[129] Vidal S, Degert J, Oberlé J & Freysz E (2010). Femtosecond optical pulse shaping for tunable terahertz pulse generation. *Journal of the Optical Society of America B*, 27(5):1044. DOI:10.1364/JOSAB.27.001044.

[130] Bader M, Heugel S, Chekhov AL, Sondermann M & Leuchs G (2013). Efficient coupling to an optical resonator by exploiting time-reversal symmetry. *New Journal of Physics*, 15(12):123008. DOI:10.1088/1367-2630/15/12/123008.

[131] Srivathsan B, Gulati GK, Cerè A, Chng B & Kurtsiefer C (2014). Reversing the Temporal Envelope of a Heralded Single Photon using a Cavity. *Physical Review Letters*, 113(16):163601. DOI:10.1103/PhysRevLett.113.163601.

[132] Liu C, Sun Y, Zhao L, Zhang S, Loy MMT & Du S (2014). Efficiently Loading a Single Photon into a Single-Sided Fabry-Perot Cavity. *Physical Review Letters*, 113(13):133601. DOI:10.1103/PhysRevLett.113.133601.

[133] Kues M, Reimer C, Roztocki P, Cortés LR, Sciara S, Wetzel B, Zhang Y, Cino A, Chu ST, Little BE, Moss DJ, Caspani L, Azaña J & Morandotti R (2017). On-chip generation of high-dimensional entangled quantum states and their coherent control. *Nature*, 546(7660):622–626. DOI:10.1038/nature22986.

[134] Imany P, Jaramillo-Villegas JA, Odele OD, Han K, Leaird DE, Lukens JM, Lougovski P, Qi M & Weiner AM (2018). 50-GHz-spaced comb of high-dimensional frequency-bin entangled photons from an on-chip silicon nitride microresonator. *Optics Express*, 26(2):1825. DOI:10.1364/OE.26.001825.

[135] Weiner AM, Heritage JP & Thurston RN (1986). Synthesis of phase-coherent, picosecond optical square pulses. *Optics Letters*, 11(3):153. DOI:10.1364/ol.11.000153.

[136] Weiner AM, Leaird DE, Patel JS & Wullert JR (1992). Programmable Shaping of Femtosecond Optical Pulses by Use of 128-Element Liquid Crystal Phase Modulator. *IEEE Journal of Quantum Electronics*, 28(4):908–920. DOI:10.1109/3.135209.

[137] Weiner AM (2000). Femtosecond pulse shaping using spatial light modulators. *Review of Scientific Instruments*, 71(5):1929–1960. DOI:10.1063/10.1150614.

[138] Clarke AM, Williams DG, Roelens MA & Eggleton BJ (2010). Reconfigurable optical pulse generator employing a fourier-domain programmable optical processor. *Journal of Lightwave Technology*, 28(1):97–103. DOI:10.1109/JLT.2009.2034872.

[139] Liu F, Huang S, Si S, Zhao G, Liu K & Zhang S (2019). Generation of picosecond pulses with variable temporal profiles and linear polarization by coherent pulse stacking in a birefringent crystal shaper. *Optics Express*, 27(2):1467. DOI:10.1364/oe.27.001467.

[140] Mittelberger DE, Muir RD, Hamamoto MY, Prantil MA & Heebner JE (2019). Frequency-to-time optical arbitrary waveform generator. *Optics Letters*, 44(11):2863. DOI:10.1364/ol.44.002863.

[141] Huh J & Azaña J (2016). In-fiber reconfigurable generation of arbitrary (asymmetric) picosecond temporal intensity waveforms by time-domain optical pulse shaping. *Optics Letters*, 41(4):693. DOI:10.1364/ol.41.000693.

[142] Park Y, Asghari MH, Ahn TJ & Azaña J (2007). Transform-limited picosecond pulse shaping based on temporal coherence synthesization. *Optics Express*, 15(15):9584. DOI:10.1364/OE.15.009584.

[143] Yin L, Wang H, Reagan BA & Rocca JJ (2019). Programmable pulse synthesizer for the generation of Joule-level picosecond laser pulses of arbitrary shape. *Optics Express*, 27(24):35325. DOI:10.1364/oe.27.035325.

[144] Liao S, Min S & Dong J (2014). On-chip optical pulse shaper for arbitrary waveform generation using optical gradient force. *European Conference on Optical Communication, ECOC*, pages 5–7. DOI:10.1109/ECOC.2014.6963847.

[145] Liao S, Ding Y, Dong J, Yang T, Chen X, Gao D & Zhang X (2015). Arbitrary waveform generator and differentiator employing an integrated optical pulse shaper. *Optics Express*, 23(9): 12161. DOI:10.1364/oe.23.012161.

[146] Liao S, Ding Y, Dong J, Yan S, Wang X & Zhang X (2016). Photonic arbitrary waveform generator based on Taylor synthesis method. *Optics Express*, 24(21):24390. DOI:10.1364/oe.24.024390.

[147] Feng S, Qin C, Shang K, Pathak S, Lai W, Guan B, Clements M, Su T, Liu G, Lu H, Scott RP & Ben Yoo SJ (2017). Rapidly reconfigurable high-fidelity optical arbitrary waveform generation in heterogeneous photonic integrated circuits. *Optics Express*, 25(8):8872. DOI:10.1364/oe.25.008872.

[148] Wang X, Zhou L, Li R, Xie J, Lu L, Wu K & Chen J (2017). Continuously tunable ultra-thin silicon waveguide optical delay line. *Optica*, 4(5):507. DOI:10.1364/OPTICA.4.000507.

[149] Zhao Y, Wang X, Gao D, Dong J & Zhang X (2019). On-chip programmable pulse processor employing cascaded MZI-MRR structure. *Frontiers of Optoelectronics*, 12(2):148–156. DOI:10.1007/s12200-018-0846-5.

[150] Zheng S, Long Y, Gao D, Luo Y, Wang L, Cao X & Wang J (2020). Chip-Scale Reconfigurable Optical Full-Field Manipulation: Enabling a Compact Grooming Photonic Signal Processor. *ACS Photonics*, 7(5):1235–1245. DOI:10.1021/acsphotonics.0c00103.

[151] Liu F, Wang T, Qiang L, Ye T, Zhang Z, Qiu M & Su Y (2008). Compact optical temporal differentiator based on silicon microring resonator. *Optics Express*, 16(20):15880. DOI:10.1364/OE.16.015880.

[152] Hu Y, Zhang L, Xiao X, Li Z, Li Y, Chu T, Su Y, Yu Y & Yu J (2012). An ultra-high-speed photonic temporal differentiator using cascaded SOI microring resonators. *Journal of Optics*, 14(6):065501. DOI:10.1088/2040-8978/14/6/065501.

[153] Fang MYS, Manipatruni S, Wierzynski C, Khosrowshahi A & DeWeese MR (2019). Design of optical neural networks with component imprecisions. *Optics Express*, 27(10):14009. DOI:10.1364/OE.27.014009.

[154] Flamini F (2021). Emergence of Biased Errors in Imperfect Optical Circuits. *Physical Review Applied*, 16(6):064038. DOI:10.1103/PhysRevApplied.16.064038.

[155] Bandyopadhyay S, Hamerly R & Englund D (2021). Hardware error correction for programmable photonics. *Optica*, 8(10):1247. DOI:10.1364/OPTICA.424052.

[156] Hacker M, Stobrawa G & Feurer T (2001). Iterative Fourier transform algorithm for phase-only pulse shaping. *Optics Express*, 9(4):191. DOI:10.1364/OE.9.000191.

[157] Omenetto F, Moores M, Luce B, Reitze D & Taylor A (2001). Femtosecond pulse delivery through single-mode optical fiber with adaptive pulse shaping. *Technical Digest. Summaries of papers presented at the Conference on Lasers and Electro-Optics. Postconference Technical Digest (IEEE Cat. No.01CH37170)*, IEEE, pages 234–235.

[158] Thomas S, Malacarne A, Fresi F, Potì L & Azaña J (2010). Fiber-based programmable picosecond optical pulse shaper. *Journal of Lightwave Technology*, 28(12):1832–1843. DOI:10.1109/JLT.2010.2048700.

[159] Saleh BEA & Teich MC (1991). *Fundamentals of photonics*. Wiley Series in Pure and Applied Optics. John Wiley & Sons, Inc., New York, USA.

[160] Bogaerts W, Pérez D, Capmany J, Miller DAB, Poon J, Englund D, Morichetti F & Melloni A (2020). Programmable photonic circuits. *Nature*, 586(7828):207–216. DOI:10.1038/s41586-020-2764-0.

[161] Little B, Sai Chu, Wei Chen, Wenlu Chen, Hryniewicz J, Gill D, King O, Johnson F, Davidson R, Donovan K & Gibson J (2006). Compact optical programmable delay lines with fast thermo-optic switching and output power balancing. *IEEE Conference Avionics Fiber-Optics and Photonics, 2006.*, IEEE, pages 68–69.

[162] Ferrera M, Razzari L, Duchesne D, Morandotti R, Yang Z, Liscidini M, Sipe JE, Chu ST, Little BE & Moss DJ (2008). Low-power continuous-wave nonlinear optics in doped silica glass integrated waveguide structures. *Nature Photonics*, 2(12):737–740. DOI:10.1038/nphoton.2008.228.

[163] Moss DJ, Morandotti R, Gaeta AL & Lipson M (2013). New CMOS-compatible platforms based on silicon nitride and Hydex for nonlinear optics. *Nature Photonics*, 7(8):597–607. DOI:10.1038/nphoton.2013.183.

[164] Baets R, Subramanian AZ, Clemmen S, Kuyken B, Bienstman P, Le Thomas N, Roelkens G, Van Thourhout D, Helin P & Severi S (2016). Silicon photonics: silicon nitride versus silicon-on-insulator. *Optical Fiber Communication Conference*, OSA, Th3J.1 pages, Washington, D.C.

[165] Liang TK & Tsang HK (2004a). Role of free carriers from two-photon absorption in Raman amplification in silicon-on-insulator waveguides. *Applied Physics Letters*, 84(15):2745–2747. DOI:10.1063/1.1702133.

[166] Liang TK & Tsang HK (2004b). Efficient Raman amplification in silicon-on-insulator waveguides. *Applied Physics Letters*, 85(16):3343–3345. DOI:10.1063/1.1807960.

[167] Pérez D, Gasulla I, Capmany J & Soref RA (2016). Reconfigurable lattice mesh designs for programmable photonic processors. *Optics Express*, 24(11):12093. DOI:10.1364/OE.24.012093.

[168] Tran MA, Komljenovic T, Hulme JC, Davenport ML & Bowers JE (2016). A Robust Method for Characterization of Optical Waveguides and Couplers. *IEEE Photonics Technology Letters*, 28(14):1517–1520. DOI:10.1109/LPT.2016.2556713.

[169] Trebino R, DeLong KW, Fittinghoff DN, Sweetser JN, Krumbügel MA, Richman BA & Kane DJ (1997). Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Review of Scientific Instruments*, 68(9):3277–3295. DOI:10.1063/1.1148286.

[170] Foster MA, Salem R, Geraghty DF, Turner-Foster AC, Lipson M & Gaeta AL (2008). Silicon-chip-based ultrafast optical oscilloscope. *Nature*, 456(7218):81–84. DOI:10.1038/nature07430.

[171] Jie Li, Hansryd J, Hedekvist P, Andrekson P & Knudsen S (2001). 300 Gbit/s eye-diagram measurement by optical sampling using fiber based parametric amplification. *OFC 2001. Optical Fiber Communication Conference and Exhibit. Technical Digest Postconference Edition (IEEE Cat. 01CH37171)*, Opt. Soc. America, volume 4, pages PD31–1 – PD31–3.

[172] Dingkang T, Jianguo Z, Yuanshan L & Wei Z (2010). Ultrashort optical pulse monitoring using asynchronous optical sampling technique in highly nonlinear fiber. *Chinese Optics Letters*, 8(7):630–633. DOI:10.3788/col20100807.0630.

[173] Antonucci L, Solinas X, Bonvalet A & Joffre M (2012). Asynchronous optical sampling with arbitrary detuning between laser repetition rates. *Optics Express*, 20(16):17928. DOI:10.1364/oe.20.017928.

[174] Hébert NB, Boudreau S, Genest J & Deschênes JD (2014). Coherent dual-comb interferometry with quasi-integer-ratio repetition rates. *Optics Express*, 22(23):29152. DOI:10.1364/oe.22.029152.

[175] Kolner B (1994). Space-time duality and the theory of temporal imaging. *IEEE Journal of Quantum Electronics*, 30(8):1951–1963. DOI:10.1109/3.301659.

[176] Schröder J, Wang F, Clarke A, Ryckeboer E, Pelusi M, Roelens MA & Eggleton BJ (2010). Aberration-free ultra-fast optical oscilloscope using a four-wave mixing based time-lens. *Optics Communications*, 283(12):2611–2614. DOI:10.1016/j.optcom.2010.02.019.

[177] Ryczkowski P, Närhi M, Billet C, Merolla JM, Genty G & Dudley JM (2018). Real-time full-field characterization of transient dissipative soliton dynamics in a mode-locked laser. *Nature Photonics*, 12(4):221–227. DOI:10.1038/s41566-018-0106-7.

[178] Tikan A, Bielawski S, Szwaj C, Randoux S & Suret P (2018). Single-shot measurement of phase and amplitude by using a heterodyne time-lens system and ultrafast digital time-holography. *Nature Photonics*, 12(4):228–234. DOI:10.1038/s41566-018-0113-8.

[179] Foster MA, Salem R, Geraghty DF, Turner-Foster AC, Lipson M & Gaeta AL (2008). Silicon-chip-based ultrafast optical oscilloscope. *Nature*, 456(7218):81–84. DOI:10.1038/nature07430.

[180] Bennett C & Kolner B (2001). Aberrations in temporal imaging. *IEEE Journal of Quantum Electronics*, 37(1):20–32. DOI:10.1109/3.892720.

[181] Weber HG & Nakazawa M (editors) (2007). *Ultrahigh-Speed Optical Transmission Technology*. Weber HG & Nakazawa M (editors). Springer Berlin Heidelberg, Berlin, Heidelberg.

[182] Qiu J, Sun K, Rochette M & Chen LR (2010). Reconfigurable All-Optical Multilogic Gate (xor, and, and or) Based on Cross-Phase Modulation in a Highly Nonlinear Fiber. *IEEE Photonics Technology Letters*, 22(16):1199–1201. DOI:10.1109/LPT.2010.2050876.

[183] Tajima K, Nakamura S & Ueno Y (2002). Semiconductor nonlinearities for ultrafast all-optical gating. *Measurement Science and Technology*, 13(11):1692–1697. DOI:10.1088/0957-0233/13/11/305.

[184] Soto H, Diaz C, Topomondzo J, Erasme D, Schares L & Guekos G (2002). All-optical AND gate implementation using cross-polarization modulation in a semiconductor optical amplifier. *IEEE Photonics Technology Letters*, 14(4):498–500. DOI:10.1109/68.992590.

[185] Kim JH, Kim BC, Byun YT, Jhon YM, Lee S, Woo DH & Kim SH (2004). All-Optical AND Gate Using Cross-Gain Modulation in Semiconductor Optical Amplifiers. *Japanese Journal of Applied Physics*, 43(2):608–610. DOI:10.1143/JJAP.43.608.

[186] Berrettini G, Simi A, Malacarne A, Bogoni A & Poti L (2006). Ultrafast integrable and reconfigurable XNOR, AND, NOR, and NOT photonic logic gate. *IEEE Photonics Technology Letters*, 18(8):917–919. DOI:10.1109/LPT.2006.873570.

[187] Kotb A (2013). AND gate based on two-photon absorption in semiconductor optical amplifier. *Optoelectronics Letters*, 9(3):181–184. DOI:10.1007/s11801-013-2418-7.

[188] OFS Fitel Denmark ApS (2014). *Datasheet: Highly Nonlinear Fiber ZDW1546*.

[189] Engelbrecht R (2014). *Nichtlineare Faseroptik*. Springer Berlin Heidelberg, Berlin, Heidelberg, 540 pages.

[190] Agrawal G (2010). *Applications of Nonlinear Fiber Optics*. 528 pages.

[191] Agrawal G (2012). *Nonlinear fiber optics*. Academic Press, fifth edition.

[192] Shmaliy Y (2006). *Continuous-time signals*. Springer Netherlands, Dordrecht, 355 pages.

[193] Alegre J, Celma S, Aldea C & Calvo B (2006). Fast-Settling Envelope Detectors. *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings*, IEEE, pages 926–929.

[194] Alexiev C, Mak JCC, Sacher WD & Poon JKS (2021). Calibrating rectangular interferometer meshes with external photodetectors. *OSA Continuum*, 4(11):2892. DOI:10.1364/OSAC.437918.

[195] Eberhart RC & Shi Y (1998). Comparison between genetic algorithms and particle swarm optimization. *Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science*. Porto V, Saravanan N, Waagen D & Eiben A (editors), Springer Berlin Heidelberg, volume 1447, pages 611–616, San Diego.

[196] Hassan R, Cohanim B, de Weck O & Venter G (2005). A Comparison of Particle Swarm Optimization and the Genetic Algorithm. *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, American Institute of Aeronautics and Astronautics, Reston, Virigina.

[197] Digani J, Hon PW & Davoyan AR (2022). Framework for Expediting Discovery of Optimal Solutions with Blackbox Algorithms in Non-Topology Photonic Inverse Design. *ACS Photonics*, 9(2):432–442. DOI:10.1021/acsphotonics.1c01819.

[198] Wetzel B, Kues M, Roztocki P, Reimer C, Godin PL, Rowley M, Little BE, Chu ST, Viktorov EA, Moss DJ, Pasquazi A, Peccianti M & Morandotti R (2018). Customizing supercontinuum generation via on-chip adaptive temporal pulse-splitting. *Nature Communications*, 9(1). DOI:10.1038/s41467-018-07141-w.

[199] Hoang VT, Boussafa Y, Sader L, Février S, Couderc V & Wetzel B (2022). Optimizing supercontinuum spectro-temporal properties by leveraging machine learning towards multi-photon microscopy. *Frontiers in Photonics*, 3. DOI:10.3389/fphot.2022.940902.

[200] Feehan JS, Yoffe SR, Brunetti E, Ryser M & Jaroszynski DA (2022). Computer-automated design of mode-locked fiber lasers. *Optics Express*, 30(3):3455. DOI:10.1364/OE.450059.

[201] Yao J & Al-Dahle A (2019). Coarse-to-fine Optimization for Speech Enhancement. *arXiv*.

[202] Chemnitz M, Baumgartl M, Meyer T, Jauregui C, Dietzek B, Popp J, Limpert J & Tünnermann A (2012). Widely tuneable fiber optical parametric amplifier for coherent anti-Stokes Raman scattering microscopy. *Optics Express*, 20(24):26583. DOI:10.1364/OE.20.026583.

[203] Chemnitz M, Scheibinger R, Gaida C, Gebhardt M, Stutzki F, Pumpe S, Kobelke J, Tünnermann A, Limpert J & Schmidt MA (2018). Thermodynamic control of soliton dynamics in liquid-core fibers. *Optica*, 5(6):695. DOI:10.1364/OPTICA.5.000695.

[204] Chemnitz M, Perron N, Fischer B, Junaid S, Schmidt MA & Morandotti R (2021). Towards telecom-compatible liquid core fibers for low-power nonlinear signal processing. *OSA Advanced Photonics Congress*, NoW3C.6 pages.

[205] Dietrich PI, Blaicher M, Reuter I, Billah M, Hoose T, Hofmann A, Caer C, Dangel R, Offrein B, Troppenz U, Moehrle M, Freude W & Koos C (2018). In situ 3D nanoprinting of free-form coupling elements for hybrid photonic integration. *Nature Photonics*, 12(4):241–247. DOI:10.1038/s41566-018-0133-4.

[206] Kum H, Lee D, Kong W, Kim H, Park Y, Kim Y, Baek Y, Bae SH, Lee K & Kim J (2019). Epitaxial growth and layer-transfer techniques for heterogeneous integration of materials for electronic and photonic devices. *Nature Electronics*, 2(10):439–450. DOI:10.1038/s41928-019-0314-2.

[207] Kaur P, Boes A, Ren G, Nguyen TG, Roelkens G & Mitchell A (2021). Hybrid and heterogeneous photonic integration. *APL Photonics*, 6(6):061102. DOI:10.1063/5.0052700.

[208] Stern B, Ji X, Okawachi Y, Gaeta AL & Lipson M (2018). Battery-operated integrated frequency comb generator. *Nature*, 562(7727):401–405. DOI:10.1038/s41586-018-0598-9.

[209] Shen B, Chang L, Liu J, Wang H, Yang QF, Xiang C, Wang RN, He J, Liu T, Xie W, Guo J, Kinghorn D, Wu L, Ji QX, Kippenberg TJ, Vahala K & Bowers JE (2020). Integrated turnkey soliton microcombs. *Nature*, 582(7812):365–369. DOI:10.1038/s41586-020-2358-x.

[210] Pasquazi A, Peccianti M, Park Y, Little BE, Chu ST, Morandotti R, Azaña J & Moss DJ (2011). Sub-picosecond phase-sensitive optical pulse characterization on a chip. *Nature Photonics*, 5(10):618–623. DOI:10.1038/nphoton.2011.199.

[211] Zhang J, Cai J, Meng Y & Meng T (2014). Genetic algorithm particle swarm optimization based hardware evolution strategy. *WSEAS Transactions on Circuits and Systems*, 13:274–283.

[212] López A, Pérez D, DasMahapatra P & Capmany J (2020). Auto-routing algorithm for field-programmable photonic gate arrays. *Optics Express*, 28(1):737. DOI:10.1364/OE.382753.

[213] Li Z, Yang S, Xiao Q, Zhang T, Li Y, Han L, Liu D, Ouyang X & Zhu J (2022). Deep reinforcement with spectrum series learning control for a mode-locked fiber laser. *Photonics Research*, 10(6):1491. DOI:10.1364/PRJ.455493.

[214] Kuprikov E, Kokhanovskiy A, Serebrennikov K & Turitsyn S (2022). Deep reinforcement learning for self-tuning laser source of dissipative solitons. *Scientific Reports*, 12(1):7185. DOI:10.1038/s41598-022-11274-w.

[215] LeCun Y, Bengio Y & Hinton G (2015). Deep learning. *Nature*, 521(7553):436–444. DOI:10.1038/nature14539.

[216] Schmidhuber J (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61:85–117. DOI:10.1016/j.neunet.2014.09.003.

[217] Qiu X, Sun T, Xu Y, Shao Y, Dai N & Huang X (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897. DOI:10.1007/s11431-020-1647-3.

[218] McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GC, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J & Shetty S (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788):89–94. DOI:10.1038/s41586-019-1799-6.

[219] Aradi S (2020). Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–20. DOI:10.1109/TITS.2020.3024655.

[220] Hopfield JJ (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558. DOI:10.1073/pnas.79.8.2554.

[221] Williams RJ & Zipser D (1989). A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, 1(2):270–280. DOI:10.1162/neco.1989.1.2.270.

[222] Connor J, Martin R & Atlas L (1994). Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2):240–254. DOI:10.1109/72.279188.

[223] Tong Z & Tanaka G (2018). Reservoir Computing with Untrained Convolutional Neural Networks for Image Recognition. *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE, pages 1289–1294.

[224] An Q, Bai K, Liu L, Shen F & Yi Y (2020). A unified information perceptron using deep reservoir computing. *Computers & Electrical Engineering*, 85:106705. DOI:10.1016/j.compeleceng.2020.106705.

[225] Gallicchio C, Micheli A & Pedrelli L (2017). Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 268:87–99. DOI:10.1016/j.neucom.2016.12.089.

[226] Ming Liang & Xiaolin Hu (2015). Recurrent convolutional neural network for object recognition. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pages 3367–3375.

[227] Gu Y, Chen Y, Liu J & Jiang X (2015). Semi-supervised deep extreme learning machine for Wi-Fi based localization. *Neurocomputing*, 166:282–293. DOI:10.1016/j.neucom.2015.04.011.

[228] Sinaga KP & Yang MS (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8:80716–80727. DOI:10.1109/ACCESS.2020.2988796.

[229] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T & Hassabis D (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489. DOI:10.1038/nature16961.

[230] Wurman PR, Barrett S, Kawamoto K, MacGlashan J, Subramanian K, Walsh TJ, Capobianco R, Devlic A, Eckert F, Fuchs F, Gilpin L, Khandelwal P, Kompella V, Lin H, MacAlpine P, Oller D, Seno T, Sherstan C, Thomure MD, Aghabozorgi H, Barrett L, Douglas R, Whitehead D, Dürr P, Stone P, Spranger M & Kitano H (2022). Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature*, 602(7896):223–228. DOI:10.1038/s41586-021-04357-7.

[231] Burkitt AN (2006). A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics*, 95(1):1–19. DOI:10.1007/s00422-006-0068-6.

[232] Bishop CM (2007). *Machine Learning and Pattern Recoginiton*. Springer Science and Business Media LLC.

[233] Glorot X, Bordes A & Bengio Y (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315—-323.

[234] Van Vaerenbergh T, Fiers M, Mechet P, Spuesens T, Kumar R, Morthier G, Schrauwen B, Dambre J & Bienstman P (2012). Cascadable excitability in microrings. *Optics Express*, 20(18):20292. DOI:10.1364/oe.20.020292.

[235] Maass W (2015). To Spike or Not to Spike: That Is the Question. *Proceedings of the IEEE*, 103(12):2219–2224. DOI:10.1109/JPROC.2015.2496679.

[236] Peng HT, Nahmias MA, De Lima TF, Tait AN, Shastri BJ & Prucnal PR (2018). Neuromorphic Photonic Integrated Circuits. *IEEE Journal of Selected Topics in Quantum Electronics*, 24(6):1–16. DOI:10.1109/JSTQE.2018.2840448.

[237] Choudhary T, Mishra V, Goswami A & Sarangapani J (2020). A comprehensive survey on model compression and acceleration. *Artificial Intelligence Review*, 53(7):5113–5155. DOI:10.1007/s10462-020-09816-7.

[238] Liang T, Glossner J, Wang L, Shi S & Zhang X (2021). Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing*, 461:370–403. DOI:10.1016/j.neucom.2021.07.045.

[239] Sze V, Chen YH, Yang TJ & Emer JS (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, 105(12):2295–2329. DOI:10.1109/JPROC.2017.2761740.

[240] Bartol TM, Bromer C, Kinney J, Chirillo MA, Bourne JN, Harris KM & Sejnowski TJ (2015). Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife*, 4. DOI:10.7554/eLife.10778.

[241] Cotman CW & Sampedro MN (1982). Brain Function, Synapse Renewal, and Plasticity. *Annual Review of Psychology*, 33(1):371–401. DOI:10.1146/annurev.ps.33.020182.002103.

[242] Bradke F & Dotti CG (2000). Differentiated neurons retain the capacity to generate axons from dendrites. *Current Biology*, 10(22):1467–1470. DOI:10.1016/S0960-9822(00)00807-1.

[243] Zucker RS & Regehr WG (2002). Short-Term Synaptic Plasticity. *Annual Review of Physiology*, 64(1):355–405. DOI:10.1146/annurev.physiol.64.092501.114547.

[244] Mongillo G, Barak O & Tsodyks M (2008). Synaptic Theory of Working Memory. *Science*, 319(5869):1543–1546. DOI:10.1126/science.1150769.

[245] Poirazi P & Mel BW (2001). Impact of Active Dendrites and Structural Plasticity on the Memory Capacity of Neural Tissue. *Neuron*, 29(3):779–796. DOI:10.1016/S0896-6273(01)00252-5.

[246] Xu Nl, Harnett MT, Williams SR, Huber D, O'Connor DH, Svoboda K & Magee JC (2012). Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature*, 492(7428):247–251. DOI:10.1038/nature11601.

[247] Hsu J (2014). IBM's new brain [News]. *IEEE Spectrum*, 51(10):17–19. DOI:10.1109/MSPEC.2014.6905473.

[248] Herculano-Houzel S (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3. DOI:10.3389/neuro.09.031.2009.

[249] Herculano-Houzel S (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, 109(Supplement_1):10661–10668. DOI:10.1073/pnas.1201895109.

[250] Furber S (2012). To build a brain. *IEEE Spectrum*, 49(8):44–49. DOI:10.1109/MSPEC.2012.6247562.

[251] Yamazaki T, Igarashi J & Yamaura H (2021). Human-scale Brain Simulation via Supercomputer: A Case Study on the Cerebellum. *Neuroscience*, 462:235–246. DOI:10.1016/j.neuroscience.2021.01.014.

[252] Kagan BJ, Kitchen AC, Tran NT, Habibollahi F, Khajehnejad M, Parker BJ, Bhat A, Rollo B, Razi A & Friston KJ (2022). In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron*, 110(23):3952–3969.e8. DOI:10.1016/j.neuron.2022.09.001.

[253] Sandin F, Khan AI, Dyer AG, Amin AHM, Indiveri G, Chicca E & Osipov E (2014). Concept Learning in Neuromorphic Vision Systems: What Can We Learn from Insects? *Journal of Software Engineering and Applications*, 07(05):387–395. DOI:10.4236/jsea.2014.75035.

[254] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J & Houlsby N (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv*, 2010.11929 pages.

[255] Denisov P & Vu NT (2019). End-to-End Multi-Speaker Speech Recognition using Speaker Embeddings and Transfer Learning. *arXiv*, 1908.04737 pages.

[256] Rosenblatt F (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408. DOI:10.1037/h0042519.

[257] Rosenblatt F (1960). Perceptron Simulation Experiments. *Proceedings of the IRE*, 48(3):301–309. DOI:10.1109/JRPROC.1960.287598.

[258] Koçak Y & Üstündağ Şiray G (2021). New activation functions for single layer feedforward neural network. *Expert Systems with Applications*, 164:113977. DOI:10.1016/j.eswa.2020.113977.

[259] Stork DG & Allen JD (1992). How to solve the N-bit parity problem with two hidden units. *Neural Networks*, 5(6):923–926. DOI:10.1016/S0893-6080(05)80088-7.

[260] Minsky M & Papert SA (1969). *Perceptrons: An Introduction to Computational Geometry*. The MIT Press.

[261] Rumelhart DE, Hinton G & Williams RJ (1985). Learning internal representations by error propagation. California University San Diego La Jolla Instititue for Cognitive Science. San Diego.

[262] Rumelhart DE, Hinton GE & Williams RJ (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536. DOI:10.1038/323533a0.

[263] Ji Z & Telgarsky M (2018). Gradient descent aligns the layers of deep linear networks. *arXiv*, 1810.02032 pages.

[264] Advani MS, Saxe AM & Sompolinsky H (2020). High-dimensional dynamics of generalization error in neural networks. *Neural Networks*, 132:428–446. DOI:10.1016/j.neunet.2020.08.022.

[265] Amit I, Matherly J, Hewlett W, Xu Z, Meshi Y & Weinberger Y (2018). Machine Learning in Cyber-Security - Problems, Challenges and Data Sets. *arXiv*, 1812.07858 pages.

[266] Haas H & Jaeger H (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80. DOI:10.1126/science.1091277.

[267] Lukoševičius M & Jaeger H (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149. DOI:10.1016/j.cosrev.2009.03.005.

[268] Huang GB, Zhu QY & Siew CK (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. *IEEE International Conference on Neural Networks*, volume 2, pages 985–990.

[269] LeCun Y (1988). A theoretical framework for back-propagation. *Proceedings of the 1988 Connectionist Models Summer School*. Touretzky D, Hinton G & Sejnowski T (editors), Morgan Kaufmann, pages 21–28, Pittsburg, PA, USA.

[270] Scellier B & Bengio Y (2017). Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation. *Frontiers in Computational Neuroscience*, 11. DOI:10.3389/fncom.2017.00024.

[271] Ernoult M, Grollier J, Querlioz D, Bengio Y & Scellier B (2020). Equilibrium Propagation with Continual Weight Updates. *arXiv*, 2005.04168 pages.

[272] Wilamowski BM & Hao Yu (2010). Neural Network Learning Without Backpropagation. *IEEE Transactions on Neural Networks*, 21(11):1793–1803. DOI:10.1109/TNN.2010.2073482.

[273] Hoerl AE & Kennard RW (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67. DOI:10.2307/1271436.

[274] Stone M (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133. DOI:10.1111/j.2517-6161.1974.tb00994.x.

[275] Koutroumbas K & Theodoridis S (2008). *Pattern Recognition*. Academic Press, fourth edition.

[276] Farhat NH, Psaltis D, Prata A & Paek E (1985). Optical implementation of the Hopfield model. *Applied Optics*, 24(10):1469. DOI:10.1364/AO.24.001469.

[277] Berger C, Collings N & Gehriger D (1998). Recurrent Optical Neural Network for the Study of Pattern Dynamics. *Proc. SPIE 3402, Optical Information Science and Technology (OIST97): Optical Memory and Neural Networks*, pages 233–244.

[278] Lau APT, Khan FN, Fan Q & Lu C (2018). Application of Machine Learning Techniques in Fiber-Optic Communication Systems. *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, OSA, numéro ii, SpW4G.1 pages, Washington, D.C.

[279] Rivenson Y & Ozcan A (2018). Toward a Thinking Microscope. *Optics and Photonics News*, 29(7):34. DOI:10.1364/OPN.29.7.000034.

[280] Rodrigues SP, Yu Z, Schmalenberg P, Lee J, Iizuka H & Dede EM (2021). Weighing in on photonic-based machine learning for automotive mobility. *Nature Photonics*, 15(2):66–67. DOI:10.1038/s41566-020-00736-0.

[281] Wertheimer ZA, Bar C & Levin A (2022). Towards machine learning for heterogeneous inverse scattering in 3D microscopy. *Optics Express*, 30(6):9854. DOI:10.1364/OE.447075.

[282] Zuo C, Qian J, Feng S, Yin W, Li Y, Fan P, Han J, Qian K & Chen Q (2022). Deep learning in optical metrology: a review. *Light: Science & Applications*, 11(1):39. DOI:10.1038/s41377-022-00714-x.

[283] Huang L, Luo R, Liu X & Hao X (2022). Spectral imaging with deep learning. *Light: Science & Applications*, 11(1):61. DOI:10.1038/s41377-022-00743-6.

[284] Liu H, Zhu J, Yin H, Yan Q, Liu H, Guan S, Cai Q, Sun J, Yao S & Wei R (2022). Extreme learning machine and genetic algorithm in quantitative analysis of sulfur hexafluoride by infrared spectroscopy. *Applied Optics*, 61(10):2834. DOI:10.1364/AO.450805.

[285] Zhang H, Lau JWZ, Wan L, Shi L, Shi Y, Cai H, Luo X, Lo G, Lee C, Kwek LC & Liu AQ (2022). Molecular Property Prediction with Photonic Chip-Based Machine Learning. *Laser & Photonics Reviews*, 2200698 pages. DOI:10.1002/lpor.202200698.

[286] Bartolozzi C, Indiveri G & Donati E (2022). Embodied neuromorphic intelligence. *Nature Communications*, 13(1):1024. DOI:10.1038/s41467-022-28487-2.

[287] Zuo Y, Cao C, Cao N, Lai X, Zeng B & Du S (2022). Optical neural network quantum state tomography. *Advanced Photonics*, 4(02). DOI:10.1117/1.AP.4.2.026004.

[288] Lim J & Psaltis D (2022). MaxwellNet: Physics-driven deep neural network training based on Maxwell's equations. *APL Photonics*, 7(1):011301. DOI:10.1063/5.0071616.

[289] Mandic D & Chambers J (2001). *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley, Chichester.

[290] Berger C & Collings N (1998). All-optical recurrent neural network. *Proc. SPIE 3490, Optics in Computing '98*, 465 pages.

[291] Vandoorne K, Dierckx W, Schrauwen B, Verstraeten D, Baets R, Bienstman P & Campenhout JV (2008). Toward optical signal processing using Photonic Reservoir Computing. *Optics Express*, 16(15):11182. DOI:10.1364/OE.16.011182.

[292] Dupont S, Petersen C, Thøgersen J, Agger C, Bang O & Keiding SR (2012). IR microscopy utilizing intense supercontinuum light source. *Optics Express*, 20(5):4887. DOI:10.1364/OE.20.004887.

[293] Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D & Soljačić M (2017). Deep learning with coherent nanophotonic circuits. *Nature Photonics*, 11(7):441–446. DOI:10.1038/nphoton.2017.93.

[294] Zhang H, Gu M, Jiang XD, Thompson J, Cai H, Paesani S, Santagati R, Laing A, Zhang Y, Yung MH, Shi YZ, Muhammad FK, Lo GQ, Luo XS, Dong B, Kwong DL, Kwek LC & Liu AQ (2021). An optical neural chip for implementing complex-valued neural network. *Nature Communications*, 12(1):457. DOI:10.1038/s41467-020-20719-7.

[295] Feldmann J, Youngblood N, Wright CD, Bhaskaran H & Pernice WHP (2019). All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature*, 569(7755):208–214. DOI:10.1038/s41586-019-1157-8.

[296] Moughames J, Porte X, Thiel M, Ulliac G, Larger L, Jacquot M, Kadic M & Brunner D (2020). Three-dimensional waveguide interconnects for scalable integration of photonic neural networks. *Optica*, 7(6):640. DOI:10.1364/OPTICA.388205.

[297] Appeltant L, Soriano MC, Van der Sande G, Danckaert J, Massar S, Dambre J, Schrauwen B, Mirasso CR & Fischer I (2011). Information processing using a single dynamical node as complex system. *Nature Communications*, 2(1):468. DOI:10.1038/ncomms1476.

[298] Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George J, Kildishev AV, Pelton M & Sorger VJ (2018). All-optical nonlinear activation function for photonic neural networks [Invited]. *Optical Materials Express*, 8(12):3851. DOI:10.1364/OME.8.003851.

[299] Pour Fard MM, Williamson IAD, Edwards M, Liu K, Pai S, Bartlett B, Minkov M, Hughes TW, Fan S & Nguyen TA (2020). Experimental realization of arbitrary activation functions for optical neural networks. *Optics Express*, 28(8):12138. DOI:10.1364/OE.391473.

[300] Wu B, Li H, Tong W, Dong J & Zhang X (2022). Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator. *Optical Materials Express*, 12(3):970. DOI:10.1364/OME.447330.

[301] Al-Qadasi MA, Chrostowski L, Shastri BJ & Shekhar S (2022). Scaling up silicon photonic-based accelerators: Challenges and opportunities. *APL Photonics*, 7(2):020902. DOI:10.1063/5.0070992.

[302] Hughes TW, Williamson IAD, Minkov M & Fan S (2019). Wave physics as an analog recurrent neural network. *Science Advances*, 5(12):eaay6946. DOI:10.1126/sciadv.aay6946.

[303] Zangeneh-Nejad F, Sounas DL, Alù A & Fleury R (2021). Analogue computing with metamaterials. *Nature Reviews Materials*, 6(3):207–225. DOI:10.1038/s41578-020-00243-2.

[304] Zhou T, Scalzo F & Jalali B (2022). Nonlinear Schrödinger Kernel for Hardware Acceleration of Machine Learning. *Journal of Lightwave Technology*, 40(5):1308–1319. DOI:10.1109/JLT.2022.3146131.

[305] Zhou T, Fang L, Yan T, Wu J, Li Y, Fan J, Wu H, Lin X & Dai Q (2020). In situ optical backpropagation training of diffractive optical neural networks. *Photonics Research*, 8(6):940. DOI:10.1364/PRJ.389553.

[306] Nakajima M, Tanaka K & Hashimoto T (2021). Neural Schrödinger Equation: Physical Law as Deep Neural Network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15. DOI:10.1109/TNNLS.2021.3120472.

[307] Hager C & Pfister HD (2021). Physics-Based Deep Learning for Fiber-Optic Communication Systems. *IEEE Journal on Selected Areas in Communications*, 39(1):280–294. DOI:10.1109/JSAC.2020.3036950.

[308] Wright LG, Onodera T, Stein MM, Wang T, Schachter DT, Hu Z & McMahon PL (2022). Deep physical neural networks trained with backpropagation. *Nature*, 601(7894):549–555. DOI:10.1038/s41586-021-04223-6.

[309] Nakajima M, Tanaka K & Hashimoto T (2021). Neural Schrödinger Equation: Physical Law as Deep Neural Network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15. DOI:10.1109/TNNLS.2021.3120472.

[310] Sunada S & Uchida A (2021). Photonic neural field on a silicon chip: large-scale, high-speed neuro-inspired computing and sensing. *Optica*, 8(11):1388. DOI:10.1364/OPTICA.434918.

[311] Michaeli L & Bahabad A (2018). Genetic algorithm driven spectral shaping of supercontinuum radiation in a photonic crystal fiber. *Journal of Optics (United Kingdom)*, 20(5):55501. DOI:10.1088/2040-8986/aab59c.

[312] Husakou AV & Herrmann J (2001). Supercontinuum generation of higher-order solitons by fission in photonic crystal fibers. *Physical Review Letters*, 87(20):203901. DOI:10.1103/PhysRevLett.87.203901.

[313] Hasegawa A & Matsumoto M (2003). *Optical solitons in fibers*. http://dx.doi.org/10.1007/978-3-540-46064-0_5.

[314] Mamyshev P, Chernikov S & Dianov E (1991). Generation of fundamental soliton trains for high-bit-rate optical fiber communication lines. *IEEE Journal of Quantum Electronics*, 27(10): 2347–2355. DOI:10.1109/3.97280.

[315] Ghadi A (2020). Phase sensitive, all-optical and self-integrated multi-logic AND, OR, XOR, and NOT gates. *Physics Letters A*, 384(22):126432. DOI:10.1016/j.physleta.2020.126432.

[316] Ghadi A (2021). All-optical computing circuits half-subtractor and comparator based on soliton interactions. *Optik*, 227:166079. DOI:10.1016/j.ijleo.2020.166079.

[317] Silva NA, Ferreira TD & Guerreiro A (2021). Reservoir computing with solitons. *New Journal of Physics*, 23(2):023013. DOI:10.1088/1367-2630/abda84.

[318] Stolen RH, Tomlinson WJ & Mollenauer LF (1983). Observation of pulse restoration at the soliton period in optical fibers. *Optics Letters*, 8(3):186. DOI:10.1364/OL.8.000186.

[319] Dudley JM, Genty G & Coen S (2006). Supercontinuum generation in photonic crystal fiber. *Reviews of Modern Physics*, 78(4):1135–1184. DOI:10.1103/RevModPhys.78.1135.

[320] Alfano RR & Shapiro SL (1970). Emission in the region 4000 to 7000 Åvia four-photon coupling in glass. *Physical Review Letters*, 24(11):584–587. DOI:10.1103/PhysRevLett.24.584.

[321] Gorbach AV & Skryabin DV (2007a). Theory of radiation trapping by the accelerating solitons in optical fibers. *Physical Review A*, 76(5):053803. DOI:10.1103/PhysRevA.76.053803.

[322] Gorbach AV & Skryabin DV (2007b). Light trapping in gravity-like potentials and expansion of supercontinuum spectra in photonic-crystal fibres. *Nature Photonics*, 1(11):653–657. DOI:10.1038/nphoton.2007.202.

[323] Udem T, Holzwarth R & Hänsch TW (2002). Optical frequency metrology. *Nature*, 416(6877): 233–237. DOI:10.1038/416233a.

[324] Inoue T & Namiki S (2008). Pulse compression techniques using highly nonlinear fibers. *Laser & Photonics Review*, 2(1-2):83–99. DOI:10.1002/lpor.200710029.

[325] Kuyken B, Billet M, Leo F, Yvind K & Pu M (2020). Octave-spanning coherent supercontinuum generation in an AlGaAs-on-insulator waveguide. *Optics Letters*, 45(3):603. DOI:10.1364/OL.45.000603.

[326] Chemnitz M (2019). *Soliton dynamics in liquid-core optical fibers*. Phd thesis, Friedrich-Schiller-Universitat Jena.

[327] Heidt A (2009). Efficient Adaptive Step Size Method for the Simulation of Supercontinuum Generation in Optical Fibers. *Journal of Lightwave Technology*, 27(18):3984–3991. DOI:10.1109/JLT.2009.2021538.

[328] Dudley JM & Coen S (2002). Coherence properties of supercontinuum spectra generated in photonic crystal and tapered optical fibers. *Optics Letters*, 27(13):1180. DOI:10.1364/OL.27.001180.

[329] Roelens M, Frisken S, Bolger J, Abakoumov D, Baxter G, Poole S & Eggleton B (2008). Dispersion Trimming in a Reconfigurable Wavelength Selective Switch. *Journal of Lightwave Technology*, 26(1):73–78. DOI:10.1109/JLT.2007.912148.

[330] Hughes TW, Minkov M, Shi Y & Fan S (2018). Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica*, 5(7):864. DOI:10.1364/OPTICA.5.000864.

[331] Guo X, Barrett TD, Wang ZM & Lvovsky AI (2021). Backpropagation through nonlinear units for the all-optical training of neural networks. *Photonics Research*, 9(3):B71. DOI:10.1364/PRJ.411104.

[332] Pai S, Sun Z, Hughes TW, Park T, Bartlett B, Williamson IAD, Minkov M, Milanizadeh M, Abebe N, Morichetti F, Melloni A, Fan S, Solgaard O & Miller DAB (2022). Experimentally realized in situ backpropagation for deep learning in nanophotonic neural networks. *arXiv*, 2205.08501 pages.

[333] Spall J, Guo X & Lvovsky AI (2022). Hybrid training of optical neural networks. *Optica*, 9(7):803. DOI:10.1364/OPTICA.456108.

[334] Zhang T, Wang J, Dan Y, Lanqiu Y, Dai J, Han X, Sun X & Xu K (2019). Efficient training and design of photonic neural network through neuroevolution. *Optics Express*, 27(26):37150. DOI:10.1364/OE.27.037150.

[335] Shao R, Zhang G & Gong X (2022). Generalized robust training scheme using genetic algorithm for optical neural networks with imprecise components. *Photonics Research*, 10(8):1868. DOI:10.1364/PRJ.449570.

[336] Brunner D, Penkovsky B, Marquez BA, Jacquot M, Fischer I & Larger L (2018). Tutorial: Photonic neural networks in delay systems. *Journal of Applied Physics*, 124(15):152004. DOI:10.1063/1.5042342.

[337] Tanaka G, Yamane T, Héroux JB, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D & Hirose A (2019). Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123. DOI:10.1016/j.neunet.2019.03.005.

[338] Lawson CL & Hanson RJ (1995). *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1 edition, 349 pages.

[339] Rashleigh S (1983). Origins and control of polarization effects in single-mode fibers. *Journal of Lightwave Technology*, 1(2):312–331. DOI:10.1109/JLT.1983.1072121.

[340] Loranger S & Kashyap R (2017). Are optical fibers really uniform? Measurement of refractive index on a centimeter scale. *Optics Letters*, 42(9):1832. DOI:10.1364/OL.42.001832.

[341] Yin H & Yang Ya (2017). Online transfer learning with extreme learning machine. *AIP Conference Proceedings 1839*, 020199 pages.

[342] Buckley SM, Tait AN, McCaughan AN & Shastri BJ (2023). Photonic online learning: a perspective. *Nanophotonics*, 12(5):833–845. DOI:10.1515/nanoph-2022-0553.

[343] Setiono R (1997). On the solution of the parity problem by a single hidden layer feedforward neural network. *Neurocomputing*, 16(3):225–235. DOI:10.1016/S0925-2312(97)00030-1.

[344] Hohil ME, Liu D & Smith SH (1999). Solving the N-bit parity problem using neural networks. *Neural Networks*, 12(9):1321–1323. DOI:10.1016/S0893-6080(99)00069-6.

[345] Hornik K, Stinchcombe M & White H (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5):551–560. DOI:10.1016/0893-6080(90)90005-6.

[346] Fisher RA (1988). *IRIS Flower Dataset*. https://archive-beta.ics.uci.edu/. 2020-06-03.

[347] Forina M, Leardi R, Armanino C & Lanteri S (1988). PARVUS - an extendible package for data exploration, classification and correlation. Institute of Pharmaceutical and Food Analysis and Technologies, Via Brigata Salerno, 16147 Genoa, Italy.

[348] Warwick N, Sellers T, Talbot S, Cawthorn A & Ford W (1994). The Population Biology of Abalone (_Haliotis_ species) in Tasmania. I. Blacklip Abalone (_H. rubra_) from the North Coast and Islands of Bass Strait. Sea Fisheries Division, Technical Report No. 48 (ISSN 1034-3288).

[349] LeCun, Yann, Jackel L, Bottou L, Cortes C, Denker J, Drucker H, Guyon I, Muller U, Sackinger E, Simard P & Vapnik V (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural Networks*, pages 261–276.

[350] Liu D, Hohil ME & Smith SH (2002). N-bit parity neural networks: new solutions based on linear programming. *Neurocomputing*, 48(1-4):477–488. DOI:10.1016/S0925-2312(01)00612-9.

[351] Lupo A, Butschek L & Massar S (2021). Photonic extreme learning machine based on frequency multiplexing. *Optics Express*, 29(18):28257. DOI:10.1364/OE.433535.

[352] Hornik K, Stinchcombe M & White H (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366. DOI:10.1016/0893-6080(89)90020-8.

[353] Huang GB, Zhu QY & Siew CK (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501. DOI:10.1016/j.neucom.2005.12.126.

[354] Minor J (1993). Parity with two layer feedforward nets. *Neural Networks*, 6(5):705–707. DOI:10.1016/S0893-6080(05)80114-5.

[355] Boscolo S, Dudley JM & Finot C (2021). Modelling self-similar parabolic pulses in optical fibres with a neural network. *Results in Optics*, 3:100066. DOI:10.1016/j.rio.2021.100066.

[356] Lupo A, Butschek L & Massar S (2021). Photonic extreme learning machine based on frequency multiplexing. *Optics Express*, 29(18):28257. DOI:10.1364/OE.433535.

[357] Lupo A & Massar S (2021). Parallel Extreme Learning Machines Based on Frequency Multiplexing. *Applied Sciences*, 12(1):214. DOI:10.3390/app12010214.

[358] Shi B, Calabretta N & Stabile R (2020). Deep Neural Network Through an InP SOA-Based Photonic Integrated Cross-Connect. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(1):1–11. DOI:10.1109/JSTQE.2019.2945548.

[359] LeCun Y, Bottou L, Bengio Y & Haffner P (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324. DOI:10.1109/5.726791.

[360] Antonik P, Marsal N & Rontani D (2020). Large-Scale Spatiotemporal Photonic Reservoir Computer for Image Classification. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(1):1–12. DOI:10.1109/JSTQE.2019.2924138.

[361] Junior OL, Delgado D, Goncalves V & Nunes U (2009). Trainable classifier-fusion schemes: An application to pedestrian detection. *2009 12th International IEEE Conference on Intelligent Transportation Systems*, IEEE, pages 1–6.

[362] Hirschberg J, Hjalmarsson A & Elhadad N (2010). "You're as Sick as You Sound": Using Computational Approaches for Modeling Speaker State to Gauge Illness and Recovery. *Advances in Speech Recognition*, Springer US, pages 305–322. Boston, MA.

[363] Schuller BW, Batliner A, Bergler C, Mascolo C, Han J, Lefter I, Kaya H, Amiriparian S, Baird A, Stappen L, Ottl S, Gerczuk M, Tzirakis P, Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Spathis D, Xia T, Cicuta P, Rothkrantz LJM, Zwerts J, Treep J & Kaandorp C (2021). The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primates. *arXiv*, 2102.13468 pages.

[364] Sharma NK, Muguli A, Krishnan P, Kumar R, Chetupalli SR & Ganapathy S (2022). Towards sound based testing of COVID-19—Summary of the first Diagnostics of COVID-19 using Acoustics (DiCOVA) Challenge. *Computer Speech & Language*, 73:101320. DOI:10.1016/j.csl.2021.101320.

[365] Han J, Xia T, Spathis D, Bondareva E, Brown C, Chauhan J, Dang T, Grammenos A, Hasthanasombat A, Floto A, Cicuta P & Mascolo C (2022). Sounds of COVID-19: exploring realistic performance of audio-based digital testing. *npj Digital Medicine*, 5(1):16. DOI:10.1038/s41746-021-00553-x.

[366] Coppock H, Akman A, Bergler C, Gerczuk M, Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Spathis D, Xia T, Cicuta P, Han J, Amiriparian S, Baird A, Stappen L, Ottl S, Tzirakis P, Batliner A, Mascolo C & Schuller BW (2022). A Summary of the ComParE COVID-19 Challenges. *arXiv*, 2202.08981 pages.

[367] Zhu Y & Falk TH (2022). Fusion of Modulation Spectral and Spectral Features with Symptom Metadata for Improved Speech-Based Covid-19 Detection. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pages 8997–9001.

[368] Pierleoni P, Pernini L, Belli A, Palma L, Valenti S & Paniccia M (2015). SVM-based fall detection method for elderly people using Android low-cost smartphones. *2015 IEEE Sensors Applications Symposium (SAS)*, IEEE, pages 1–5.

[369] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M & Duchesnay É (2012). Scikit-learn: Machine Learning in Python. *arXiv*, 1201.0490 pages.

[370] Li X & Mills M (2019). Vocal Features: From Voice Identification to Speech Recognition by Machine. *Technology and Culture*, 60(2S):S129–S160. DOI:10.1353/tech.2019.0066.

[371] Badshah AM, Ahmad J, Rahim N & Baik SW (2017). Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. *2017 International Conference on Platform Technology and Service (PlatCon)*, IEEE, pages 1–5.

[372] Larger L, Baylón-Fuentes A, Martinenghi R, Udaltsov VS, Chembo YK & Jacquot M (2017). High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification. *Physical Review X*, 7(1):1–14. DOI:10.1103/PhysRevX.7.011015.

[373] Eyben F, Wöllmer M & Schuller B (2010). Opensmile. *Proceedings of the international conference on Multimedia - MM '10*, ACM Press, 1459 pages, New York, New York, USA.

[374] Falk TH, Chan WY, Sejdic E & Chau T (2010). Spectro-Temporal Analysis of Auscultatory Sounds. *New Development in Biomedical Engineering*. Campolo D, editor, IntechOpen, chapitre 5, pages 93–104.

[375] Pearson K (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572. DOI:10.1080/14786440109462720.

[376] Wold S, Esbensen K & Geladi P (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3):37–52. DOI:10.1016/0169-7439(87)80084-9.

[377] Forrest S (1993). Genetic algorithms: principles of natural selection applied to computation. *Science*, 261(5123):872–878. DOI:10.1126/science.8346439.

[378] Cheng Y, Wang D, Zhou P & Zhang T (2017). A Survey of Model Compression and Acceleration for Deep Neural Networks. *arXiv*, 1710.09282 pages.

[379] Akyol K (2020). Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Systems with Applications*, 140:112875. DOI:10.1016/j.eswa.2019.112875.

[380] Montana DJ & Davis L (1989). Training feedforward neural networks using genetic algorithms. *IJCAI'89: Proceedings of the 11th international joint conference on Artificial intelligence*, pages 762–767.

[381] Eberhart R & Kennedy J (1995). A new optimizer using particle swarm theory. *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, IEEE, pages 39–43.

[382] Yao P & Wang XZ (2018). Weight Learning in Weighted ELM Classification Model Based on Genetic Algorithms. *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, IEEE, pages 370–377.

[383] Mousavirad SJ, Schaefer G, Jalali SMJ & Korovin I (2020). A benchmark of recent population-based metaheuristic algorithms for multi-layer neural network training. *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, ACM, pages 1402–1408, New York, NY, USA.

[384] Hieu Trung Huynh & Yonggwan Won (2008). Evolutionary algorithm for training compact single hidden layer feedforward neural networks. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, IEEE, pages 3028–3033.

[385] Madry A, Makelov A, Schmidt L, Tsipras D & Vladu A (2017). Towards Deep Learning Models Resistant to Adversarial Attacks. *arXiv*, 1706.06083 pages.

[386] Su J, Vargas DV & Sakurai K (2019). One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841. DOI:10.1109/TEVC.2019.2890858.

[387] Jiao S, Song Z & Xiang S (2023). Adversarial Attacks on an Optical Neural Network. *IEEE Journal of Selected Topics in Quantum Electronics*, 29(2: Optical Computing):1–6. DOI:10.1109/JSTQE.2022.3207056.

[388] Li B, Wei Z, Wu J, Yu S, Zhang T, Zhu C, Zheng D, Guo W, Zhao C & Zhang J (2023). Machine Learning-enabled Globally Guaranteed Evolutionary Computation. *Research Square (preprint)*. DOI:https://doi.org/10.21203/rs.3.rs-2156781/v1.

[389] Brunner D, Soriano MC, Mirasso CR & Fischer I (2013). Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature Communications*, 4(1):1364. DOI:10.1038/ncomms2368.

[390] Szabo AD, Ribeiro V, Gaur CB, Ali AA, Mussot A, Quiquempois Y, Bouwmans G & Doran NJ (2021). Dual-Polarization C+L-Band Wavelength Conversion in a Twin-Core Highly Nonlinear Fibre. *2021 Optical Fiber Communications Conference and Exhibition, OFC 2021 - Proceedings*, pages 30–32.

[391] Cohen E, Malka D, Shemer A, Shahmoon A, Zalevsky Z & London M (2016). Neural networks within multi-core optic fibers. *Scientific Reports*, 6(1):29080. DOI:10.1038/srep29080.

[392] Gan L, Shen L, Tang M, Xing C, Li Y, Ke C, Tong W, Li B, Fu S & Liu D (2018). Investigation of channel model for weakly coupled multicore fiber. *Optics Express*, 26(5):5182. DOI:10.1364/OE.26.005182.

[393] Alves TMF & Cartaxo AVT (2019). Decorrelation Bandwidth of Intercore Crosstalk in Weakly Coupled Multicore Fibers With Multiple Interfering Cores. *Journal of Lightwave Technology*, 37(3):744–754. DOI:10.1109/JLT.2018.2880076.

[394] Mercy Kingsta R & Shantha Selvakumari R (2019). A review on coupled and uncoupled multicore fibers for future ultra-high capacity optical communication. *Optik*, 199:163341. DOI:10.1016/j.ijleo.2019.163341.

[395] Chemnitz M, Gebhardt M, Gaida C, Stutzki F, Kobelke J, Limpert J, Tünnermann A & Schmidt MA (2017). Hybrid soliton dynamics in liquid-core fibres. *Nature Communications*, 8(1):42. DOI:10.1038/s41467-017-00033-5.

[396] Junaid S, Schaarschmidt K, Chemnitz M, Chambonneau M, Nolte S & Schmidt MA (2020). Tailoring modulation instabilities and four-wave mixing in dispersion-managed composite liquid-core fibers. *Optics Express*, 28(3):3097. DOI:10.1364/OE.382639.

[397] Katumba A, Heyvaert J, Schneider B, Uvin S, Dambre J & Bienstman P (2018). Low-Loss Photonic Reservoir Computing with Multimode Photonic Integrated Circuits. *Scientific Reports*, 8(1):1–10. DOI:10.1038/s41598-018-21011-x.

[398] Sun Y, Zitelli M, Ferraro M, Mangini F, Parra-Rivas P & Wabnitz S (2022). Multimode soliton collisions in graded-index optical fibers. *Optics Express*, 30(12):21710. DOI:10.1364/OE.459447.

[399] Lafforgue C, Montesinos Ballester M, Dinh Ttd, Le Roux X, Cassan E, Marris-Morini D, Alonso Ramos C & Vivien L (2022). Supercontinuum generation in silicon photonics platforms. *Photonics Research*, 10(3):A43–A56. DOI:10.1364/PRJ.445304.

[400] May S, Clerici M & Sorel M (2021). Supercontinuum generation in dispersion engineered AlGaAs-on-insulator waveguides. *Scientific Reports*, 11(1):2052. DOI:10.1038/s41598-021-81555-3.

[401] Jankowski M, Langrock C, Desiatov B, Marandi A, Wang C, Zhang M, Phillips CR, Lončar M & Fejer MM (2020). Ultrabroadband nonlinear optics in nanophotonic periodically poled lithium niobate waveguides. *Optica*, 7(1):40. DOI:10.1364/OPTICA.7.000040.

[402] Wang X & Wada N (2007). Spectral phase encoding of ultra-short optical pulse in time domain for OCDMA application. *Optics Express*, 15(12):7319. DOI:10.1364/OE.15.007319.

[403] Diddams SA, Hollberg L & Mbele V (2007). Molecular fingerprinting with the resolved modes of a femtosecond laser frequency comb. *Nature*, 445(7128):627–630. DOI:10.1038/nature05524.

[404] Goda K, Tsia KK & Jalali B (2009). Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature*, 458(7242):1145–1149. DOI:10.1038/nature07980.

[405] Godin T, Sader L, Khodadad Kashi A, Hanzard PH, Hideur A, Moss DJ, Morandotti R, Genty G, Dudley JM, Pasquazi A, Kues M & Wetzel B (2022). Recent advances on time-stretch dispersive Fourier transform and its applications. *Advances in Physics: X*, 7(1). DOI:10.1080/23746149.2022.2067487.

[406] Shen TSR & Lau APT (2011). Fiber nonlinearity compensation using extreme learning machine for DSP-based coherent communication systems. *16th Opto-Electronics and Communications Conference*, pages 816–817, Kaohsiung, Taiwan.

[407] Stern B, Ji X, Okawachi Y, Gaeta AL & Lipson M (2018). Battery-operated integrated frequency comb generator. *Nature*, 562(7727):401–405. DOI:10.1038/s41586-018-0598-9.

[408] Bao C, Suh MG & Vahala K (2019). Microresonator soliton dual-comb imaging. *Optica*, 6(9):1110. DOI:10.1364/OPTICA.6.001110.

[409] Op de Beeck C, Haq B, Elsinger L, Gocalinska A, Pelucchi E, Corbett B, Roelkens G & Kuyken B (2020). Heterogeneous III-V on silicon nitride amplifiers and lasers via microtransfer printing. *Optica*, 7(5):386. DOI:10.1364/OPTICA.382989.

[410] Zhang L, Chen J, Ma C, Li W, Qi Z & Xue N (2021). Research Progress on On-Chip Fourier Transform Spectrometer. *Laser & Photonics Reviews*, 2100016 pages. DOI:10.1002/lpor.202100016.

[411] Li A & Fainman Y (2021a). Integrated Silicon Fourier Transform Spectrometer with Broad Bandwidth and Ultra-High Resolution. *Laser & Photonics Reviews*, 2000358 pages. DOI:10.1002/lpor.202000358.

[412] Li A & Fainman Y (2021b). On-chip spectrometers using stratified waveguide filters. *Nature Communications*, 12(1):2704. DOI:10.1038/s41467-021-23001-6.

[413] Huang C, Bilodeau S, Ferreira de Lima T, Tait AN, Ma PY, Blow EC, Jha A, Peng HT, Shastri BJ & Prucnal PR (2020). Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits. *APL Photonics*, 5(4):040803. DOI:10.1063/1.5144121.

[414] Zhang W, Huang C, Peng HT, Bilodeau S, Jha A, Blow E, de Lima TF, Shastri BJ & Prucnal P (2022). Silicon microring synapses enable photonic deep learning beyond 9-bit precision. *Optica*, 9(5):579. DOI:10.1364/OPTICA.446100.

[415] Xiang C, Jin W & Bowers JE (2022). Silicon nitride passive and active photonic integrated circuits: trends and prospects. *Photonics Research*, 10(6):A82. DOI:10.1364/PRJ.452936.

[416] Bu T, Kumar S, Zhang H, Huang I & Huang YP (2020). Single-pixel pattern recognition with coherent nonlinear optics. *Optics Letters*, 45(24):6771. DOI:10.1364/OL.411564.

[417] Bu T, Zhang H, Kumar S, Jin M, Kumar P & Huang Y (2022). Efficient optical reservoir computing for parallel data processing. *Optics Letters*, 47(15):3784. DOI:10.1364/OL.464288.

[418] Basani JR, Heuck M, Englund DR & Krastanov S (2022). All-Photonic Artificial Neural Network Processor Via Non-linear Optics. *arXiv*, 2205.08608 pages.

[419] Gao H, Chen GFR, Xing P, Choi JW & Tan DTH (2021). 3D printed on-chip microtoroid resonators and nested spiral photonic devices. *Photonics Research*, 9(9):1803. DOI:10.1364/PRJ.430801.

[420] Wang R, Wang P, Luo G, Yu H, Zhou X, Zhang Y & Pan J (2020). Silicon-based Optical Neural Network Chip Based on Coherent Detection. *Asia Communications and Photonics Conference/International Conference on Information Photonics and Optical Communications 2020 (ACP/IPOC)*, OSA, T2D.4 pages, Washington, D.C.

[421] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J & Chintala S (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv*, 1912.01703 pages.

Appendix A

Pulse Shaping

In this appendix, the characterization of equipment for the autonomous pulse-shaper is reported. In particular, I shall focus on the radio-frequency (RF) measurements of the two sources for the determination of repetition-rate (required for the sampling calibration), as well as the pulsewidth measurement of the sampling source.
Additionally, the characterization of one delay line switch is shown as an example.

A.1 Source Characterization

For the source characterization, the RF spectrum was measured using a HP E4407B RF Spectrum Analyzer. For the shaping source (Pritel FFL) the RF output port was directly used, while for the sampling source (Menlo FC) the optical monitoring port ($P_{avg} \approx 0.5\,\mathrm{mW}$) and an additional photodiode (Lab buddy DSC-R402) was used. Figure A.1 shows the obtained RF spectra.

For the pulse characterization of the sampling source, the pulse duration was evaluated using an intensity autocorrelator (Femtochrome FR-103XL). The frequency comb source emits uncompensated optical pulses $>2\,\mathrm{ps}$, and to ensure the sampling criterion $\Delta\tau_{Sampling} \ll \Delta\tau_{Target}$ it was characterized before entering the highly-nonlinear fiber. Figure A.2 shows the results.

The pulse duration characterization yielded a value of $\approx 4.4\,\mathrm{ps}$, which is much smaller than the shaping source ($\approx 21\,\mathrm{ps}$), thus validating the use of this source for the sampling experiment.

A.2 Photonic Chip

For the characterization of the sample as a switch, a tunable continuous-wave laser (Tunics Plus) and an optical power-meter (Thorlabs PM100D) was used. In particular, the laser was coupled on the input port, and the optical power was read out at the tap port of the short path (see Fig. A.3a). Subsequently, the switch voltage was swept from -3.8 V to +3.8 V and the variation in power was measured. Afterwards, the previous switch was set to the short path (i.e., all power in the short

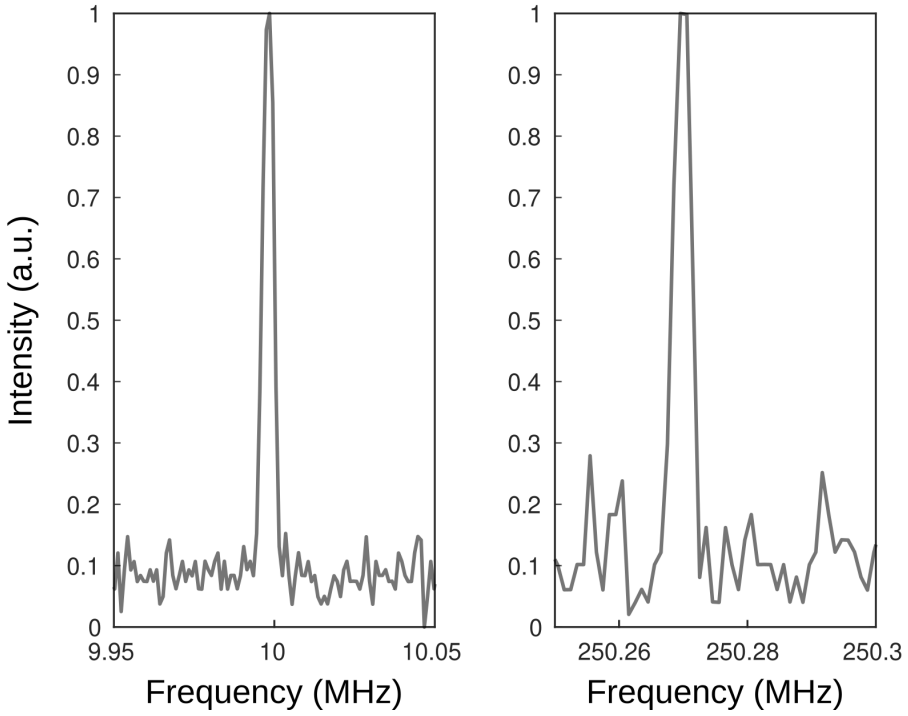


Figure A.1: RF measurement of both laser sources. a) Shaping source operating at 9.999 MHz, b) Sampling source operating at 250.27 MHz.

path) and the procedure was repeated for the next switch. Fig. A.3 shows the experimental setup and results for the first two switches.

To fully characterize the available delays a different experiment was conducted. Here, the frequency comb (Menlo FC) was used as a source and an additional dispersion compensating fiber (Thorlabs DCF-38) was used to compensate the pulse-duration of $\approx$ 100 fs. This short pulse-duration allows for the characterization of the individual delays using two switches with a 50/50 splitting ratio (i.e., generating double pulses spaced by the individual delay) and avoid any overlapping of pulses. An autocorrelator (Femtochrome FR-103XL) was used to characterize the delays. Figure A.4 depicts the results for the first delay (1 ps) and obtained delays vs. the targeted (bit-wise) delays.

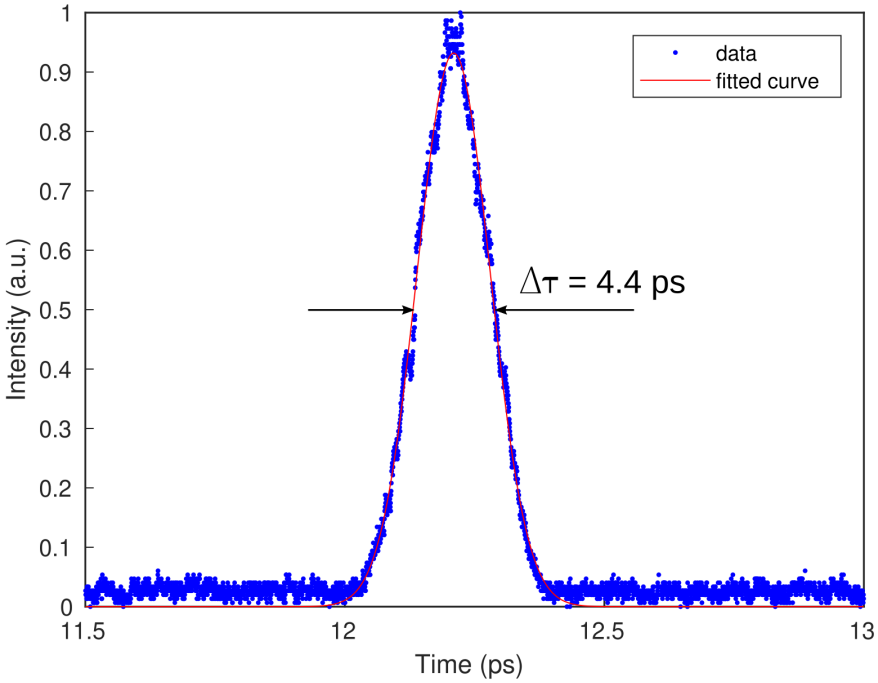


Figure A.2: Pulse duration measurement of the sampling source. For the calculation of the pulse-width, a calibration factor was applied (0.707×37.6 ps/ms) to the retrieved FWHM of the Gaussian fit (0.165 ps).

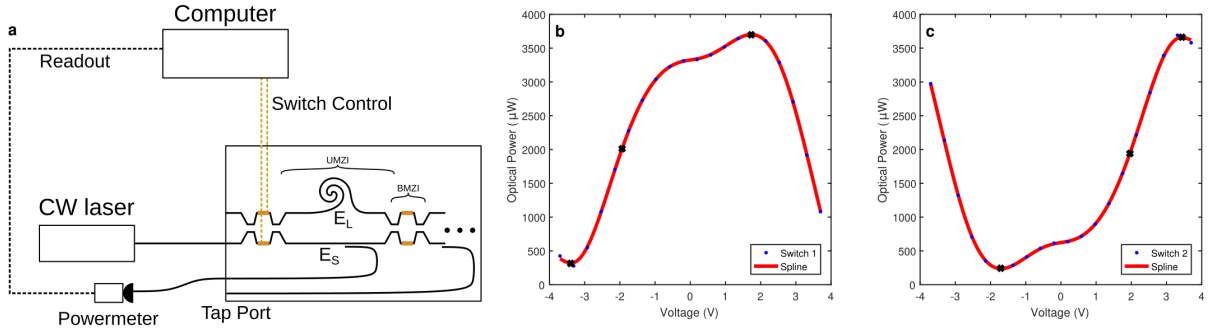


Figure A.3: Voltage characterization of the split-and-delay line switches. a) experimental setup, b) and c) characterization for the first and second switch, respectively. Blue dots represent measurement points, the solid red line a spline interpolation, and the black crosses the minimum (all power to the delay path), half-maximum and maximum (all power to the short path) points of the spline.

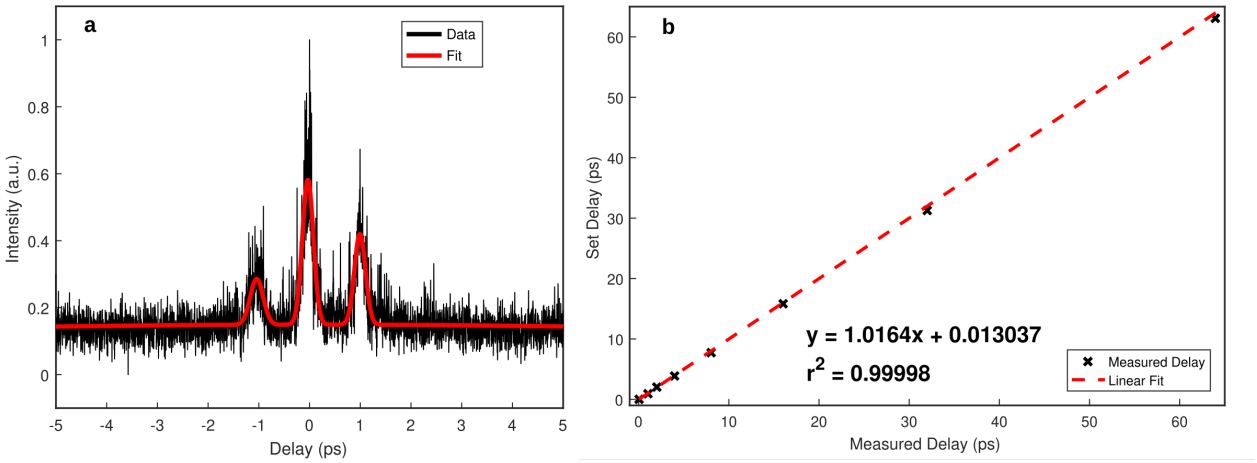


Figure A.4: Delay characterization of the split-and-delay line. a) Autocorrelation trace for the first delay (correction factor applied), returned delay of ≈ 1.02 ps. b) Characterization for subsequent delays up to 64 ps (limited by the autocorrelator).

Appendix B

Transient Optical Neural Network

B.1 Simulations

For experiment, the output of the frequency comb source (Menlo FC) was compressed to ≈ 85.6 fs (close to the transform limit) using dispersion compensating fibers (Thorlabs DCF-38) and subsequently measured with an OSA (Ando AQ6317B).

Simulations were carried out for both polarization axes using the GNLSE and split-step algorithm. Wavelength dependent parameters from the manufacturer were used for the dispersion on both axes and the effective area[1]. The simulation and experimental results are shown in Fig. B.1.

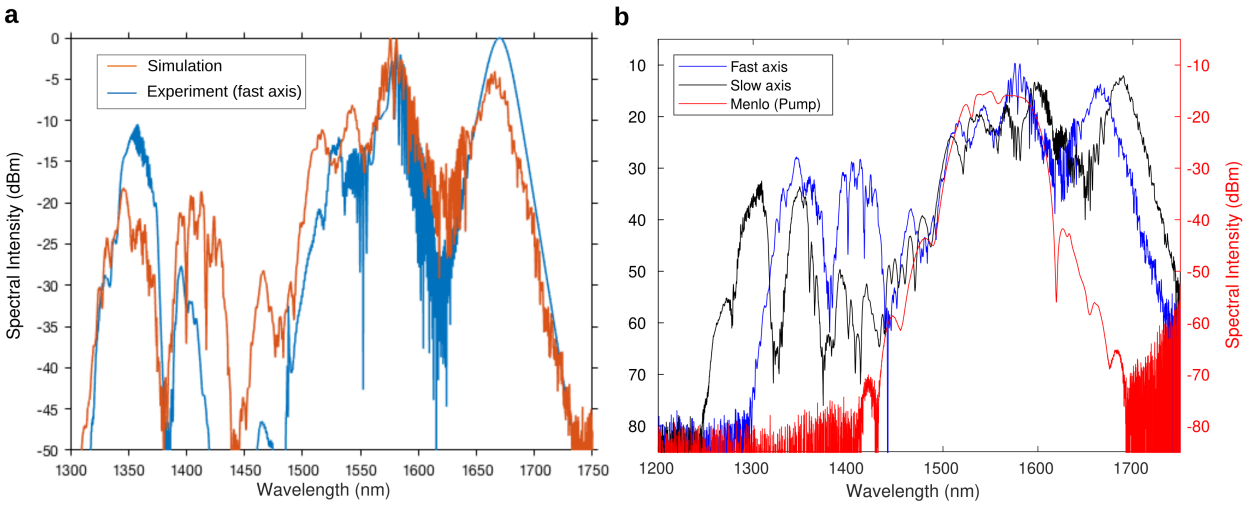


Figure B.1: Comparison between simulated SCG and experimental measured SCG. a) simulated and experimental result (fast axis), b) experimentally measured SCG for both axes.

While there is still uncertainty in the SCG simulaton due to lack of wavelength dependent parameters, the simulations show a better overlap to the experimental results from the fast-axis

[1]Not included due to confidentiality

compared to the slow axis. As such, for the implemented system the fast axis was used for the experiments and future numerical studies.

B.2 Optical Neural Network Comparison

Table B.1 summarizes the results from different optical neural network implementations for the same tasks as presented in this work.

Platform	Task	Test Accuracy	Nonlinearity
This work	Sinc	0.0465 (rmse)	Soliton Dynamics
	Abalone	0.0686 (rmse)	
	Iris	100%	
	Wine	100%	
	MNIST	86.7%	
On-chip accelerator[294]	Iris	97.4%	Digital function
On-chip accelerator[420]	Iris	96.67%	Digital function
EOM ELM[351]	Iris	93.9%	PD ($I \propto E^2$)
	Wine	97.5%	
On-chip accelerator accelerator[420]	Wine	97.22%	Digital function
EOM ELM[357]	MNIST	90-92%	PD ($I \propto E^2$)
Free-space ELM[57]	MNIST	92.2%	PD (saturation)
	Abalone	0.12 (rmse)	
MM fiber[56]	Abalone	0.126 (rmse)	MM coupling
	Sinc	0.0671 (rmse)	

Table B.1: Benchmark comparison of different optical neuromorphic processing approaches. ELM = extreme learning machine, EOM = electro-optic modulator, MM = multi-mode, PD = photodiode

B.3 Digital Neural Networks

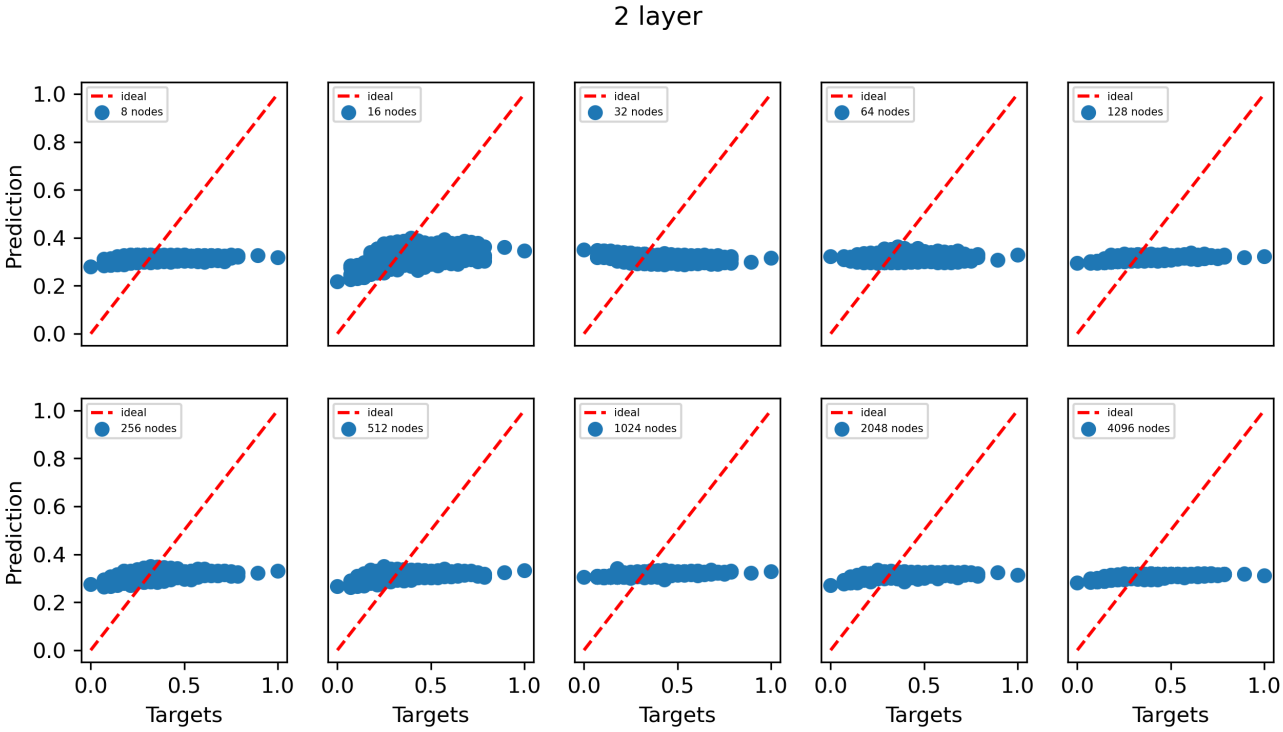
In order to illustrate the performance issue associated to the Abalone task using classical neural networks, a straight-forward multi-layer neural network with 2 layers[2] was implemented, see Figure B.2. In particular, two different loss metrics were investigated: The MSE loss akin to the optical implementation in Chapter 3 and the Kullback-Leibler (KL) divergence[3]. The KL divergence measures the similarity between the predicted sample distribution and the true sample distribution. Additionally, numerous node configurations per layer were investigated. For this, a selected number of nodes for the first layer were chosen, and the second layer was set to 'number of nodes in first

[2] 1 to 3 layers was tested with 2 being the most performative

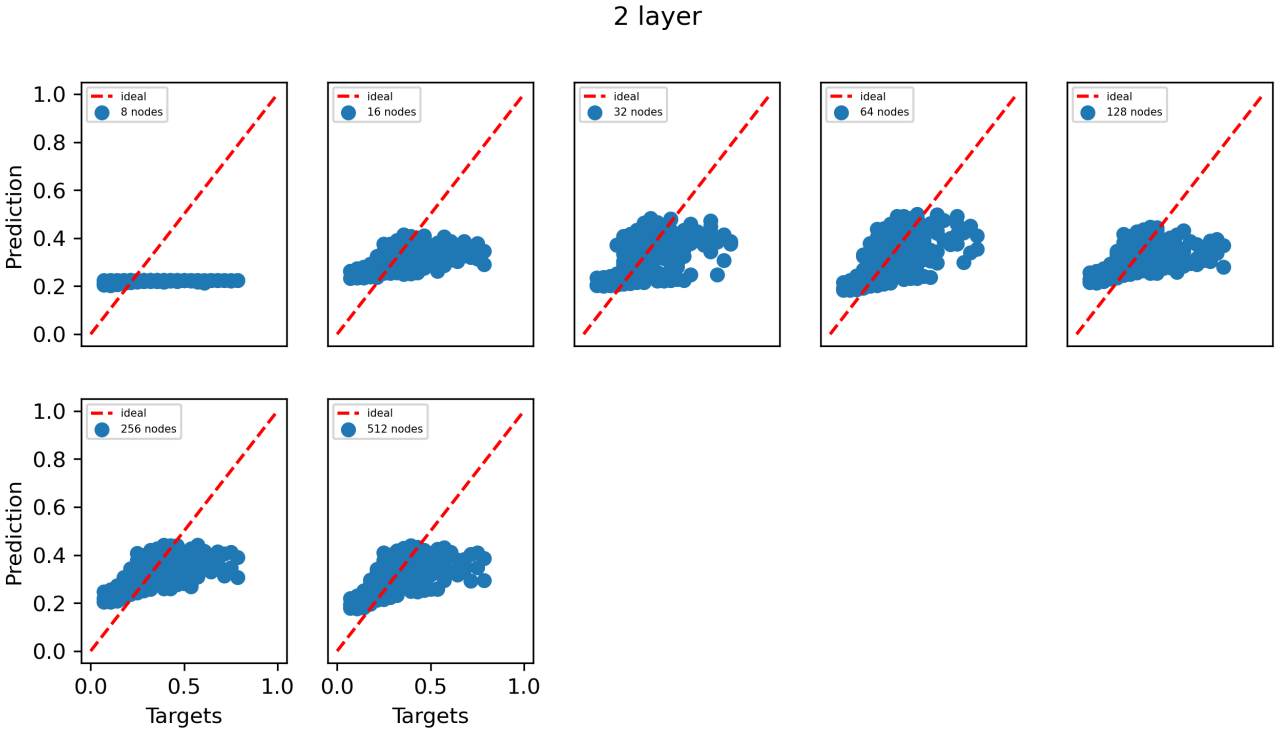
[3] The PyTorch function KLDivLoss was used [421]

layer divided by 8'. For the activation function the hyperbolic tangent (tanh[...]) was used, for training a stochastic gradient descent (SGD) algorithm was used with a learning rate of 0.0001, the batch size was 16 for the MSE loss and 128 for the KL divergence loss. Overfitting was prevented by means of a 5-fold cross validation. For both digital comparisons, the same data partition as in the optical experiment was used.
In case of the MSE loss (Fig. B.2a), the neural network evidently converges to a point where it finds the average in the data (around 0.3), while also achieving relatively good RMSE values (best case: 0.119 RMSE for 1024 nodes in the first layer) but no good generalization (i.e., a data distribution closer to the ideal solution). The main reason for this is the data distribution provided for this particular set, as the majority of the data centers around 0.3. Therefore, a solution where the majority of data is fitted almost perfectly is preferred for this metric by the neural network as the outliers do not significantly impact the MSE value. On the other hand (Fig. B.2b), more complex metrics such as the KL divergence can achieve significant better performance in such cases (best case: 0.095 RMSE for 128 nodes in the first layer). However, it is noteworthy that this increase in performance comes at the cost of increased computational complexity during training. Contrary, the optical neural network was able to achieve a much better performance (see Fig. 3.20, RMSE of 0.0766) using only the MSE loss.

For the digital comparison of the COVID-19 results, a support vector machine (SVM) model using the SCIKIT-LEARN toolkit with the sub-class LINEARSVC [369] was implemented. The optimization of hyper-parameters was conducted with a 5-fold cross-validation scheme and an ℓ^2 penalty as the loss function. The optimal penalty parameter C was determined by exhaustively experimenting values from 10^-5 to 10 and a value of 0.05 was found to be the optimum.



(a) Results for MSE loss.



(b) Results for KL-divergence loss.

Figure B.2: Abalone results for different software multi-layer neural networks. a) Results for MSE loss metric. b) Results for KL-divergence metric.

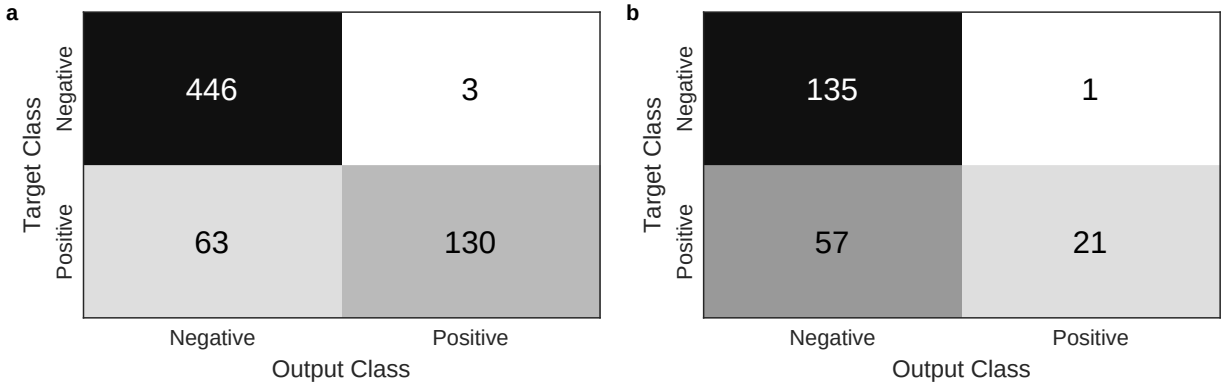


Figure B.3: SVM results for the COVID-19 task (absolute numbers). a Training results (UAR = 0.8334), b Test results (UAR = 0.6309)

Appendix C

Publication List

C.1 Journals

J11 B. MacLellan, P. Roztocki, J. Belleville, L. Romero Cortés, K. Ruscitti, **B. Fischer**, J. Azaña, and R. Morandotti, "Inverse design of fibre-based photonic systems," *submitted* (2023)

J10 **B. Fischer**, M. Chemnitz, Y. Zhu, N. Perron, P. Roztocki, B. MacLellan, L. di Lauro, A. Aadhi, C. Rimoldi, T. Falk and R. Morandotti, "Transient Optical Neural Emulation," *submitted, under review* (2023).

J9 S. Sciara, F. Nosfrati, H. Yu, P. Roztocki, **B. Fischer**, M. Chemnitz, C. Reimer, L. Romero Cortés, W. Munro, D. Moss, R. Lo Franco, A. Cino, L. Caspani, M. Kues, Z. Wang, J. Azaña, and R. Morandotti, "Generation of integrated quantum frequency combs on microring resonators for the realization of complex entangled two-photon states," *submitted to IEEE Journal of Selected Topics in Quantum Electronics* (*invited*) (2022).

J8 **B. Fischer**, M. Chemnitz, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "Autonomous on-chip interferometry for reconfigurable optical waveform generation," Optica **8** (10), pp. 1268-1276 (2021).

J7 P. Roztocki, B. MacLellan, M. Islam, C. Reimer, **B. Fischer**, S. Sciara, R. Helsten, Y. Jestin, A. Cino, S. T. Chu, B. Little, D. J. Moss, M. Kues, and R. Morandotti, "Arbitrary phase access for stable fiber interferometers," Laser & Photonics Reviews **15** (7), 2000524 (2021).

J6 S. Sciara, P. Roztocki, **B. Fischer**, C. Reimer, L. Romero Cortés, W. Munro, D. Moss, A. Cino, L. Caspani, M. Kues, J. Azaña, and R. Morandotti, "Scalable and effective multi-level entangled photon states: a promising tool to boost quantum technologies," Nanophotonics **10** (18), pp. 4447-4465 (2021).

J5 Y. Zhang, M. Kues, P. Roztocki, C. Reimer, **B. Fischer**, B. MacLellan, A. Bisianov, U. Peschel, B.E. Little, S.T. Chu, D.J. Moss, L. Caspani, and R. Morandotti, "Induced Photon Correlations Through the Overlap of Two Four-Wave Mixing Processes in Integrated Cavities," Laser & Photonics Reviews **14** (9), 2000128 (2020).

J4 C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J.

Azaña, M. Kues, and R. Morandotti, "High-dimensional one-way quantum processing implemented on d-level cluster states," Nature Physics **15**, 148-153 (2019).

J3 S. Sciara, P. Roztocki, C. Rimoldi, M. Chemnitz, C. Reimer, **B. Fischer**, L. Romero Cortés, W. J. Munro, D.J. Moss, L. Caspani, J. Azaña, M. Kues, and R. Morandotti, "Quantum Frequency Combs for Generation and Processing of Complex Entangled Photon States," IEEE Photonics Technology Letters **31** (23), pp.1862-1865 (*invited*) (2019).

J2 P. Roztocki, S. Sciara, C. Reimer, L. Romero Cortés, Y. Zhang, B. Wetzel, M. Islam, **B. Fischer**, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, J. Azaña, M. Kues, and R. Morandotti, "Complex quantum state generation and coherent control based on integrated frequency combs," IEEE Journal of Lightwave Technology **37** (2), pp. 338-344 (*invited*) (2018).

J1 C. Reimer, Y. Zhang, P. Roztocki, S. Sciara, L. Romero Cortés, M. Islam, **B. Fischer**, B. Wetzel, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, J. Azaña, M. Kues, R. Morandotti, "On-chip frequency combs and telecommunications signal processing meet quantum optics," Front. Optoelectron. **11** (2), pp. 134-137, (*invited*) (2018).

C.2 Book Chapters

B2 C. Rimoldi, **B. Fischer**, L. di Lauro, M. Chemnitz, A. Pasquazi, D.J. Moss, R. Morandotti, „Dissipative solitons in microresonators," in Dissipative Optical Solitons, Springer (2022); ISBN: 9783030974923.

B1 **B. Fischer**, P. Roztocki, C. Reimer, S. Sciara, Y. Zhang, M. Islam, L. Romero Cortés, S. Bharadwaj, D.J. Moss, J. Azaña, L. Caspani, M. Kues, R. Morandotti, "Nonlinear and quantum effects in integrated microcavities," in Nonlinear Meta Optics, CRC Press (2020); ISBN: 9781138576544

C.3 Patents

P5 A. Rahim, L. di Lauro, **B. Fischer**, P. Dimitriev, P. Roztocki, A. Eshaghi, Y. Jestin, R. Morandotti, "Method and apparatus for optical information processing," CA 20,22050451 (patent pending).

P4 **B. Fischer**, M. Chemnitz, B. Maclellan, P. Roztocki, J. Azaña, Y. Jestin, and R. Morandotti, "System and method for arbitrary optical waveform generation," US 63/201,378 (patent pending).

P3 B. MacLellan, P. Roztocki, J. van Howe, L. Romero Cortes, **B. Fischer**, Y. Jestin, R. Morandotti, "Method and system for design of photonic systems," US 17/303,705 (patent pending).

P2 **B. Fischer**, P. Roztocki, M. Kues, M. Chemnitz, C. Rimoldi, B. MacLellan, L. Romero Cortes, J. Azaña, Y. Jestin, and R. Morandotti, "System and method for optical information processing with a reconfigurable nonlinear optical network," US 63/051,435 (patent pending).

P1 M. Islam, P. Roztocki, M. Kues, C. Reimer, **B. Fischer**, S. Sciara, R. Helsten, Y. Zhang, Y. Jestin, R. Morandotti, "System and method for active stabilization at an arbitrary phase of multiport fiber-based interferometers," US 11,506,477 (granted).

C.4 Conferences

** As presenting author*

C45 L. Di Lauro, S. Sciara, A. Aadhi, I. Alamgir, P. Dimitriev, **B. Fischer**, S. Kamali, R. Fazili, C. Mazoukh, A. Eshaghi, B.E. Little, S.T. Chu, D.J. Moss, and R. Morandotti, "Integrated Photonics for AI-assisted Telecom Signal Processing," Photonics and Electromagnetics Research Symposium (PIERS), 2023.

C44 S. Sciara, H. Yu, M. Chemnitz, N. Montaut, **B. Fischer**, B. Crockett, B. Wetzel, B.E. Little, S.T. Chu, D.J. Moss, J. Azaña, Z. Wang, and R. Morandotti, "Picosecond-spaced time-entangled photonic qudits on chip and their use for quantum communications," 2023 Conference on Lasers and Electro-Optics - CLEO, 2023.

C43 C. Mazoukh, L. Di Lauro, **B. Fischer**, A. Aadhi, I. Alamgir, A. Eshaghi, B.E. Little, S.T. Chu, D.J. Moss, and R. Morandotti, "Machine-Learning Enabled Optimization for Robust Soliton Crystal Generation in Microring Resonators," 2023 Conference on Lasers and Electro-Optics - CLEO, 2023.

C42 N. Perron, M. Chemnitz, **B. Fischer**, S. Junaid, M. A. Schmidt, and R. Morandotti, "All-optical digital processing in Carbon Disulfide liquid core optical fiber," 2023 Conference on Lasers and Electro-Optics - CLEO, 2023.

C41 N. Montaut, P. Roztocki, H. Yu, M. Chemnitz, S. Sciara, Y. Jestin, B. MacLellan, **B. Fischer**, M. Kues, C. Reimer, L. Romero Cortes, B. Wetzel, Y. Zhang, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W.J. Munro, J. Azaña, and R. Morandotti, "Telecom-compatible, affordable and scalable quantum technologies," IEEE Photonics Conference (IPC), 2022.

C40 N. Montaut, P. Roztocki, H. Yu, M. Chemnitz, S. Sciara, Y. Jestin, B. MacLellan, **B. Fischer**, M. Kues, C. Reimer, L. Romero Cortes, B. Wetzel, Y. Zhang, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W.J. Munro, J. Azaña, and R. Morandotti, "Scalable architectures for CMOS and telecom-compatible quantum information processing," SPIE Photonics West OPTO, San Francisco (CA), USA, 2023, *invited*.

C39 S. Sciara, H. Yu, M. Chemnitz, **B. Fischer**, P. Roztocki, B. Crockett, C. Reimer, L. Caspani, M. Kues, W. Munro, S. Chu, B. Little, D. Moss, Z. Wang, J. Azaña, and R. Morandotti, "On-Chip Generation and Telecom-Compatible Processing of Complex Photonic Systems in Time and Frequency," Frontiers in Optics - Proceedings Frontiers in Optics + Laser Science APS/DLS, 2022, FW1B.1, *invited*.

C38 V. T. Hoang, **B. Fischer**, M. Chemnitz, Y. Boussafa, L. Sader, P. Roztocki, B. MacLellan, C. Reimer, S. Fevrier, V. Couderc, B. Little, S. Chu, E. Viktorov, D. Moss, M. Kues, J. Azaña, A. Pasquazi, M. Peccianti, R. Morandotti, and B. Wetzel, "Smart Photonics: From On-chip Pulse Processing to Nonlinear Fiber Propagation Dynamics Optimization," 2nd Global Summit and Expo on Lasers, Optics and Photonics (GSELOP2022), 2022.

C37 P. Dmitriev, L. Di Lauro, **B. Fischer**, A. Aadhi, M. Chemnitz, E. Viktorov, A. Kovalev, and R. Morandotti, "Analysis of time-delay photonic reservoirs for neuromorphic computing," Photonics North, 2022.

C36 H. Yu, M. Chemnitz, S. Sciara, **B. Fischer**, B. Crockett, P. Roztocki, B. E. Little, S. T. Chu, D. J. Moss, J. Azaña, Z. Wang, and R. Morandotti, "On-chip generation and full characterization of time-bin entanglement," Photonics North, 2022.

C35 N. Perron, M. Chemnitz, **B. Fischer**, S. Junaid, M. A. Schmidt, and R. Morandotti, "Low-power picosecond to femtosecond pulse conversion through efficient shock-front formation in monolithic liquid-core optical fibres," Photonics North, 2022.

C34 **B. Fischer**, M. Chemnitz, B. Wetzel, P. Roztocki, B. MacLellan, C. Reimer, B. Little, S. Chu, E. Viktorov, D. Moss, M. Kues, J. Azaña, A. Pasquazi, M. Peccianti, and R. Morandotti, "User-friendly, reconfigurable all-optical signal processing with integrated photonics," 3rd URSI AT-AP-RASC, 2022, We-D-AI-PM4-1, *invited*.

C33 **B. Fischer***, M. Chemnitz, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "Scalable, Autonomous On-Chip Picosecond Pulse-Shaping Enabled by Smart Optimization," 2022 Conference on Lasers and Electro-Optics - CLEO, 2022, STh2E.5.

C32 M. Chemnitz, **B. Fischer**, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "On-chip temporal coherence synthesis for classical and quantum waveform processing," 2022 Conference on Lasers and Electro-Optics - CLEO, 2022, SF3G.2.

C31 M. Chemnitz, H. Yu, S. Sciara, **B. Fischer**, P. Roztocki, B. Crockett, C. Reimer, L. Caspani, M. Kues, W. Munro, S. Chu, B. Little, D. Moss, Z. Wang, J. Azaña, and R. Morandotti, " Telecom-compatible, on-chip generation and processing of complex photon states in time and frequency ," SPIE OPTO - Photonics West, 2022, Integrated Optics: Devices, Materials, and Technologies XXVI 12004, 42-45.

C30 M. Chemnitz, N. Perron, **B. Fischer**, S. Junaid, M. A. Schmidt, and R. Morandotti, "Towards telecom-compatible liquid core fibers for low-power nonlinear signal processing," OSA Advanced Photonics Congress, 2021, NoW3C.6.

C29 M. Chemnitz, **B. Fischer***, B. MacLellan, P. Roztocki, R. Helsten, B. Wetzel, B. E. Little, S. T. Chu, D. J. Moss, J. Azaña, and R. Morandotti, "All-optical Sampling for Adaptive On-Chip Picosecond Pulse-Shaping," OSA Nonlinear Optics 2021, Washington, D.C., 2021, NF2A.1.

C28 **B. Fischer**, M. Chemnitz, B. MacLellan, P. Roztocki, B. Wetzel, B. Little, S. Chu, D. Moss, J. Azaña, and R. Morandotti, "Picosecond Pulse Shaping via On-Chip Interferometry," Frontiers in Optics - Proceedings Frontiers in Optics + Laser Science APS/DLS, 2021, FM1B.5.

C27 B. Maclellan, P. Roztocki, M. Islam, C. Reimer, **B. Fischer**, S. Sciara, R. Helsten, Y. Jestin, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, M. Kues, and R. Morandotti, "Fiber Interferometers for Time-domain Quantum Optics," Conference on Lasers and Electro-Optics (CLEO), 2021, JTu3A.131.

C26 N. Perron, M. Chemnitz, **B. Fischer**, S. Junaid, M. A. Schmidt, and R. Morandotti, "Highly efficient cross-phase modulation in monolithic CS2 liquid-core fibre," Photonics North, 2021.

C25 H. Yu, M. Chemnitz, S. Sciara, **B. Fischer**, B. Crockett, P. Roztocki, B. E. Little, S. T. Chu, D. J. Moss, J. Azaña, Z. Wang, and R. Morandotti, "On-chip generation of telecommunication

compatible ultrafast time-bin entangled qubits," Conference on Lasers and Electro-Optics (CLEO), 2021, FTu4G.3.

C24 H. Yu, M. Chemnitz, S. Sciara, **B. Fischer**, B. Crockett, P. Roztocki, B. E. Little, S. T. Chu, D. J. Moss, J. Azaña, Z. Wang, and R. Morandotti, "Generation of ultrafast-spacing time-bin entangled qubit on chip," Photonics North, 2021.

C23 H. Yu, M. Chemnitz, S. Sciara, **B. Fischer**, B. Crockett, P. Roztocki, B. E. Little, S. T. Chu, D. J. Moss, J. Azaña, Z. Wang, and R. Morandotti, "On-chip generation and characterization of densely-spaced time-bin entangled qubits," OSA Advanced Photonics Congress, 2021, SpF2E.6.

22 S. Sciara, H. Yu, M. Chemnitz, **B. Fischer**, P. Roztocki, B. Crockett, C. Reimer, B. Little, L. Caspani, W. Munro, S. Chu, D. Moss, J. Azaña, M. Kues, Z. Wang, R. Morandotti, "On-chip time and frequency modes for the generation and processing of complex photon states," IEEE Photonics Conference (IPC), 2021.

C21 P. Roztocki, M. Chemnitz, B. MacLellan, S. Sciara, C. Reimer, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, W. J. Munro, J. Azaña, M. Kues, and R. Morandotti, "Designing Time and Frequency Entanglement for Generation of High-Dimensional Photon Cluster States," 22nd International Conference on Transparent Optical Networks (ICTON), Bari, Italy, 2020.

C20 **B. Fischer***, A. Aadhi, C. Rimoldi, P. Roztocki, L. di Lauro, M. Chemnitz, A.V. Kovalev, S. Chu, B. Little, D. J. Moss, E.A. Viktorov, M. Kues, and R. Morandotti, "Optical frequency comb generation by hybrid mode-locking in a nested cavity scheme," OSA Advanced Photonics Congress, Virtual Conference, 2020, JTu3A.6.

C19 A. Aadhi, P. Roztocki, A. V. Kovalev, M. Kues, **B. Fischer**, C. Reimer, Y. Zhang, T. Wang, B.E. Little, S.T. Chu, Z. Wang, D.J. Moss, E.A. Viktorov, and R. Morandotti, "Distinct laser dynamics from a single figure-eight laser with an integrated nonlinear waveguide," OSA Advanced Photonics Congress, Virtual Conference, 2020.

C18 P. Roztocki, B. MacLellan, M. Islam, C. Reimer, **B. Fischer**, S. Sciara, R. Helsten, Y. Jestin, A. Cino, S. T. Chu, B. Little, D. J. Moss, M. Kues, and R. Morandotti, "Phase retrieval in fiber-based interferometers", OSA Advanced Photonics Congress, Virtual Conference, 2020.

C17 S. Sciara, C. Reimer, P. Roztocki, **B. Fischer**, L. Romero Cortés, D. J. Moss, L. Caspani, W. J. Munro, J. Azaña, M. Kues, and R. Morandotti, "Generation and coherent manipulation of complex entangled photon states based on integrated quantum frequency combs," OSA Advanced Photonics Congress, Virtual Conference, 2020.

C16 P. Roztocki, M. Kues, Y. Zhang, C. Reimer, **B. Fischer**, B. MacLellan, A. Bisianov, U. Peschel, B.E. Little, S.T. Chu, D.J. Moss, L. Caspani, and R. Morandotti, "Induced photon correlations by the superposition of two four-wave mixing processes on a photonic chip," OSA Advanced Photonics Congress, Virtual Conference, 2020.

C15 **B. Fischer***, A. Aadhi, C. Rimoldi, P. Roztocki, L. di Lauro, M. Chemnitz, S. Chu, B. Little, D. J. Moss, and R. Morandotti, "Hybrid mode-locking in a nested cavity scheme," 22nd Photonics North, Virtual Conference, 2020.

C14 P. Roztocki, M. Kues, Y. Zhang, C. Reimer, **B. Fischer**, B. MacLellan, A. Bisianov, U. Peschel, B.E. Little, S.T. Chu, D.J. Moss, L. Caspani, and R. Morandotti, "Induced photon correlations by the superposition of two four-wave mixing processes on a photonic chip," 22nd Photonics North, Virtual Conference, 2020.

C13 **B. Fischer***, C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, M. Kues, and R. Morandotti, "High-dimensional d-level cluster states with on-chip frequency combs," SPIE Photonics West OPTO, San Francisco (CA), USA, 2019, *invited*.

C12 M. Kues, S. Sciara, P. Roztocki, **B. Fischer**, C. Reimer, M. Islam, L. Romero Cortés, Y. Zhang, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, and R. Morandotti, "On-chip quantum frequency comb for complex photon state generation," SPIE Photonics West LASE, San Francisco (CA), USA, 2019, *invited*.

C11 R. G. Krämer, **B. Fischer**, C. Matzdorf, T. A. Goebel, D. Richter, R. Morandotti, and S. Nolte, "High contrast ultrashort pulse written transmission filter based on Moiré-fiber Bragg grating," SPIE Photonics West LASE, San Francisco (CA), USA, 2020.

C10 P. Roztocki, M. Islam, C. Reimer, **B. Fischer**, S. Sciara, Y. Zhang, D. J. Moss, M. Kues and R. Morandotti, "Removing phase ambiguity in fiber-based interferometers for coherent time-bin operations', 2019 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC), Munich, Germany, 2019.

C9 P. Roztocki, C. Reimer, S. Sciara, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, W. J. Munro, J. Azaña, M. Kues, and R. Morandotti, "Hyper-entanglement in time and frequency", 2019 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC), Munich, Germany, 2019.

C8 M. Kues, C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, and R. Morandotti, "D-dimensional frequency-time entangled cluster states with on-chip frequency combs," OSA Advanced Photonic Congress (AP), Paper No. ITh2A.4, Burlingame (CA), USA, 2019.

C7 P. Roztocki, C. Reimer, S. Sciara, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer***, S. Loranger, R. Kashyap, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, W. J. Munro, J. Azaña, M. Kues, and R. Morandotti, "Kerr combs and telecommunications components for the generation and high-dimensional quantum processing of d-level cluster states," 21st International Conference on Transparent Optical Networks (ICTON), Angers, France, 2019.

C6 **B. Fischer***, P. Roztocki, M. Islam, C. Reimer, S. Sciara, Y. Zhang, D.J. Moss, M. Kues, and R. Morandotti, "Phase Retrieval in fiber-based interferometers," Photonics North, Quebec City, 2019.

C5 M. Kues, C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortes, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B. E. Little, D. J. Moss, L. Caspani, W. J. Munro, J. Azana, and R. Morandotti, "D-dimensional frequency-time entangled cluster states with on-chip frequency combs," Integrated Photonics Research, Silicon and Nanophotonics, Burlingame, California (USA), 2019.

C4 S. Sciara, C. Reimer, M. Kues, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, and R. Morandotti, "On-chip Generation, Coherent Control and Processing of Complex Entangled Photon States," IEEE Photonics Society Summer Topical Meeting Series (SUM), 2019, *invited*.

C3 M. Kues, C. Reimer, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, and R. Morandotti, "High-dimensional one-way quantum processing enabled by optical d-level cluster states," Quantum Information and Measurement (QIM) V: Quantum Technologies, Rome, Italy, 2019.

C2 C. Reimer, M. Kues, S. Sciara, P. Roztocki, M. Islam, L. Romero Cortés, Y. Zhang, **B. Fischer**, S. Loranger, R. Kashyap, A. Cino, S.T. Chu, B.E. Little, D.J. Moss, L. Caspani, W. Munro, J. Azaña, and R. Morandotti, "High-dimensional one-way quantum computation operations with on-chip optical d-level cluster states," Conference on Lasers and Electro-Optics (CLEO), San Jose, USA, 2019.

C1 Y. Zhang, M. Islam, P. Roztocki, C. Reimer, S. Sciara, **B. Fischer**, Y. Bromberg, L. Caspani, S.T. Chu, B.E. Little, D.J. Moss, M. Kues, and R. Morandotti, "On-chip emissions of four-photon qubit states," Photonics North, Montreal, Canada, 2018.

Appendix D

Synopsis

D.1 Introduction

L'intelligence artificielle (IA) et les algorithmes d'apprentissage machine (ML) révolutionnent actuellement plusieurs domaines d'application, promettant l'avènement d'une quatrième révolution industrielle [1, 2]. Il n'est pas surprenant que les techniques et les algorithmes d'apprentissage automatique soient rapidement devenus l'élément clé l'épine de nombreux services et produits technologiques de tous les jours, permettant la création de nouveaux appareils et services aux consommateurs, tels que des appareils et des assistants pour la maison intelligente ou des voitures autonomes. En outre, ces dernières années, ces approches sont également devenues un outil essentiel dans les domaines de la science et de l'industrie, par exemple dans les sciences physiques telles que la microscopie et l'optique ultrarapide [3, 4], le développement de nouveaux matériaux et médicaments [5–8] ou le diagnostic médical [9, 10], pour n'en citer que quelques-uns.

Étant donné la quantité toujours croissante de données qui accompagne une telle révolution, un domaine de recherche et d'application qui peut sans aucun doute bénéficier de cette évolution est le traitement des signaux optiques [11, 12]. En effet, les concepts d'IA (y compris les algorithmes ML et les techniques d'optimisation guidée [13]) peuvent contribuer à améliorer les performances des lignes de télécommunication grâce à un post-traitement numérique avancé ou, idéalement, à un matériel de traitement du signal entièrement optique plus efficace [11, 12, 14, 15]. La fusion des techniques d'IA et de ML avec la photonique est souvent appelée la photonique intelligente, et peut être généralement divisée en deux branches principales, comme illustrées dans la Fig. 1.1:

- *"L'apprentissage automatique pour la photonique"* qui décrit l'utilisation de l'apprentissage automatique et des techniques d'optimisation pour la photonique.
- *"Photonic machine learning"*, décrivant la mise en œuvre de systèmes optiques d'apprentissage automatique.

Toutefois, il convient de noter plusieurs des concepts issus de cette division ne sont pas mutuellement exclusive et qu'il y existe multiples chevauchements.

En effet, des progrès rapides et prometteurs ont été réalisés dans le domaine de la photonique intelligente. Cependant, les mises en œuvre effectuées jusqu'à présent utilisent souvent des conceptions de systèmes hautement spécifiques aux applications ou des puces photoniques intégrées complexes. Bien que ces systèmes et puces photoniques puissent atteindre des performances remarquables, le développement de tels systèmes complexes est souvent prohibitif en matière de ressources utilisées, de temps et de couts. Par conséquent, elles ne sont accessibles qu'à une poignée de groupes. Un tel avancement limite de toute évidence la progression opportune de la technologie, ce qui à son tour peut potentiellement être préjudiciable pour l'ensemble du domaine de la recherche [64]. Par conséquent, de nouvelles approches pour le domaine de la photonique intelligente sont nécessaires. D'abord, il est possible de réutiliser les composants et des dispositifs photoniques existants pour de nouvelles applications, en plus d'utiliser des effets optiques disponibles via ces composants, le tout joint à des designs de systèmes spécifiques. Cette thèse vise à proposer et à démontrer une nouvelle stratégie dans ce domaine. Elle pose la question de savoir dans quelle mesure les composants optiques standards et les dispositifs optiques polyvalents peuvent être réutilisés et renforcés par l'utilisation d'algorithmes d'apprentissage automatique en combinaison avec l'optique non linéaire. Il est frappant de constater que, bien que les effets optiques non linéaires soient largement utilisés dans le traitement classique du signal [65] et qu'ils aient été proposés très tôt dans le contexte de l'apprentissage automatique photonique [66], ils ont rarement été exploités jusqu'à présent pour des implémentations en *photoniques intelligentes*. Cela soulève à son tour plusieurs questions:

- **L'optique non linéaire peut-elle constituer un outil polyvalent pour le domaine de la photonique intelligente?**
- Que peut faire l'apprentissage automatique pour la photonique et que peut faire la photonique pour l'apprentissage automatique?
  - Les composants intégrés existants peuvent-ils être réorientés vers de nouveaux domaines d'application en utilisant des approches d'apprentissage automatique?
  - Les composants photoniques existants sur étagère peuvent-ils être réutilisés pour de nouveaux concepts informatiques (neuromorphiques)?

Le traitement des signaux optiques peut bénéficier doublement de la mise en œuvre des techniques de ML en conjonction avec les effets optiques non linéaires. Premièrement, ils permettent une utilisation plus variée des composants existants en utilisant des algorithmes *intelligent* [13, 67]. Deuxièmement, ils allouent l'élaboration de nouveau système de traitement d'information optique, souvent inhabituel, tel que les systèmes informatiques neuromorphiques (c'est-à-dire inspirés du cerveau) [68, 69], qui sont capables de traiter l'information très efficacement. Par conséquent, les méthodes de ML et les algorithmes d'optimisation peuvent fournir une nouvelle méthodologie et également améliorer les systèmes existants au-delà de leur utilisation et de leur conception prévues [15, 70].
Cependant, les solutions de surveillance efficaces et peu couteuses constituent un défi majeur pour l'application des dispositifs intelligents en dehors du laboratoire. En effet, la plupart des implémentations photoniques intelligentes utilisent jusqu'à présent des diagnostics standards, mais souvent couteux et complexes, tels que la reconstruction plein champ [39, 71], les oscilloscopes ultrarapides ou en temps réel [34, 72], ou les analyseurs de radiofréquences [30, 31]. Ces techniques sont également souvent limitées en matière de bande passante ($\ll 100\,\mathrm{GHz}$), d'où la nécessité de développer et de mettre en œuvre des solutions de mesure à la volée efficaces, peu couteuses, mais adaptées. Ces techniques sont d'un grand intérêt, car elles permettent d'atteindre des taux de traitement à la vitesse de la lumière sans conversion électrooptique inefficace dans des empreintes compactes [14, 65, 70]. L'optique non linéaire dans les guides d'ondes compacts offre des solutions élégantes et efficaces pour surmonter cette limitation.

En effet, la réalisation de nouveaux schémas de calcul est un sujet de grand intérêt ces dernières années étant donné l'impact environnemental croissant des approches numériques d'IA et de ML [73]. En particulier, les approches d'apprentissage automatique sont souvent réalisées à l'aide d'une architecture informatique classique, qui est majoritairement inefficace en raison d'un manque de parallélisme ou d'un besoin d'énergie pour le stockage et le transfert des données, ce qui entraine une mauvaise utilisation des ressources dans ces approches logicielles. Par exemple, l'entrainement de réseaux neuronaux de pointe, comme pour le traitement du langage naturel, peut produire autant de CO2 que plusieurs vols de New York à San Francisco [73]. Cette évolution limite de toute évidence la portée et l'échelle des problèmes auxquels l'apprentissage automatique peut être appliqué, par exemple dans le contexte du "big data" ou des applications portables. Le matériel de photonique neuromorphique (inspiré du cerveau) peut jouer un rôle important pour surmonter les limitations actuelles en fournissant un traitement de l'information économe en énergie et ultrarapide à la vitesse de la lumière [69], car il peut effectuer certaines opérations mathématiques de manière totalement passive ou à une puissance extrêmement faible (femto à attojoule par opération [70]). De plus, l'utilisation de la photonique permet un haut degré de parallélisme en exploitant les différents degrés de liberté dont dispose la photonique [74, 75]. Bien que de nombreuses approches prometteuses aient été développées [16, 76–78], de nombreuses solutions sont souvent négligées au profit de conventions existantes dans des domaines dont les nouvelles méthodes sont empruntées (par exemple, ici l'informatique), ce qui aboutit souvent à des solutions conçues avec une complexité toujours croissante plutôt qu'à la recherche d'innovation et de principes peu communs [79].

D.2 Structure de la these

L'objectif de la thèse est d'étudier l'utilisation des effets optiques non linéaires pour la mise en œuvre de dispositifs de *photoniques intelligente* pour le traitement efficace et évolutif des signaux

entièrement optiques. En particulier, le travail se concentre sur l'exploration de l'utilisation de techniques optiques non linéaires pour la surveillance et l'inférence optique, en mettant l'accent sur la réutilisation des composants intégrés et à base de fibres existants. À cette fin, une solution de surveillance d'échantillonnage optique basée sur le mélange dégénéré à quatre ondes est étudiée pour permettre une mise en forme autonome des impulsions sur la puce dans le domaine de la picoseconde (Chapitre 2). Par la suite, une mise en œuvre simple, mais puissante de l'émulateur de réseau neuronal tout optique, constitué seulement de fibre optique, basé sur la fission de solitons est présentée et étudiée (Chapitre 3). Ces deux applications distinctes ont été rendues possibles grâce à l'utilisation d'équipements et de méthodes très similaires, à savoir des fibres hautement non linéaires, des filtres programmables et des routines d'optimisation guidées, démontrant ainsi le potentiel économique d'équipements communs ayant la capacité d'être reconfigurés. Chaque chapitre contient une introduction distincte dans le domaine de la recherche avec une revue détaillée de la littérature, et l'introduction du contexte scientifique nécessaire à la compréhension des systèmes étudiés, ainsi qu'une conclusion intermédiaire.

En détail, le chapitre 2 se concentre sur la réalisation d'un conformateur d'impulsions optiques autonome opérant dans la picoseconde et basé sur l'approche de synthèse de cohérence temporelle. Pour cette mise en œuvre, une puce optique existante est réutilisée en utilisant deux techniques d'optimisation guidée courantes, telles que les algorithmes génétiques et l'optimisation par essaims de particules. Une fonctionnalité clé pour une mise en œuvre réussie fut l'utilisation d'une méthode d'échantillonnage entièrement optique basé sur le mélange de quatre ondes dégénérées dans une fibre hautement non linéaire. Celle-ci permettait une mesure directe des différentes formes temporelles des impulsions optiques. Enfin, les performances de la mise en forme des impulsions sont étudiées en détail et une comparaison entre deux algorithmes d'optimisation couramment utilisés est effectuée.

Le chapitre 3 se concentre sur la mise en œuvre d'un réseau neuronal transitoire entièrement optique, seulement constitué de fibre, basé sur la structure de la machine à apprentissage extrême. Le système est d'abord décrit sur le plan théorique et installé en utilisant uniquement des composants de fibre disponibles sur le marché, tout en réutilisant la méthodologie développée dans le chapitre 2. La machine à apprentissage extrême exploite de manière unique la dynamique complexe de fission des solitons comme ressource de calcul sous un design efficace et compact. Les performances du système sont ensuite démontrées en exécutant plusieurs tâches de référence de réseaux neuronaux. De plus, la classification de test COVID-19 a été effectuée à titre indicatif d'application réelle de cette méthode.

Enfin, le chapitre 4 fournit un bref sommaire du travail présenté. Dans une perspective, les lacunes des systèmes actuels du domaine sont discutées et des solutions potentielles sont présentées.

D.3 Mise en forme d'impulsions sur puce reconfigurable

D.3.1 Contexte et méthodologie

Le contrôle des propriétés temporelles des impulsions optiques présente un grand intérêt pour différentes applications, car elles peuvent impacter considérablement les performances d'un système[121, 122]. Par exemple, l'efficacité des interactions non linéaires dépend fortement de la durée et de la forme de l'impulsion (parmi d'autres facteurs tels que l'adaptation de phase ou la puissance de

crête[123]). Augmenter l'efficacité est particulièrement intéressant lors d'applications spécifiques. Notamment, pour le traitement tout optique des signaux dans le domaine des télécommunications[65], l'efficacité des processus d'amplification paramétrique non linéaire dépend fortement de la forme d'onde du signal[124]. Contrairement, pour réduire les distorsions du signal pendant la transmission de données optiques, il est de mise de générer des impulsions paraboliques[125, 126]. D'autres exemples incluent des applications pour une spectroscopie améliorée[127], la génération optimisée de rayonnement Terahertz[128, 129] ou en optique quantique, où l'adaptation des impulsions optiques permet d'augmenter l'efficacité du couplage aux cavités optiques[130–132]. Ceci peut être d'un grand intérêt pour les plates-formes intégrées compactes, qui utilisent des microrésonateurs comme sources d'états quantiques[133, 134].

Initialement, le contrôle de l'enveloppe temporelle de l'impulsion s'accompagnait du concept de cartographie espace-fréquence [135, 136]. Ces premiers travaux se concentraient principalement sur la dispersion spatiale de l'impulsion entrante, puis sur la manipulation ultérieure des composantes spectrales individuelles (également connue sous le nom de synthèse de Fourier, comme illustré à la figure 2.4) pour obtenir les formes d'impulsions souhaitées en utilisant, par exemple, des modulateurs de lumière spatiaux [136, 137]. Bien que ce concept ait trouvé son application dans des dispositifs commerciaux[1] [138], il est limité au régime femto- et picoseconde en raison des exigences croissantes sur les éléments dispersifs, la praticabilité (c'est-à-dire la taille) et le coût. Par conséquent, il existe un intérêt technologique pour réaliser la mise en forme de formes d'onde picosecondes dans des plateformes efficaces et compactes, afin d'améliorer les performances d'applications existantes et futures.

Une approche alternative aux méthodes de synthèse de Fourier est la mise en forme d'impulsions basée sur la synthèse par cohérence temporelle (TCS) [142, 143]. Dans ce cas, la forme d'onde cible est générée par la superposition cohérente de copies d'impulsions retardées et pondérées en amplitude (Δt_n et a_n dans la Fig. 2.4). L'un des avantages de la TCS est qu'elle peut fonctionner sans avoir à concevoir spécifiquement un système, contrairement aux méthodes susmentionnées, car la TCS utilise des interféromètres. Seuls l'accès aux délais relatifs et le contrôle de l'amplitude sont nécessaires (en espace libre). Cette méthode a été réalisée pour la première fois en 2007 [142] en utilisant une cascade d'interféromètres de Michelson en espace libre [**?**].

Pour la mise en œuvre d'un système de photoniques intelligentes, il est essentiel de disposer d'une lecture efficace et ultra-rapide de la sortie du système afin de l'utilisée pour optimiser vers une sortie cible. Pour l'applications de mise en forme d'impulsions (arbitraires), la lecture non ambiguë de l'information temporelle avec une précision suffisamment élevée est nécessaire. Pour cela, il n'existe qu'un nombre limité de techniques.
Plusieurs techniques ont été développées pour mesurer les impulsions optiques, allant de la détection directe à l'aide de photodiodes et de convertisseurs analogiques-numériques, à la reconstruction complexe plein champ des impulsions optiques [123]. Alors que le premier cas est limité aux équipements radiofréquence de pointe qui sont très couteux[2] (largeurs de bande analogiques d'environ 110 GHz), cette dernière est une technique assez complexe qui nécessite des algorithmes d'extraction supplémentaires, ce qui peut entraîner des mesures lentes [169]. De plus, les méthodes de reconstruction plein champ ne fonctionnent souvent que sur des intervalles de temps limités, généralement dans le domaine de la femtoseconde jusqu'à quelques pico-secondes ($< 10\,\mathrm{ps}$). Néanmoins, les premières

[1] (par exemple, les Waveshapers de Finisar)

[2] pour capturer la véritable forme de l'impulsion, il faut des détecteurs dont le temps de réponse est nettement inférieur à la largeur de l'impulsion mesurée [159]

mises en œuvre de mise en forme d'impulsions ont souvent utilisé ces techniques [39, 71]. D'autres méthodes, comme l'autocorrélation [159], sont peu coûteuses et simples à mettre en œuvre. Elles ont donc été utilisées pour des applications d'apprentissage automatique, notamment dans le contexte des lasers intelligents [32, 72]. Cependant, une autocorrélation renvoie toujours un signal symétrique et présente donc une ambiguïté quant à la forme d'onde, c'est-à-dire que différentes formes d'onde de sortie peuvent donner lieu à des traces d'autocorrélation identiques ou très similaires.
En conclusion, afin d'utiliser efficacement les techniques d'apprentissage automatique dans les plates-formes expérimentales, il est nécessaire de mettre en place des systèmes de détection rapides, simples, non ambigus et précis.
L'échantillonnage optique (voir Fig. 2.10) est une méthode simple, mais puissante pour mesurer des impulsions picosecondes indépendamment de la phase de l'impulsion. En raison de la simplicité de ces montages et de la grande largeur de bande de détection réalisable ($\gg 100\,\mathrm{GHz}$), ils ont été largement étudiés en tant que récepteurs dans les applications de télécommunications [171, 181]. Cependant, leur utilisation dans des systèmes de photoniques intelligentes n'a pas encore été étudiée.
L'élément central du schéma d'échantillonnage est une porte ET (optique) qui combine un signal d'échantillonnage et le signal testé. En optique, une impulsion d'échantillonnage et l'impulsion cible sont superposées dans un milieu non linéaire, à partir duquel les processus paramétriques génèrent un nouveau signal proportionnel au signal cible. Afin de reconstruire l'impulsion originale, l'impulsion d'échantillonnage doit être balayée sur l'impulsion cible. Ceci peut être réalisé en accordant un délai ou, plus simplement, en désaccordant les taux de répétition du signal cible par rapport à l'impulsion d'échantillonnage, ce que l'on appelle l'échantillonnage asynchrone [171, 173, 174]. Ce dernier a été utilisé dans ce travail, où la source d'échantillonnage et la source de mise en forme étaient désaccordées par un multiple entier plus une petite différence. Le signal nouvellement généré est généré à un taux beaucoup plus faible, qui correspond au désaccord entre l'impulsion cible et l'impulsion d'échantillonnage (ou des multiples plus une petite différence [174]). Ainsi, l'impulsion cible peut être mesurée à l'aide d'électronique peu coûteuse et à faible bande passante ($\ll$GHz).
Une porte ET optique peut être réalisée de différentes manières, comme une fibre hautement non linéaire utilisant des effets paramétriques, par ex., mélange à quatre ondes [171, 172] ou modulation de phase [182], ou divers effets dans les amplificateurs optiques à semi-conducteurs [183–187] (par exemple, modulation de polarisation croisée et de gain, ou absorption à deux photons). Dans le travail présenté, tout l'échantillonnage optique a été réalisé en exploitant le mélange dégénéré à quatre ondes dans une fibre hautement non linéaire [40].
En exploitant la non-linéarité paramétrique de la fibre hautement non linéaire, deux photons de pompe et un photon de signal peuvent être convertis en un photon idler symétrique à la pompe, comme illustré sur la figure 2.11. La génération du photon idler à un point d'échantillonnage donné N est proportionnelle au carré de la puissance de la pompe [189–191]. Pour la caractérisation du schéma d'échantillonnage mis en œuvre, une source de pompe picoseconde (Pritel FFL) avec une durée d'impulsion de $\approx$20 ps a été utilisée. Le signal obtenu à partir du schéma d'échantillonnage optique réalisé a ensuite été comparé à un autocorrélateur commercial (Femtochrome FR-103XL) afin d'évaluer ses performances, comme le montre la figure 2.13. Dans l'ensemble, les largeurs d'impulsion récupérées à la demi-largeur maximale (FWHM) correspondent très bien (($\Delta\tau_{FWHM,1,AC} = 21.85\,\mathrm{ps}$ et $\Delta\tau_{FWHM,1,Sampling} = 21.26\,\mathrm{ps}$)).

La commutation rapide dans les guides d'ondes confinant la lumière, promet des applications efficaces et nouvelles en termes de photonique intelligente avec une empreinte très compact grâce à l'utilisation de photonique intégrée interconnectée électroniquement. La puce photonique utilisée dans ce travail a été conçue à l'origine comme un commutateur optique [161]. Elle a ensuite été réutilisée avec succès, en combinaison avec des techniques d'optimisation, pour permettre la

génération de lumière non linéaire personnalisée [42].
La puce optique est basée sur du Hydex (voir Tableau 2.1 pour les spécifications), un matériau propriétaire aux propriétés comparables à celles du silicium-oxy-nitride (SiON) [162, 163] et est une plateforme de matériau compatible avec les CMOS (complementary-metal-oxide-semiconductor) qui permet de produire en masse des dispositifs dotés de vastes fonctionnalités (commutateurs thermo-optiques [161], résonateurs microanneaux [162], guides d'ondes en spirale (>50 cm) [163]) sur une petite surface. L'échantillon spécifique est constitué d'interféromètres concaténés, qui peuvent être ajustés électroniquement pour offrir des retards croissants par bit.

Afin de contrôler efficacement les rapports de division des interféromètres, deux algorithmes d'optimisation différents sont étudiés. En particulier, l'algorithme génétique (GA) et l'algorithme d'optimisation par essaim de particules (PSO) sont étudiés du point de vue de leurs performances, telles que la fidélité de la forme d'onde réalisable et le comportement de convergence. Le principe sous-jacent de l'algorithme génétique est l'exploitation de la fonctionnalité inhérente à l'évolution [88], telle que la création de descendants d'une population initiale en vue d'améliorer le système biologique, c'est-à-dire l'objectif du système. Grâce à la génération et au test itératifs de différents ensembles de gènes (populations), l'algorithme est capable de trouver des points optimaux dans des espaces de recherche hautement dimensionnels, comme l'illustre la Fig. 2.2. Le PSO, quant à lui, s'inspire du comportement social des essaims d'animaux [109]. L'algorithme fonctionne de manière similaire à la façon dont les bancs de poissons ou les essaims d'oiseaux recherchent de la nourriture. Le PSO est particulièrement intéressant en optique, car il peut optimiser des problèmes de grands paramètres sans connaissance des gradients, à un coût de calcul minimal par rapport à d'autres algorithmes, et son utilisation est généralement plus intuitive que celle des AG, par exemple. Contrairement à l'AG, le PSO repose sur des multiplications vectorielles simples (voir eq. 2.1), et non sur des opérations élaborées comme l'AG (c'est-à-dire la mutation ou le croisement). La principale fonction objectif utilisée dans ce travail est une similarité en cosinus modifiée (voir eq. 2.12), qui compare la forme d'onde renvoyée avec la forme d'onde cible. En outre, pour une comparaison quantitative entre différentes formes d'onde et différents paramètres, un facteur de qualité (Q) est introduit (voir eq. 2.14).

La configuration complète de la mise en forme intelligente des impulsions est décrite dans la Fig. 2.14. La source laser est un laser à fibre commerciale (Pritel FFL) et la source d'échantillonnage est un peigne de fréquence commercial (Menlo FC1500-250-WG). La sortie de la source laser est directement introduite dans l'un des ports d'entrée (In_1) de la puce optique. Compte tenu de la courte durée de l'impulsion d'entrée (environ 20 ps) par rapport aux délais disponibles, le nombre de délais utilisés a été limité à $N = 5$ (c'est-à-dire six commutateurs correspondant à 1 ps à 16 ps) et un miroir de Faraday couplé à une fibre a été utilisé pour réfléchir la forme d'onde à la sortie de la puce (Out_1), c'est-à-dire que les formes d'onde se propagent à travers la puce deux fois [139]. Cela augmente à son tour la plage de retard disponible et les combinaisons dans la plage de la durée de la pompe. Les connexions de la puce utilisées sont indiquées comme définies dans la Fig. 2.5. Après avoir quitté l'échantillon (In_2), l'impulsion mise en forme est combinée dans un coupleur de fibre standard 2×2 3 dB avec le signal provenant de la source d'échantillonnage avant d'entrer dans la fibre hautement non linéaire. Enfin, le signal parasite généré est filtré et amplifié avant d'être détecté.

D.3.2 Résultats

Les résultats expérimentaux pour le cas de formes d'onde non contraintes (c'est-à-dire que la largeur de bande passante de la forme d'onde peut varier) sont illustrés à la Fig. 2.15. Dans cette configuration, l'algorithme renvoie la forme d'onde qui présente le meilleur chevauchement avec la fonction cible, quelle que soit la largeur d'impulsion. Les paramètres utilisés pour l'algorithme PSO sont résumés dans le Tableau 2.4. Comme le montre la Fig. 2.15, une large gamme de formes d'onde cibles peut être obtenue avec l'échantillon de puce utilisé, allant jusqu'à plus de 50 ps (dans le cas de la Fig. 2.15h) en utilisant seulement cinq délais. Cependant, l'absence de contrôle de phase supplémentaire entraîne des modulations notables dans les enveloppes des formes d'onde renvoyées (voir Fig. 2.21). La figure 2.17 montre les résultats simulés de la polyvalence et de l'évolutivité du conformateur d'impulsions mis en œuvre. Pour cela, trois durées d'impulsion d'entrée différentes et plusieurs durées d'impulsion de sortie choisies aléatoirement ont été choisies. Dans l'ensemble, une bonne fidélité de la forme d'onde peut être obtenue pour toutes les durées cibles testées, comme le montrent les formes d'onde les plus performantes de la Figure 2.17c-h. L'amélioration de la fidélité de la mise en forme grâce à l'inclusion du décalage de phase est évidente sur la figure 2.21. En effet, dans tous les cas, la fidélité de la forme d'onde a été améliorée à partir d'une faible augmentation de $1,36\%$ pour la forme d'onde triangulaire, jusqu'à plus de 13% pour le flattop, et les performances détaillées pour chaque forme d'onde cible sont résumées dans le Tableau 2.8.

Cependant, comme la mise en œuvre la plus courante dans les études expérimentales repose sur l'AG, les performances de PSO et de l'AG sont comparées dans cette section pour les deux cas, la simulation et l'expérience. Pour étudier les performances des deux algorithmes, le modèle de simulation a été utilisé dans un premier temps. Les paramètres de l'algorithme utilisés pour la comparaison sont résumés dans le Tableau 2.6. Pour la forme d'onde, une dent de scie négative a été sélectionnée avec une FWHM non contrainte, et comme critère d'évaluation général, la valeur de perte des résultats de la Fig. 2.15g a été utilisée ($f_{loss} = 1.1351$, Q=8.81) comme critère de seuil. En utilisant ce seuil, la performance de l'algorithme a été évaluée comme le point où le seuil a été dépassé en termes de temps d'exécution et de populations requises. En outre, pour déterminer si les algorithmes (GA et PSO) présentent un quelconque avantage, l'étude a également été répétée avec un algorithme de recherche aléatoire (Kolen2010, Hendrix2015). Une recherche aléatoire est une approche utile où un nombre M de jeux de paramètres aléatoires sont générés et testés sur N itérations et le meilleur résultat est conservé. Cela permet d'évaluer si en choisissant des paramètres aléatoires à plusieurs reprises, les mêmes résultats qu'un algorithme élaboré peuvent être obtenus. Pour la recherche aléatoire, M=60 et N=100 ont été utilisés, fournissant ainsi un nombre de paramètres testés similaire à celui des deux algorithmes d'optimisation. La figure 2.18 montre les résultats de la valeur de perte pour les deux tailles de population et d'essaim différentes. Comme on peut le constater, les deux approches algorithmiques surpassent de manière significative la recherche aléatoire (Fig.2.18 inset), validant ainsi l'utilisation d'approches d'optimisation élaborées. En revanche, les différences entre les deux algorithmes sont moins importantes. On peut voir que dans tous les cas, le PSO converge plus rapidement (c'est-à-dire qu'il atteint plus vite les valeurs de perte les plus faibles). Ceci est particulièrement important dans les implémentations expérimentales où le temps total de mesure et de détection varie avec le nombre d'individus testés. Les performances des algorithmes sont présentées dans le tableau 2.7. Notamment, l'OSP est plus performant pour les petites populations par rapport à l'algorithme génétique, tant au niveau du temps d'exécution que de la valeur de perte finale, tandis que pour les grandes populations, les valeurs de perte finale sont presque égales. Les résultats expérimentaux de mise en forme pour les deux algorithmes sont présentés dans la Fig. 2.19. Les deux algorithmes sont globalement performants, cependant le PSO

atteint une fidélité de forme d'onde légèrement meilleure (Q=9.08, $f_{loss} = 1.1010$) par rapport à l'AG (Q=8.68, $f_{loss} = 1.1527$). Comme dans les simulations précédentes, le PSO a utilisé toutes les itérations de décrochage (2000) alors que le GA a nécessité toutes les itérations (15). Par rapport au nombre total de populations testées (9800 pour le PSO, 7500 pour le GA), le PSO confirme les résultats de la simulation, à savoir qu'il converge également plus rapidement dans l'expérience (meilleure performance pour le même nombre d'itérations).

Des tendances similaires ont également été signalées pour d'autres plates-formes intégrées [116] ou pour la conception inverse de composants [197]. Le principe fondamental des algorithmes peut expliquer ce phénomène. La petite taille de l'essaim dans le cas du PSO conduit à une exploration plus rapide de l'espace de recherche, tandis que la petite taille de la population de l'AG manque de diversité génétique, d'où une faible exploration. Il faut noter à ce stade que cela dépend généralement beaucoup du problème (des exceptions ont été trouvées dans le passé), mais dans le cas de problèmes inconnus ou sans contraintes [195], le PSO peut être un meilleur choix que le GA.

D.3.3 Conclusion

En conclusion, une nouvelle approche pour une mise en forme intégrée d'impulsions picosecondes continues basée sur la synthèse de cohérence temporelle a été étudiée et démontrée avec succès pour la première fois. Le dispositif est basé sur une concaténation d'interféromètres Mach-Zehnder contrôlables individuellement sur une plateforme en verre à faible perte.

En particulier, la mise en œuvre d'un schéma d'échantillonnage tout-optique rentable a été un élément clé pour l'acquisition rapide et sans ambiguïté de données à faible consommation d'énergie et à faible coût d'équipement. En tant que tels, les diagnostics basés sur l'optique non linéaire peuvent être un outil précieux pour le domaine de recherche émergent de la photonique intelligente afin de permettre une application hors laboratoire à un coût et une complexité réduits.

Les résultats obtenus indiquent que, malgré l'utilisation de la seule optimisation du rapport d'amplitude, de bons résultats de mise en forme des impulsions peuvent être obtenus pour une large gamme de durées d'impulsion d'entrée et de sortie. En effet, les simulations montrent que la plate-forme peut atteindre des durées d'impulsion de sortie de plus de 100 ps (FWHM) avec seulement cinq retards. Il a également été démontré que l'inclusion d'un contrôle de phase supplémentaire est capable d'améliorer davantage la forme d'onde au prix d'une plus grande complexité du système (c'est-à-dire, plus d'électronique comme les DACs nécessaires).

Comme nous l'avons démontré, l'utilisation d'un algorithme simple d'optimisation par essaims de particules peut surpasser l'algorithme génétique couramment utilisé pour un problème non contraint (c'est-à-dire qu'il n'existe pas de bonne estimation initiale). La convergence plus rapide du PSO peut être particulièrement intéressante pour l'optique ultra-rapide (par exemple, les lasers intelligents), où la dynamique du système est rapide. Néanmoins, il convient de souligner qu'il n'y a pas de généralité dans ce domaine. L'utilisation d'un algorithme plutôt qu'un autre dépend souvent fortement du problème et doit être étudiée en détail pour un problème spécifique [111, 195, 196]. De même, une combinaison de différentes approches peut convenir à des problèmes plus complexes [211, 212].

D.4 Réseau neuronal tout-optique

Les concepts d'apprentissage automatique, en particulier ceux basés sur les réseaux neuronaux (RN), sont le principal moteur d'une "quatrième révolution industrielle" [1], qui a déjà un impact sur le traitement de l'information dans un large éventail de technologies [2, 215, 216]. En effet, les réseaux neuronaux sont convaincants pour trouver des modèles cachés dans de grandes quantités de données, pour interpoler à partir de ces données ou pour reconnaître des caractéristiques spécifiques. Plus précisément, les réseaux neuronaux permettent d'automatiser des processus statiques et dynamiques et de développer de nouvelles applications "intelligentes", par exemple pour le traitement du langage naturel [217], le diagnostic medical [218] ou les véhicules autonomes [219], pour n'en citer que quelques-unes.

D.4.1 Contexte

L'épine dorsale de la plupart des réseaux neuronaux est constituée par les neurones artificiels (AN), qui visent à imiter le comportement des neurones d'un cerveau (humain). En bref, les neurones biologiques reçoivent des informations d'entrée (pondérées), qui sont ensuite additionnées et finalement traitées de manière non linéaire avant d'atteindre le ou les neurones suivants. Une distinction importante entre un neurone biologique et un neurone artificiel est le type de traitement non linéaire. Alors que les neurones biologiques fonctionnent généralement comme des neurones à pointes (c'est-à-dire qu'ils se déclenchent lorsqu'ils sont activés ou non) [231], un neurone artificiel peut présenter une plus grande variété de fonctions d'activation, allant d'un seuil "dur", comme les fonctions à pas du modèle de perceptron, à des fonctions plus continues (c'est-à-dire qu'ils se déclenchent un peu), comme les fonctions sigmoïdes [232] ou ReLU (rectifier l'unité linéaire) [233]. Dans le passé, les neurones uniques ont été essentiellement implémentés de deux manières. Soit comme un perceptron (c'est-à-dire en utilisant une fonction d'activation en forme de pas [256, 257]), soit comme un neurone artificiel avec une fonction d'activation différente. En fait, les deux concepts présentent d'importantes similitudes et diffèrent principalement par le choix de la fonction non linéaire et des algorithmes d'apprentissage, et sont donc souvent utilisés comme synonymes. La figure 3.1 illustre un seul neurone artificiel.
Il convient de noter qu'essentiellement n'importe quelle fonction (non linéaire) peut être utilisée comme fonction d'activation, mais dans le respect de certaines contraintes [258]. Des fonctions plus complexes ont été utilisées par le passé pour des problèmes spécifiques [259], mais elles nécessitent une puissance de calcul plus importante pour l'évaluation et l'apprentissage. Par conséquent, pour les mises en œuvre pratiques, on utilise souvent des fonctions non linéaires simples, mais suffisantes (telles que celles de l'éq. 3.3).

Bien que le neurone unique (et le perceptron) se soit avéré être un outil simple mais puissant dans le domaine de l'apprentissage automatique, il atteint ses limites lors de la gestion de tâches (significativement) non linéaires. Pour de telles tâches (par exemple la parité à n bits), un seul perceptron/neurone artificiel est incapable de résoudre la tâche [260]. C'est pourquoi des réseaux multicouches ont été développés [259, 261, 262]. Il convient de noter à ce stade que les perceptrons multicouches (MLP) et les réseaux neuronaux à action directe ne diffèrent à nouveau que par la fonction d'activation utilisée (activation discontinue ou continue, respectivement), bien que MLP soit souvent utilisé comme synonyme de réseaux multicouches. La figure 3.2. illustre une architecture générale de réseau multicouche. Comme dans le cas d'un neurone unique, dans un réseau

multicouche, chaque neurone reçoit un signal d'entrée pondéré, mais au lieu d'une sortie unique, ils sont connectés aux neurones de la couche suivante. Si tous les neurones sont connectés entre eux, le réseau est appelé réseau de neurones entièrement connecté, tandis qu'un nombre limité de connexions est appelé réseau de neurones clairsemé. L'optimisation des connexions est appelée "élagage" [238]. Si le nombre de couches de traitement (sans compter l'entrée et la sortie comme une couche, respectivement en bleu et en vert sur la figure) dépasse une couche, cette terminologie est ambiguë dans la littérature car certains comptent également les couches d'entrée et/ou de sortie, ou spécifient le nombre de couches indépendamment, par exemple, réseau à 2 couches, à 3 couches, etc.), on parle souvent de réseau neuronal profond, et des couches suivantes de couches cachées (les neurones d'une couche cachée étant appelés *unités cachées*) [232].

La non-linéarité joue un rôle central dans les implémentations de réseaux neuronaux. En l'absence de toute non-linéarité, tout réseau neuronal n'effectuerait qu'une mise en correspondance linéaire des données d'entrée, quelle que soit l'architecture ou la taille du réseau (par exemple, plusieurs couches linéaires peuvent être condensées en une seule couche linéaire). Ceci est d'autant plus important que la plupart des problèmes du "monde réel" sont non linéaires par nature, et que les réseaux linéaires ne sont donc pas en mesure d'effectuer des opérations significatives sur ces données (par exemple, pour les tâches de classification). Le choix de la fonction non linéaire est important en soi [258], car la majorité des algorithmes de formation (par exemple, la descente de gradient stochastique) nécessitent des fonctions différentiables pour une convergence optimale. Notamment, les réseaux non linéaires de très grande taille (et donc complexes) peuvent offrir des avantages supplémentaires, tels qu'une forte capacité de généralisation même avec des données de taille limitée ou l'évitement de l'utilisation de méthodes de régularisation [264].
Pour illustrer la capacité de séparation non linéaire du réseau neuronal, un réseau neuronal multicouche simple est mis en œuvre pour résoudre la tâche XOR[3], avec une couche cachée contenant deux unités cachées. L'activation des unités cachées est une fonction tangente hyperbolique ($\tanh(\cdot)$) et l'unité de sortie utilise une fonction sigmoïde. L'apprentissage est effectué pendant 100 époques. En évaluant la sortie et les poids des nœuds cachés après la formation, la transformation des données peut être visualisée (Fig. 3.3c). Cette petite expérience montre clairement que la non-linéarité des réseaux neuronaux a un impact considérable sur la transformation des données et constitue donc une caractéristique cruciale pour l'application réussie des réseaux neuronaux au traitement des données du monde réel.

L'objectif ultime de l'apprentissage du réseau est l'ajustement des poids $\mathbf{w}$ du réseau afin de minimiser l'erreur entre la sortie du système $\mathbf{Y}_{Output}$ et la vérité terrain $\mathbf{Y}_{Target}$ et d'atteindre la " généralisation " (c'est-à-dire la prédiction réussie de données non vues après l'apprentissage). De multiples facteurs doivent être pris en compte pour la formation: Le choix i) de la fonction d'activation non linéaire, ii) de la fonction de perte (par exemple, la métrique de la distance, la métrique de l'entropie, etc.), et iii) de l'algorithme d'apprentissage (par exemple, la rétropropagation des erreurs, la descente du gradient, la propagation de l'équilibre, etc.)
Aujourd'hui, la plupart des approches utilisent des méthodes basées sur le gradient, telles que la descente de gradient stochastique et la rétropropagation par erreur [269]. Cependant, l'utilisation de ces algorithmes nécessite que i) les fonctions non linéaires utilisées soient différentiables, et ii) l'accès aux nœuds individuels des réseaux soit possible. Il est important de noter que si la méthode de rétropropagation est très populaire, elle n'est pas un algorithme très efficace [270]. Dans les implémentations de réseaux neuronaux actuelles, la phase de formation est la partie la

[3]Exemple basé sur le code de `https://gist.github.com/CihanSoylu/6967249574192728a9fba367065e8949` (consulté le 27 janvier 2022)

plus inefficace et la plus longue [73]. Cela s'explique en partie par le fait que la rétropropagation est un concept " artificiel " (dérivé du calcul [239]) qui n'a pas d'équivalent biologique [270]. En tant que tel, l'inefficacité de la formation des réseaux neuronaux peut être partiellement attribuée aux concepts d'apprentissage choisis et des efforts importants sont déployés pour étudier des concepts de formation plus efficaces (qui présentent un intérêt particulier pour les réseaux neuronaux physiques) tels que l'apprentissage par l'équilibre [270, 271] ou le calcul de l'erreur directe [272]. En outre, une conséquence importante de l'utilisation de l'algorithme de rétropropagation est la nécessité d'un " accès " aux nœuds afin de calculer le gradient par nœud. Si ce problème est moins important pour les implémentations numériques, il peut s'avérer très difficile pour les implémentations de réseaux neuronaux physiques.

Enfin, un aspect important de la formation de tout réseau neuronal est le choix correct des hyperparamètres de formation, tels que les taux d'apprentissage η, la taille des échantillons/réseaux, les méthodes de régularisation ou de validation croisée [273, 274]. Si les hyperparamètres sont mal (ou pas du tout) choisis, le réseau neuronal présente des performances inadéquates. Les défauts de formation courants comprennent le sous-ajustement et le surajustement, comme l'illustre la Fig. 3.4 dans le cas d'un problème de régression [275]. De manière remarquable, des études récentes montrent que les réseaux non linéaires de très grande taille peuvent éviter le surajustement même sans régularisation [264]. Pourtant, la plupart des réseaux neuronaux pratiques fonctionnent dans un régime de paramètres (par exemple, le nombre de nœuds/couches) où ces problèmes sont apparents.

D.4.2 Réseaux neuronaux photoniques: État de l'art

Le développement de réseaux neuronaux photoniques n'est pas une innovation récente et suscite de l'intérêt depuis plusieurs dizaines d'années [276, 277]. La principale raison de l'étude des plates-formes photoniques réside dans les avantages prometteurs inhérents à l'optique, tels que la transmission passive, le haut degré de parallélisme (par exemple, l'exploitation des différents degrés de liberté de la lumière), les opérations de champ intrinsèques (par exemple, les multiplications vectorielles) et les non-linéarités intrinsèques des matériaux [70]. Ainsi, les réseaux de neurones physiques basés sur une plateforme photonique ont le potentiel pour un traitement du signal inspiré du cerveau tout-optique très efficace. La mise en œuvre réussie de tels réseaux (voir Fig. 3.5) présente un grand intérêt pour de nombreux domaines d'application.
Essentiellement, la mise en œuvre des réseaux neuronaux photoniques peut être classée en trois architectures principales (voir Fig. 3.5) en fonction de leur type de traitement de l'information: i) les réseaux de type feed-forward, ii) les réseaux de type récurrent et iii) les réseaux neuromorphiques. Les deux premières architectures de réseau sont bien connues de l'informatique, tandis que les implémentations neuromorphiques s'inspirent de concepts biologiques tels que les neurones à pointes. Bien que l'objectif ultime soit la mise en œuvre d'un réseau de neurones tout optique, la réalisation d'architectures de type feed-forward est particulièrement difficile. Bien que certains réseaux neuronaux entièrement optiques aient été démontrés pour des ordinateurs à réservoir [45, 47] et des réseaux neuronaux diffractifs [52], la plupart des travaux se concentrent principalement sur la mise en œuvre de blocs de construction spécifiques (soit les fonctions d'activation non linéaires) [298–300], ou encore sur des multiplications matricielles efficaces [59, 62, 63, 293][4]). En conséquence, la majorité des implémentations sont de facto de nature opto-électronique [298], ce qui ajoute un

[4]Une revue complète des différentes implémentations peut être trouvée dans les Refs. [16, 76–78]

goulot d'étranglement en termes d'efficacité, de vitesse de traitement et d'évolutivité [301], même si elles ont montré des résultats remarquables.

D.4.3 Réseaux neuronaux basés sur la physique

L'un des principaux problèmes des concepts proposés précédemment est la mise en œuvre de la non-linéarité et/ou de configurations complexes (c'est-à-dire des architectures multicouches). La plupart des implémentations susmentionnées (en particulier le feed-forward physique) sont des implémentations à une seule couche qui n'effectuent la multiplication matricielle que de manière optique, tandis que la transformation non linéaire est effectuée numériquement sur un ordinateur. Cela conduit à un goulot d'étranglement dans la plupart de ces approches, car les conversions OEO (optique-électrique-optique) sont intrinsèquement inefficaces (connues sous le nom de *"goulot d'étranglement électro-optique"*), ce qui pose des limites strictes à l'évolutivité du système [301]. Ainsi, les recherches récentes se concentrent sur la mise en œuvre de concepts qui présentent un comportement non linéaire intrinsèque du système. Cela inclut par exemple l'utilisation de matériaux non linéaires "artificiels" [302, 303], ainsi que des effets optiques non linéaires [56, 304]. Les avantages de ces systèmes sont les suivants: i) en général, une dynamique plus complexe (nécessaire pour des problèmes plus complexes), ii) une efficacité énergétique (la transformation des données se fait de manière passive pendant la propagation), et iii) une moindre complexité du système (la plupart des conceptions jusqu'à présent sont des conceptions passives et/ou à composant unique). Par conséquent, les réseaux tout-optiques promettent un traitement plus efficace et, à ce titre, les réseaux inspirés de la physique ont suscité un intérêt rapide [54, 302]. Cependant, ces mises en œuvre inspirées de la physique manquent souvent de représentations spécifiques des nœuds, ce qui nécessite des méthodes d'apprentissage différentes, que ce soit pour la conception du système ou pour l'apprentissage réel d'une plateforme [302, 303, 305].
Un type particulier de réseaux neuronaux inspirés de la physique est celui des réseaux neuronaux à base d'ondes (*Réseaux de Schrödinger* [56, 304, 306]), ou d'autres réseaux neuronaux à base d'ondes physiques [54, 302, 307, 308]). Ces types de réseaux rompent avec les conventions courantes car ils utilisent la dynamique complexe des ondes physiques pour le traitement neuromorphique analogique sans confiner artificiellement le matériel optique dans l'un des formats de réseau connus (réseaux neuronaux à action directe ou récurrente). En général, ces systèmes s'appuient sur des schémas d'apprentissage en entrée ou en sortie (tels que le cadre du machine learning extreme [54, 268]), qui peuvent être à la fois très efficaces en termes de temps et d'énergie, tout en offrant la puissance de calcul des réseaux neuronaux profonds dans une certaine mesure.
Des réalisations expérimentales récentes sont par exemple basées sur des approches multimodes spatiales dans des fibres [56] utilisant le couplage de modes non linéaires comme mécanisme clé. Ces travaux ont donné des résultats remarquables dans une variété de tâches d'inférence. Pourtant, les multimodes manquent de perspectives pour améliorer la compacité, la latence ou la consommation d'énergie. Parmi les autres approches, on peut citer les systèmes à espace libre (free-space systems) [57] qui ne présentent qu'une non-linéarité limitée (fournie par le gain de saturation de la photodiode) ou les approches de type "boîte noire" [304]. Cependant, si ces premières mises en œuvre ont donné des résultats prometteurs, elles restent limitées en termes de méthodologie et de compréhension des principes sous-jacents. Le rôle que la dynamique non linéaire complexe dans les guides d'ondes peut jouer dans la mise en œuvre de processeurs neuromorphiques basés sur la physique reste particulièrement flou.

D.4.4 Apprentissage automatique tout-optique par les ondes - Dynamique des solitons pour les réseaux neuronaux optiques transitoires

L'informatique ondulatoire semble être un concept prometteur pour la mise en œuvre de réseaux neuronaux optiques analogiques, mais elle présente souvent des inconvénients en termes de reconfigurabilité [56, 57], d'efficacité énergétique [56] ou de réalisation expérimentale [302, 309]. En effet, l'un des "chevaux de bataille" de l'optique non linéaire, les fibres/guides d'ondes hautement non linéaires, n'a été étudié que récemment [304]. Si ces premiers résultats semblent très prometteurs, ces mises en œuvre ne fonctionnent que dans des régimes très limités et/ou traitent le système comme une "boîte noire". Par exemple, elles n'utilisent que des bandes spectrales très limitées et reposent sur des transformations non linéaires uniques [304]. Ils n'utilisent donc pas la gamme rapide de dynamiques non linéaires disponibles sur ces plates-formes (par exemple, l'élargissement spectral non linéaire par soliton-fission), et manquent donc d'évolutivité en réseau sous la forme de nombreuses cases de sortie spectrale. Jusqu'à présent, la mise en œuvre de ces systèmes s'est largement concentrée sur des concepts spatio-temporels exploitant le couplage et la dynamique multimodes [56, 310], qui sont soit encombrants, soit difficiles à mettre à l'échelle et inefficaces sur le plan énergétique. À ce titre, l'exploitation de l'espace spectro-temporel présente un intérêt particulier car elle permet une mise à l'échelle dans un encombrement compact. En outre, ces procédés peuvent souvent être facilement adaptés pour fonctionner dans différents régimes de sortie [42, 311]. Par exemple, le système lui-même ne change pas si l'on considère plus de bacs spectraux de sortie ou si l'on modifie les paramètres des impulsions d'entrée, offrant ainsi un degré de liberté facilement évolutif. Les premières mises en œuvre ont utilisé le codage d'amplitude dans un régime opérationnel limité (bande L optique) [304], mais il manque encore une compréhension plus approfondie i) de la dynamique sous-jacente et de l'adaptation à des réseaux neuronaux plus "généralisés" et ii) des performances et de l'influence de la dépendance de phase.

La dynamique des solitons est un cadre particulièrement attrayant car elle offre des processus non linéaires très complexes qui peuvent être mis en œuvre par le biais de composants facilement accessibles et disponibles sur étagère [312, 313]. La forte dépendance de phase de ces dynamiques a également été étudiée récemment en vue de la mise en œuvre de portes logiques tout-optiques [315] ou de circuits informatiques optiques [316]. Par conséquent, les solitons ont également été proposés comme un outil fondamental pour l'informatique neuromorphique [54, 317]. Par conséquent, la compréhension et la mise en œuvre de tels systèmes neuromorphiques basés sur les solitons permettront une large gamme d'applications à faible consommation d'énergie, en particulier dans les domaines de la vision artificielle, de la détection et de la spectroscopie, où les sources laser ultra-rapides et les HNLF sont déjà largement répandues.

La génération d'un supercontinuum (SC) par l'utilisation de guides d'ondes hautement non linéaires (par exemple des guides d'ondes intégrés ou des fibres) fait partie des phénomènes non linéaires les plus complexes observés à ce jour dans le domaine des fibres optiques. Depuis la découverte de la génération de SC dans le verre massif [320] en 1970, des efforts considérables ont été déployés pour comprendre et modéliser la dynamique de ce phénomène non linéaire complexe. En détail, l'élargissement est obtenu par la SC, médiée par des effets tels que la non-linéarité de Kerr, la diffusion Raman, la modulation de phase croisée et la modulation de phase propre (XPM et SPM, respectivement), ou les instabilités de modulation (MI). La SPM permet de générer des solitons (d'ordre supérieur) [313] qui sont d'une grande importance pour le processus d'élargissement via la fission de solitons [312]. En effet, la fission de solitons est l'un des mécanismes d'élargissement les plus importants pour la génération de supercontinuum cohérent à large bande [312]. Essentiellement,

les perturbations pendant la propagation conduisent à une rupture du soliton supérieur qui entraîne l'éjection de solitons fondamentaux. Les termes de dispersion d'ordre supérieur assurent l'adaptation de phase de ces solitons, ce qui entraîne la génération d'un rayonnement non-solitonique (c'est-à-dire des ondes dispersives) du côté des basses fréquences (décalées en bleu). Au cours de la propagation, les solitons se déplacent davantage vers les basses fréquences (ce que l'on appelle le décalage auto-fréquence, voir Fig. 3.6) et, par conséquent, les ondes dispersives générées apparaissent également à des fréquences décalées, ce qui conduit à des spectres de sortie à très large bande. Il est important de noter que le décalage vers le bleu des ondes dispersives ne se produit que si le soliton et les ondes dispersives se chevauchent dans le temps [321, 322]. La cohérence est une propriété importante pour l'objectif de cette mise en œuvre spécifique, car la cohérence du supercontinuum décrit la stabilité d'une impulsion à l'autre et constitue donc une mesure de la reproductibilité spectrale. Comme une cohérence élevée implique que la sortie pour une entrée donnée reste la même même, même sous l'influence du bruit, cela fournit l'exigence de stabilité importante pour les applications d'apprentissage automatique (c'est-à-dire que la même entrée conduira à la même sortie).

L'outil fondamental pour la modélisation de la génération de supercontinuum est la propagation non linéaire des ondes basée sur l'équation non linéaire généralisée de Schrödinger (GNLSE), qui permet de modéliser avec précision la dynamique non linéaire tout au long de la propagation dans un milieu optique [191]. Une approche fiable et largement utilisée pour résoudre la propagation non linéaire d'impulsions optiques intenses est l'algorithme de Fourier à étapes fractionnées [189, 319], comme l'illustre la Fig. 3.7. En résumé, l'évolution d'une impulsion le long du guide d'ondes est divisée en une séquence alternée d'étapes linéaires et non linéaires.
L'approche de Fourier en deux étapes bien établie pour la modélisation de la propagation des impulsions dans les fibres optiques [191] ressemble aux fonctionnalités de base d'un réseau neuronal numérique, c'est-à-dire une séquence de transformations linéaires et non linéaires comme illustré dans la Fig. 3.8!.
En effet, des études récentes ont exploré cette analogie dans des cadres de télécommunication à grande echelle [307], ainsi que dans des études numériques sur la mise en œuvre de réseaux neuronaux optiques [309], mais il manque encore une démonstration expérimentale et un aperçu des régimes opérationnels potentiels. Dans le cadre du GNLSE, les étapes linéaires peuvent être décrites par la dispersion linéaire et l'opérateur de Fourier (correspondant à une sommation dans le domaine des fréquences) et l'étape non linéaire est décrite par l'opérateur non linéaire (une puissance cubique du champ dans le domaine temporel). Dans cette image, les poids du réseau correspondent à la constante de propagation dispersive bêta, le seuil d'activation non linéaire au coefficient non linéaire, et le biais à la distribution de puissance du champ entrant (avant masquage ou encodage avec les données).
Contrairement aux réseaux neuronaux classiques, la mise à jour active des poids d'un réseau neuronal optique à base de fibres est difficile, mais pas impossible à mettre en œuvre. Par exemple, la distribution de l'indice de réfraction d'une fibre peut être conçue par des simulations pour servir de poids afin d'exécuter des tâches spécifiques du réseau neuronal [309]. Cependant, même sans un tel apprentissage *in silico*, les paramètres des fibres disponibles dans le commerce (dispersion, coefficient non linéaire, etc.) imposent une pondération aléatoire intrinsèque pour l'étape dispersive (c'est-à-dire linéaire), ce qui permet d'utiliser des paradigmes d'apprentissage tels que les machines d'apprentissage extrêmes [268] (voir la section suivante pour plus de détails). Dans cette image, un réseau neuronal virtuel peut être construit sur la base de nœuds de réseau transitoires projetés dans un espace spectro-temporel bidimensionnel. Une entrée de champ optique dans le réseau peut être représentée comme une sélection de nœuds dans le temps et la fréquence portant des poids de

nœud non nuls, tandis que la dispersion et la non-linéarité du guide d'ondes définiront l'évolution du champ optique dans le temps et la fréquence, respectivement.

D.4.5 Réalisation expérimentale

Le concept et la configuration expérimentale permettant de vérifier les capacités de calcul neuromorphique de la fission de solitons sont illustrés dans la Figure 3.9. Tous les composants de la configuration sont des fibres à maintien de polarisation. Les données sont collectées à l'aide d'un analyseur de spectre optique standard (Ando AQ6317B) et lues par un ordinateur, qui est également utilisé pour mettre à jour les données d'entraînement sur le Waveshaper. En particulier, par rapport aux approches de modulation spatiale de la lumière (SLM), le nombre de cases de codage est limité avec un Waveshaper, mais ces dispositifs sont couplés à des fibres et disponibles dans le commerce, et ont prouvé leur fiabilité et leur stabilité à long terme dans le secteur des télécommunications. Pour la lecture spectrale, différentes configurations (résolution spectrale, points d'échantillonnage) ont été utilisées en fonction des tâches.

Optimisation des systèmes intelligents

Le régime de fission des solitons présente des exigences strictes en matière de durée d'impulsion afin de fonctionner dans un régime cohérent (c'est-à-dire répétable). Par conséquent, la dispersion du système de traitement doit être soigneusement étudiée afin de maintenir une largeur d'impulsion inférieure à environ 150 fs, ce qui constitue un défi en soi dans un système de fibres de plusieurs mètres de long. C'est pourquoi des DCF ont été utilisés dans le dispositif afin de réduire la durée d'impulsion à la limite de transformation donnée par la largeur de bande de la mise en forme d'ondes. Cependant, étant donné que la configuration implique des fibres PM de différents fabricants (ce qui entraîne des coefficients de dispersion d'ordre supérieur différents) et des longueurs de fibres internes inconnues (par exemple, à l'intérieur du laser ou de l'EDFA), une compensation exacte ne peut pas être obtenue par les DCF seuls. Comme l'illustre la Fig. 3.10a, la compensation avec un DCF ne permet d'obtenir qu'un délai d'environ 375 fs. Bien qu'il s'agisse déjà d'une bonne compression après plus de 20 m de fibres et une largeur d'impulsion initiale > 2 ps, ce n'est toujours pas suffisant pour la dynamique cohérente de fission de soliton [312, 319, 328]. Par conséquent, afin de comprimer davantage l'impulsion (à $<$150 fs), un schéma de compensation d'impulsion active est d'abord mis en œuvre. Le contrôle actif de la phase a été réalisé au moyen d'un dispositif de mise en forme des ondes (Waveshaper[329]), qui permet d'imprimer un profil de phase arbitraire pour compenser les termes de dispersion d'ordre supérieur.
Le masque de phase personnalisé a été implémenté à l'aide d'un polynôme d'ordre 5^{th}, et les coefficients du polynôme ont été optimisés à l'aide d'une optimisation par essaims de particules, semblable à celle utilisée dans le chapitre 2, pour atteindre la durée d'impulsion la plus courte sur un autocorrélateur, similaire aux approches précédentes avec GAs [36, 157]. Les connaissances acquises lors de l'utilisation d'un PSO pour la mise en forme d'impulsions optiques dans le chapitre 2, ont été réutilisées ici en combinaison avec une mise en forme d'impulsions commerciale (i.e., le Waveshaper). Les hyper-paramètres utilisés pour le PSO sont résumés dans le Tableau 3.2. La lecture de la durée d'impulsion a été effectuée au moyen d'un autocorrélateur commercial (Femtochrome FR-103XL) et d'un oscilloscope (Tektronix TDS5052B). La durée d'impulsion mesurée a été évaluée à l'aide d'un ajustement gaussien à deux termes. L'équation 3.12 montre la fonction de perte utilisée. Ici, FWHM est la largeur totale à mi-hauteur de l'ajustement, et Width_{1/e^2} est la largeur $1/e^2$ de l'ajustement. La FWHM est pondérée plus fortement dans la fonction de perte ($(\dots)^4$) et

une pénalité est ajoutée pour le fond sous la forme de la largeur d'impulsion $1/e^2$.
Par conséquent, une compression d'impulsion de 375 fs à environ 140 fs a été obtenue, comme le montre la Fig. 3.10b. Le profil de phase obtenu par la routine d'optimisation intelligente est illustré sur la Fig. 3.10c.

Caractérisation du système

Un aspect important de ce concept est la sensibilité des systèmes à la phase optique de l'impulsion d'entrée (c'est-à-dire le support d'information de la plate-forme développée). Afin d'étudier ce point, une expérience simple de "décalage de bits" a été mise en œuvre. Pour cela, un codage à un seul bit avec une largeur de bande de 1.0 nm (centré à 1553 nm, environ le centre de la pompe) a été utilisé. Ensuite, l'amplitude du bit unique a été modifiée de manière linéaire dans la plage $[0.1\pi, 1\pi]$, et les spectres correspondants ont été mesurés. La Fig. 3.11 illustre la sensibilité de phase du système pour des valeurs de phase croissantes.
Il est intéressant de noter que les effets du codage de phase ne sont que limités dans la région de la pompe (entre 1500 nm et 1600 nm), où des concepts récents sont utilizes [78]. Au contraire, le régime de rayonnement solitonique ($>$ 1600 nm) et non-solitonique ($<$ 1400 nm) présente une variation spectrale significative pour différentes valeurs de phase d'entrée. Cette forte dépendance de la phase de la sortie spectrale est un aspect fondamental de cette mise en œuvre. En particulier, le "mappage" non linéaire (c'est-à-dire que l'entrée de phase à la pompe affecte le spectre bien au-delà de la région de la pompe) est crucial pour la discrimination des problèmes séparables hautement non linéaires.
Pour étudier plus en détail l'impact de la phase, un balayage supplémentaire de la longueur d'onde du bit unique a été effectué. Pour ce faire, le bit a été déplacé de manière itérative dans la largeur de bande disponible de la mise en forme d'onde (environ 1528 nm à environ 1600 nm) avec un pas de 2.0 nm. Pour chaque position binaire individuelle, le spectre de sortie a été enregistré. Le balayage a été répété pour différentes valeurs de phase dans la plage $[0.1\pi, 1\pi]$, et les résultats pour quatre valeurs de phase différentes sont illustrés sur la Fig. 3.12.
D'après ces résultats, il est évident que pour des valeurs de phase relativement faibles (environ 0.1π), aucune variation spectrale significative ne se produit, quelle que soit la position du bit d'entrée. Une fois que la valeur de phase augmente, une variance plus élevée de la sortie spectrale peut être observée, indiquant un meilleur régime opérationnel qui permet potentiellement la distinction spectrale de différentes entrées. Cependant, il est important de noter que l'impulsion optique subit davantage de perturbations (dans le domaine temporel) lorsque la phase augmente, ce qui, en fin de compte, diminue la cohérence du système et donc sa précision. La Fig. 3.13 montre les traces d'autocorrélation pour des impulsions d'entrée codées avec différentes valeurs de phase. Pour cette mesure, un chiffre aléatoire provenant de l'ensemble de données MNIST (sous-échantillonné à 1×100, voir la section suivante pour plus de détails) a été codé avec deux valeurs de phase différentes (π et $\pi/8$) et comparé au cas de référence sans codage actif.
Comme l'indiquent les traces d'autocorrélation, pour des valeurs de phase relativement faibles (ici $\pi/8$), seuls de petits changements dans l'impulsion optique sont observables (le FWHM s'élargit légèrement de $\approx$ 140 fs à $\approx$ 155 fs). D'autre part, pour des valeurs de phase plus importantes (ici π), l'impulsion optique est affectée plus fortement, ce qui entraîne une augmentation significative de la largeur de bande passante (de $\approx$ 155 fs à $\approx$ 250 fs) et de plus, une division de l'impulsion de $\approx$ 1.5 ps est observable (une double impulsion entraîne des pics "latéraux" symétriques à l'impulsion principale dans l'autocorrélation).

Formation au reseau

Comme nous l'avons souligné au début de ce chapitre, l'apprentissage des réseaux neuronaux est généralement effectué par des algorithmes de rétropropagation d'erreur [215, 269]. Par conséquent, de nombreux homologues physiques suivent une stratégie similaire [16, 77]. Cependant, la mise en œuvre des méthodes de rétropropagation des erreurs dans les plateformes physiques est très difficile, car les gradients ne peuvent pas être mesurés directement [330–332]. Il est intéressant de noter que des concepts tels que les machines d'apprentissage extrême (ELM) ne nécessitent pas de simulations numériques élaborées pour l'apprentissage, mais utilisent plutôt des méthodes de régression linéaire directes [268]. Les ELM suivent un principe analogue à celui des méthodes de calcul par réservoir (reservoir computing [336, 337]), où un réservoir (ou réseau) très complexe de nœuds interconnectés de manière aléatoire est exploité pour la transformation des données, tandis que l'apprentissage est uniquement effectué par régression linéaire au niveau de la couche de sortie. La principale différence entre les implémentations RC et ELM est l'absence de mémoire de réseau intrinsèque. Les deux, RC et ELM, utilisent généralement des algorithmes simples tels que la régression linéaire [338] ou sa version régularisée, la régression dite de Ridge (également connue sous le nom de régularisation de Tikhonov) [273]. Cette dernière introduit un paramètre de régularisation λ afin d'éviter un ajustement excessif[5] (c'est-à-dire l'ajustement du comportement du bruit plutôt que de l'information). Par conséquent, dans la mise en œuvre proposée d'un ordinateur à ondes optiques, le cadre ELM est adopté pour la formation (c'est-à-dire la régression linéaire avec $\lambda = 0$ pour déterminer la matrice de poids de la couche de sortie).
La régression linéaire est combinée à une recherche directe des cases de lecture sur les spectres enregistrés (appelée *equal-search* dans ce qui suit) pour déterminer les positions des cases les plus performantes de la sortie large bande. La Fig. 3.14 illustre l'algorithme de recherche égale.

Pour un nombre donné de cases de lecture (déterminé par exemple par le nombre de classes ou trouvé par des recherches itératives), des cases équidistantes sont sélectionnées sur la base d'une longueur d'onde de départ définie (première position de case) et d'un espacement de case défini. Ces deux paramètres sont balayés de manière itérative sur les spectres d'entraînement enregistrés (voir Fig. 3.14bc). Pour chaque combinaison, la fonction de perte mise en œuvre (ici l'erreur quadratique moyenne - EQM) est évaluée à l'aide d'une validation croisée n-fold[274]. La validation croisée n-fold divise l'ensemble de formation en n lots égaux dont n-1 sont utilisés pour la formation et le lot restant est utilisé pour le test. Ensuite, les lots sont itérés et réévalués pour toutes les combinaisons, comme l'illustre la figure 3.15. Enfin, la combinaison de bacs la plus performante (en termes d'EQM) est sélectionnée. L'ensemble de test non vu est ensuite appliqué à cette configuration de lecture, c'est-à-dire $\mathbf{Y}_{Predicted} = \mathbf{W}_{out}\mathbf{X}_{Test}$. Pour les tâches de classification effectuées, un codage à un coup a été utilisé comme illustré dans la figure 3.16. Les colonnes individuelles correspondent à chaque classe et pour la récupération d'une prédiction unique, la sélection "*winner takes it all*" (c'est-à-dire $argmax(\mathbf{Y}_{Predicted})$) a été utilisée pour déterminer l'indice de la colonne présentant la valeur la plus élevée et donc la classe gagnante.

Résultats de calcul

Au fil des ans, une pléthore de tests de référence pour l'évaluation des réseaux neuronaux a été rendue publique. Afin d'évaluer le système développé et de permettre une comparaison avec d'autres implémentations optiques, un certain nombre d'ensembles ont été utilisés. En détail, l'évaluation comparative est divisée en deux types d'ensembles différents. Tout d'abord, il y a des tests plus fondamentaux ("benchmarks indépendants de la tâche") qui peuvent être effectués sans aucune bibliothèque de données préparée, comme le test de parité n-Bit[343, 344] ou l'approximation de fonction

[5]Pour $\lambda = 0$, l'équation représente les moindres carrés ordinaires

universelle[345]. Ces deux tests permettent d'étudier les aspects fondamentaux et l'évolutivité potentielle de la plate-forme développée. Ensuite, des tests plus orientés vers les applications ("benchmarks dépendants de la tâche") sont utilisés pour évaluer la plate-forme pour des types de tâches spécifiques comme la catégorisation basée sur des entrées numériques ou la classification d'images. Ainsi, trois tâches différentes couramment utilisées dans l'apprentissage automatique optique ont été testées[6]: le jeu de données IRIS de reconnaissance de fleurs[346], le jeu de données WINE[347], le jeu de données Abalone[348], ainsi que le célèbre jeu de données MNIST de reconnaissance de chiffres manuscrits[349]. Notamment, ces ensembles sont de complexité et de structure variables, ce qui permet également une étude plus rigoureuse des performances spécifiques.
Contrairement à des implémentations récentes [304], le codage de l'information de la plateforme démontrée est basé sur un codage spectral de phase et un fonctionnement en régime soliton-fission. Notamment, cette configuration évite toute perte de puissance due au schéma de codage. Les informations (c'est-à-dire les caractéristiques) sont converties en valeurs de phase spectrale, puis imprimées sur le spectre d'impulsion à l'aide d'un filtre programmable (Finisar Waveshaper 1000X). Le filtre programmable présente une largeur de bande de transmission couvrant les bandes optiques C et L, ce qui évite la troncature spectrale et donc l'élargissement de l'impulsion.

Repères indépendants de la tâche

Le problème de parité à n bits, c'est-à-dire la généralisation de l'opération XOR (voir l'équation 3.14), qui devient de plus en plus difficile à résoudre dans les réseaux neuronaux pour des longueurs de bits croissantes n[343, 344]. Notamment, la tâche XOR n'est pas soluble avec un seul perceptron[232]. En général, la précision de la résolution du test de parité sur n bits à l'aide de réseaux neuronaux est en corrélation directe avec le nombre de neurones de la couche cachée du réseau. Il a été constaté que le nombre de neurones peut être réduit en utilisant des non-linéarités plus sophistiquées que la sigmoïde ou la $tanh(\cdot)$[259]. *Donc, le problème de parité à n bits peut être considéré comme une mesure des capacités de séparation non linéaire d'un processeur neuromorphique et fournit un moyen direct de mettre à l'échelle la complexité du problème.*
Plusieurs itérations paramétriques du problème de parité à n bits ont été réalisées. Les performances ont été étudiées en fonction de la longueur binaire n, de la puissance optique d'entrée (en termes de nombre de solitons N, voir éq. 3.7) et du nombre de bins de lecture m. Les résultats sont présentés sur les figures 3.17 et 3.18. En détail, pour les résultats de la Fig. 3.17, le système a été entraîné pour différents niveaux de puissance d'entrée (exprimés par le nombre de solitons N), et une configuration de lecture correspondant à la longueur de bit d'entrée (c'est-à-dire *longueur de bit n = n bins de lecture*).
Il est évident, d'après la Fig. 3.17, que pour un réglage constant donné du système (c'est-à-dire la puissance d'entrée et la configuration de lecture), les performances se détériorent pour des longueurs de bits croissantes, comme le prévoient les travaux antérieurs [343]. Il est intéressant de noter qu'en augmentant seulement la puissance d'entrée du système, ce qui est équivalent à une augmentation de la non-linéarité du système, les performances peuvent être améliorées. Ce comportement est étudié plus en détail en augmentant le nombre de bins de lecture du *bin de lecture = longueur de bit* initial jusqu'à 125 bins de lecture pour deux configurations de puissance d'entrée différentes (faible non-linéarité N=2 et forte non-linéarité N=5). Les résultats de cette étude paramétrique sont présentés

[6]Bien que les références originales soient fournies, les ensembles de données ont été obtenus à partir du dépôt d'apprentissage automatique Irving de l'Université de Californie (IRIS, WINE, Abalone) à l'adresse suivante: `https://archive.ics.uci.edu/ml/index.php`, et de kaggle (MNIST) à l'adresse `https://www.kaggle.com/datasets/hojjatk/mnist-dataset`

sur la Fig. 3.18. Dans tous les cas, une augmentation du nombre de bins de lecture au-delà de la taille de la caractéristique d'entrée améliore considérablement les performances du système. Par conséquent, le nombre de cases de lecture utilisées pour la formation constitue un hyperparamètre supplémentaire. De plus, la Fig. 3.17 confirme la tendance selon laquelle une augmentation de la non-linéarité du système entraîne une augmentation des performances pour les tâches hautement non linéaires. D'un point de vue pratique, une réduction des cases de lecture peut être obtenue en augmentant la non-linéarité du système (c'est-à-dire que moins de cases sont nécessaires pour la même précision), ce qui est particulièrement important si des filtres programmables sont envisagés pour la formation [351], évitant ainsi les longs temps d'acquisition d'un OSA et accélérant donc le taux de traitement du système.

Un choix populaire dans l'informatique neuromorphique et l'apprentissage automatique pour l'étude du théorème d'approximation universelle des fonctions [345, 352], c'est-à-dire la capacité du réseau neuronal à représenter toute fonction complexe donnée. Pour les ELM, la régression sincère normalisée est souvent utilisée pour étudier ce théorème [54, 56, 353].
Ici, le codage a été effectué en utilisant un codage aléatoire (c'est-à-dire qu'un chiffre aléatoire de l'ensemble de données MNIST a été utilisé comme signal " de base ") avec des valeurs de phase aléatoires tirées d'une distribution uniforme dans la plage $[-\pi, \pi]$. Au total, 1000 valeurs de phase (=échantillons) ont été mesurées pour cette tâche.

Une erreur quadratique moyenne (RMSE) extraordinairement faible (pour les systèmes physiques) de 0.0656 a été obtenue, ce qui rend l'approche basée sur la fission de solitons similaire aux récentes approches ELM multimodales sur fibre [56] (RMSE de 0.0671). Cependant, le régime soliton-fission mis en œuvre fonctionne avec une énergie d'impulsion inférieure de deux ordres de grandeur (c'est-à-dire environ 90 pJ par rapport à 35 nJ). À titre de comparaison, les ELM numériques, en l'absence de bruit, peuvent atteindre une précision de test de RMSE = 0.0097 avec seulement 20 nœuds [353]. Ces résultats suggèrent que le concept de soliton-fission mis en œuvre peut être interprété comme un réseau neuronal complexe comportant au moins une couche cachée. La résolution du problème de parité à n bits nécessite un réseau à une seule couche cachée [343] ou un réseau multicouche [354] (s'il n'existe pas de connexions directes entrée-sortie). De même, la capacité d'approximation universelle de fonctions nécessite en général des réseaux multicouches [345, 352]. Notamment, la propagation à travers la fibre non linéaire ne peut être décrite par une seule fonction [191, 319] ni par une seule couche cachée [80, 355]. *Par conséquent, la résolution du test de parité à n bits pour $N > 2$ et la démonstration de la capacité d'approximation de fonctions non linéaires par propagation non linéaire indiquent la présence d'une topologie de réseau multicouche, contrairement aux travaux antérieurs qui traitaient les systèmes à base de fibres comme des "boîtes noires" non linéaires ou des transformateurs [56, 304].*

Repères dépendants de la tâche

La tâche Abalone est une tâche de régression multiparamétrique de taille moyenne[348], contenant 4177 échantillons avec huit caractéristiques d'entrée. L'objectif de cette tâche est de prédire l'âge de différentes espèces d'escargots de mer. Initialement, l'âge de ces escargots est déterminé en coupant la coquille en deux et en comptant le nombre d'anneaux (comme pour un arbre). Bien que ce processus soit très long et invasif, la détermination de l'âge sur la base d'autres facteurs tels que la taille ou le poids constitue une alternative viable. En tant que tel, le jeu de données présente huit caractéristiques d'entrée telles que le sexe (codage binaire), la longueur/diamètre/taille et plusieurs

mesures de poids dans le but de déterminer l'âge.
Pour permettre la comparaison avec les implémentations précédentes [56, 57], les étiquettes des données ont été normalisées dans la plage [0,1]. En outre, pour permettre des temps de mesure et une précision raisonnables (voir la discussion autour de la Fig. 3.22), la taille de l'ensemble de données a été réduite à 2000 échantillons (y compris l'âge minimum et maximum). Enfin, la RMSE a été calculée pour l'évaluation des performances. La Fig. 3.20 montre les performances obtenues pour la tâche de prédiction de l'âge des ormeaux. Une erreur quadratique moyenne de 0.0766 a pu être obtenue pour l'ensemble de test, ce qui est nettement mieux que les approches précédentes (erreur quadratique moyenne d'environ 0.12 dans le cas de [56, 57]). De plus, par rapport aux systèmes multimodes spatiaux [56], l'énergie d'impulsion requise est nettement inférieure (90 pJ contre 35 nJ), ce qui souligne l'efficacité énergétique de l'approche soliton-fission. Cependant, alors que l'approximation des données fonctionne bien pour les jeunes âges, on peut voir qu'après environ 0.6, il y a une plus grande déviation. En effet, une telle déviation a également été observée dans des études antérieures [56] et pourrait être causée soit par les données elles-mêmes (c'est-à-dire le manque d'échantillons d'âge élevé), soit par un manque de complexité du système.

La reconnaissance de fleurs IRIS est un point de référence très pertinent pour tester de nouveaux classificateurs [346]. Il consiste à mesurer la longueur et la largeur des pétales et des sépales de trois espèces différentes de fleurs d'Iris: Iris Setosa, Iris Versicolor, Iris Virginica. *Bien que l'ensemble de données soit relativement petit, avec seulement 150 échantillons et 3 classes, il s'agit d'un problème non linéairement séparable (c'est-à-dire que deux classes peuvent être linéairement séparées, alors que la troisième ne l'est pas).* Par conséquent, il s'agit d'un repère commun utilisé dans les implémentations physiques [95, 294, 356–358].
Les caractéristiques individuelles ont été normalisées par maximum d'ensemble par caractéristique (c'est-à-dire caractéristique/max(caractéristique)). La Fig. 3.21(a,b) montre les résultats de la classification IRIS, où la diagonale fait référence aux classes correctement identifiées, tandis que les éléments hors diagonale correspondent aux fausses prédictions. Il est remarquable que le système atteigne une précision de test de 100 % pour les données non vues. Cette forte séparabilité des données provient probablement de la transformation hautement non linéaire qui suit la dynamique complexe de la propagation des impulsions.

La tâche Vin[347] est une référence populaire pour les nouveaux classificateurs car il s'agit d'un problème bien posé, bien que non difficile. Elle comprend 178 échantillons avec 13 caractéristiques d'entrée (telles que le taux d'alcool ou le taux de phénol) et l'objectif est de classer trois types de vin différents en fonction des caractéristiques d'entrée fournies. *La tâche WINE est un type de problème de classification populaire pour les nouveaux classificateurs car elle est considérée comme un problème bien posé, peu complexe et avec une bonne distribution des classes.*
Pour l'ensemble de données WINE, la même normalisation que pour l'ensemble de données IRIS a été appliquée. La Fig. 3.21(c,d) montre les résultats de la classification WINE. Comme pour le test IRIS, une précision de 100 % a pu être obtenue pour les données non vues, bien que les données d'entraînement contiennent une classification erronée dans ce cas. Néanmoins, le système mis en œuvre présente des performances de pointe pour ces tâches et, notamment, aucun système optique à ce jour n'a atteint une précision aussi élevée pour l'un ou l'autre de ces ensembles de données (voir App. B.2).

Une tâche plus complexe (en termes de dimensionnalité d'entrée et de sortie) est la reconnaissance de chiffres manuscrits MNIST. Les chiffres manuscrits sont basés sur l'ensemble modifié de l'Institut national des normes et de la technologie (cet ensemble est communément appelé MNIST).

L'ensemble contient un total de 70000 chiffres manuscrits (60000 images d'entraînement et 10000 images de test) des chiffres 0 à 9[349]. Bien que cette tâche ne soit pas très difficile pour les approches de réseaux numériques (qui atteignent jusqu'à 100 % de précision), elle constitue une référence établie pour les nouveaux systèmes physiques. *Notamment, il ne s'agit pas d'un ensemble de données très non linéaires et les classificateurs linéaires peuvent déjà atteindre une précision élevée[359].*
Afin d'étudier l'effet des différents paramètres du système sur les performances, une recherche itérative des paramètres a d'abord été effectuée. Pour cela, un ensemble de données limité à 3000 images aléatoires a été sélectionné, tout en assurant une distribution égale entre les classes. La Fig. 3.22 résume les résultats. Cette étude multiparamétrique permet de constater que les performances du système dépendent de plusieurs paramètres. Notamment, dans l'implémentation actuelle du système, un compromis supplémentaire existe entre la performance du système et le temps d'acquisition des données. Par conséquent, ce n'est pas la meilleure résolution (0.125 nm, temps d'acquisition des données d'environ 2 minutes pour un intervalle de 550 nm) qui a été choisie pour les tests de référence, mais des résolutions plus faibles (0.5 nm, temps d'acquisition des données de 1 minute) pour permettre un temps de mesure raisonnable. Enfin, la Fig. 3.23 montre les résultats de la tâche MNIST pour le jeu de paramètres optimisé (amplitude de phase de $\pi/8$, 94 bins de lecture, résolution de 0.5 nm).
Le système a atteint une performance d'environ 87 % pour les données de test non vues, ce qui correspond à la moyenne des systèmes optiques. Une explication potentielle de la moindre performance dans cette tâche pourrait être la forte transformation non linéaire des données due au processus de fission des solitons. En effet, le MNIST est déjà soluble à un haut degré avec un classificateur numérique linéaire [359]. Il est remarquable de constater que les plus hautes précisions rapportées pour les systèmes optiques ne présentent que de faibles transformations non linéaires, comme les photodiodes [57, 360]. Cela suggère que i) la tâche du MNIST est de faible non-linéarité et ii) la forte non-linéarité médiée par le processus de fission des solitons pourrait être préjudiciable aux problèmes de faible non-linéarité. En effet, un manque de linéarité dans le système peut être l'une des principales raisons de la mauvaise performance dans la tâche MNIST, comme des études précédentes l'ont mis en évidence pour les systèmes basés sur la génération de seconde harmonique [308]. En outre, l'algorithme de "recherche égale" peut également imposer des restrictions quant aux solutions potentielles, limitant ainsi les performances réalisables. Dans ce cas, un algorithme de recherche non équidistant plus complexe pourrait contribuer à améliorer les résultats.

Afin de démontrer la performance de cette approche pour un problème du monde réel, le système est appliqué à une tâche de classification audio dans le cadre de la surveillance de la santé. En effet, l'utilisation d'enregistrements audio pour l'évaluation de la santé est une application très souhaitée [362], offrant une méthode peu coûteuse et continue pour la surveillance de la santé à distance et décentralisée. La reconnaissance autonome d'échantillons audio est particulièrement intéressante pour l'identification de maladies respiratoires comme les infections COVID-19 [363–365], où l'utilisation de telles méthodes peut accélérer les tests, tout en réduisant les analyses coûteuses comme les radiographies ou les tomographies par ordinateur [365]. En particulier, les machines à vecteurs de support (SVM) ont montré des performances remarquables pour cette tâche [363, 366, 367] et présentent également un grand intérêt pour les applications à distance (par exemple, pour une utilisation sur les téléphones mobiles [368]) compte tenu de leur efficacité et de leur complexité réduite [232, 369].
Pour la démonstration particulière du système basé sur la fission de solitons pour la classification audio, le jeu de données fourni par le défi *INTERSPEECH 2021* est utilisé et les performances sont comparées à une implémentation récente de SVM [367]. Le jeu de données est constitué

de 856 échantillons audio (étiquetés) provenant d'individus avec et sans infection par le COVID-19. L'objectif de cette tâche est d'identifier les individus infectés par rapport aux individus sains uniquement sur la base des échantillons audio fournis. Dans ce qui suit, la méthode d'extraction des caractéristiques est brièvement décrite (voir Fig. 3.24): Tout d'abord, le signal audio brut a été converti en un spectrogramme au moyen d'une transformée de Fourier à court terme (STFT) générant une carte bidimensionnelle (ou image). Le spectrogramme représente les différentes composantes de fréquence à différents moments de la parole. Par la suite, une autre transformée de Fourier a été mise en œuvre pour générer ce que l'on appelle des spectrogrammes de modulation, qui sont particulièrement adaptés aux données du monde réel (c'est-à-dire aux données affectées par le bruit environnemental). Le spectrogramme de modulation représente le taux de changement des composantes de fréquence individuelles pendant la parole. Pour fournir au système des caractéristiques utilisables (c'est-à-dire des caractéristiques de taille < 400), le spectrogramme de modulation est binné en utilisant dix bins (taille de 1 Hz) pour la fréquence de modulation (f_{mod}) et vingt bins (taille de 400 Hz) sur la fréquence conventionnelle (f), générant la carte finale 20×20. Enfin, une extraction de caractéristiques en deux étapes est appliquée pour extraire 30 caractéristiques des cartes. Dans un premier temps, des descripteurs spectraux de bas niveau (LLD) tels que l'entropie, le centroïde et le kurtosis ont été extraits des cartes générées, ce qui a donné 320 LLD par signal audio ($2 \times (8 \times 20)$). Dans un deuxième temps, 30 caractéristiques ont été extraites des 720 LLD présélectionnés (320 LLD plus 400 énergies du spectrogramme de modulation binné) au moyen d'une analyse en composantes principales [375, 376]. Les caractéristiques extraites ont ensuite été normalisées (mise à l'échelle des caractéristiques min-max) et codées de manière similaire aux tâches précédentes sur la bande C optique, puis les données (c'est-à-dire les spectres de sortie pour chaque échantillon COVID) ont été mesurées. La Fig. 3.25 montre les résultats obtenus. De manière remarquable, le système optique (0.7133) surpasse le SVM implémenté (0.6309, voir Annexe B.3), et dépasse également la valeur de référence (0.709) pour cette tâche. Une valeur pour l'UAR supérieure à 0.5 indique un résultat supérieur à une supposition aléatoire. La valeur de référence pour le défi COVID-19 est de 0.709 pour les données vocales [366].

D.4.6 Conclusion

En résumé, un nouveau type de processeur optique basé sur une dynamique de fission de soliton très complexe a été implémenté pour la première fois et caractérisé en utilisant différentes tâches de référence. Pour la majorité des tâches, le système présente des performances supérieures à l'état de l'art (voir le tableau annexe B.1 pour une comparaison). En particulier, le système semble bien adapté aux tâches hautement non linéaires ou complexes telles que la classification IRIS ou COVID-19, alors qu'il est moins performant pour les tâches plus linéaires telles que la tâche populaire MNIST. Les résultats obtenus pour toutes les tâches effectuées sont résumés dans le tableau 3.4. Un résultat particulièrement intéressant de ces tâches de référence est la position spectrale des bins de lecture. En effet, les positions semblent indiquer que les tâches plus complexes (par exemple, IRIS ou COVID-19) donnent lieu à des cases de lecture éloignées de la pompe en raison d'une forte transformation non linéaire (c'est-à-dire la génération d'ondes dispersives), tandis que les tâches plus linéaires (par exemple, WINE ou MNIST) les cases de lecture sont plus proches de la pompe (c'est-à-dire un régime plus linéaire). Cette découverte suggère qu'un tel système photonique pourrait même être utilisé pour évaluer le degré de non-linéarité pour une tâche donnée, ce qui est une information très utile pour la communauté de l'apprentissage automatique numérique également, car elle pourrait aider à estimer le nombre de couches cachées (c'est-à-dire la gamme dynamique

d'un NN) pour des applications particulières (ou des ensembles de données) et donc potentiellement réduire les complexités du réseau et la consommation d'énergie.

D.5 Conclusion et perspectives

En conclusion, ce travail a exploré la possibilité d'utiliser la photonique non linéaire, l'apprentissage automatique et les techniques d'optimisation guidée pour la mise en œuvre de deux dispositifs photoniques intelligentes basés sur des composants sur étagère très similaires et des dispositifs réaffectés.

En particulier, l'utilisation de techniques d'optimisation permet non seulement de générer des sorties de dispositif personnalisées (par exemple, la mise en forme d'impulsion autonome du chapitre 2), mais aussi de réutiliser des composants existants pour de nouvelles applications (par exemple, la même puce a été utilisée pour la génération de lumière non linéaire [42]). Il est intéressant de noter que la combinaison de ces méthodes d'optimisation avec l'optique intégrée pourrait permettre de créer de nouveaux outils pour améliorer les performances des dispositifs en réduisant l'impact de leur fabrication imparfaite. En effet, les stratégies visant à réduire l'impact des tolérances de fabrication sont d'un grand intérêt pour améliorer les performances des réseaux de neurones optiques et des applications d'optique quantique [153–155]. En tant que telles, les méthodes présentées peuvent fournir un outil essentiel dans le domaine de la photonique intelligente, où le succès pourrait reposer sur la mise en œuvre de solutions de contrôle peu coûteuses et peu complexes. Ces méthodes sont cruciales pour le déploiement réussi de ces techniques dans des scénarios du monde réel, où un fonctionnement abordable, compact ou efficace du dispositif est requis. Cependant, il reste des questions ouvertes à résoudre dans ce type d'applications. En détail, alors que les algorithmes d'optimisation sont bien adaptés pour trouver un point de fonctionnement optimal, ils ne sont généralement pas bien adaptés pour une mise en œuvre à long terme (par exemple, la stabilisation du système). Les systèmes présentent souvent une détérioration des performances s'ils sont exploités pendant une période prolongée en raison, par exemple, de la dérive causée par des changements dans l'environnement tels que la température ou les vibrations. Dans de tels cas, les chercheurs ont exploré l'utilisation d'algorithmes de récupération en combinaison avec des algorithmes de recherche [31] ou des stratégies d'apprentissage par renforcement qui permettent un fonctionnement à long terme en surveillant constamment l'état de sortie [102]. Ainsi, des approches hybrides combinant des algorithmes de recherche et de stabilisation pourraient être essentielles à un déploiement réussi.

De même, il a été démontré qu'une fibre standard disponible dans le commerce, similaire à celle utilisée pour l'échantillonnage optique au chapitre 2, est capable d'effectuer des tâches d'apprentissage automatique au chapitre 3. L'exploitation de la riche dynamique des composants existants a déjà été exploitée ces dernières années pour la mise en œuvre d'approches optiques d'apprentissage automatique dans le contexte du calcul de réservoir citepBrunner2013. Par conséquent, l'utilisation de ces composants prêts à l'emploi permet une mise en œuvre simple des concepts d'apprentissage automatique optique par rapport aux circuits photoniques hautement complexes [63, 293].
Le système mis en œuvre présente également plusieurs possibilités d'améliorations futures, en plus de l'approche de formation en ligne discutée précédemment: Premièrement, l'utilisation de fibres multicœurs peut permettre un haut degré de parallélisme où l'information peut être traitée dans différents

cœurs sans augmenter l'empreinte du système [390]. De telles fibres pourraient permettre d'atteindre des débits de données TBit comparables à ceux des récentes plateformes intégrées [62, 63]. De plus, les fibres multi-cœurs peuvent apporter une complexité supplémentaire aux systèmes. En particulier, les récents procédés de fibres multicœurs couplées pourraient permettre d'ajouter une dimension supplémentaire à une fraction seulement de la puissance par rapport aux approches multimodes [391–394]. Il est intéressant de noter que les fibres de matériaux hybrides, telles que les fibres à cœur liquide, pourraient également constituer des options appropriées pour de tels schémas de traitement non linéaire, étant donné leurs propriétés uniques telles que la dispersion facilement accordable grâce à l'utilisation de différents liquides, le contrôle adaptatif de la dynamique grâce à des gradients de température, ou le fonctionnement à faible puissance [203, 204, 326, 395, 396]. Notamment, le concept mis en œuvre peut également être combiné avec des approches multimodes existantes [56, 310, 397] car la combinaison d'un tel système avec une dynamique de soliton multimode [398] pourrait offrir une dynamique non linéaire encore plus riche par rapport au régime monomode. Cependant, une telle augmentation de la dynamique non linéaire pourrait se faire au détriment de l'efficacité du dispositif.

Deuxièmement, la mise en œuvre du système en vue d'un fonctionnement en une seule fois peut permettre un calcul très efficace, où la consommation d'énergie pour le traitement des données n'est déterminée que par l'énergie d'impulsion utilisée [7] (environ fJ-pJ s'il est intégré [399–401]). Cette opération à un seul coup pourrait être employée en utilisant des techniques d'étirement du temps en combinaison avec des modulateurs de phase électro-optiques [402] ou l'imagerie dispersive [403, 404]. La combinaison de ces méthodes avec des mesures spectrales en temps réel telles que les transformées de Fourier dispersives [177, 178, 404, 405] peut permettre un fonctionnement en une seule prise. Étant donné les vitesses élevées de ces systèmes, la formation nécessitera probablement des réseaux de portes programmables sur le terrain au lieu des microcontrôleurs/ordinateurs standard [58, 91].

Enfin, le transfert de l'ensemble du système sur une seule puce permettra une meilleure stabilité environnementale, une production de masse potentielle, et offrira une plus grande flexibilité pour la conception des guides d'ondes (notamment en termes de profils de dispersion). Chacun des éléments utilisés peut par exemple être réalisé dans une plateforme en silicium et en nitrure de silicium, depuis les sources femtosecondes efficaces [209, 407], en passant par les schémas de codage et d'imagerie [150, 408], à l'amplification [409], aux schémas de mesure spectrale [410–412] et à la pondération optique [356] pour la couche de lecture [150, 413, 414]. Mais l'intégration hybride de différentes plateformes matérielles peut également constituer une excellente stratégie pour tirer parti des avantages spécifiques de certains matériaux [415]. Enfin, ce travail démontre l'étendue de l'applicabilité de l'optique non linéaire à la résolution de tâches clés dans la mise en œuvre de dispositifs photoniques intelligents.

Perspective - Quelle est la prochaine étape?

L'utilisation d'approches d'apprentissage automatique et d'optimisation guidée, associée aux effets optiques non linéaires, fournit une base appropriée pour mettre en œuvre diverses applications photoniques *smart* basées à la fois sur l'infrastructure photonique existante et sur des circuits intégrés hautement spécifiques. Les progrès récents dans le domaine de l'apprentissage automatique photonique portent également de plus en plus sur l'utilisation de différents effets non linéaires pour un calcul optique hautement efficace. Il s'agit par exemple d'effets de type $\chi^{(2)}$ comme la génération de second harmonique [308, 416, 417] ou d'effets de type $\chi^{(3)}$ comme le mélange à quatre ondes [418].

[7] dans le système démontré, l'analyseur de spectre fait la moyenne de milliers d'impulsions

De plus, les approches décrites dans les chapitres 2 et 3 pourraient être utilisées de manière complémentaire avec de nouvelles méthodes de fabrication telles que l'impression 3D photonique. Les dispositifs qui en résultent ont suscité un grand intérêt ces dernières années pour la mise en œuvre de guides d'ondes, de composants sur puce, d'éléments diffractifs ou de structures de couplage pour les applications d'apprentissage automatique [52, 205, 296, 419]. À ce titre, l'impression 3D pourrait constituer un outil essentiel, en combinaison avec les infrastructures et les composants disponibles sur étagère, pour permettre un prototypage rapide et un déploiement accéléré de la technologie, et donc des méthodologies complémentaires pour des circuits complexes sur puce en temps voulu [64].
La combinaison de ces approches pour tirer parti des différents aspects de chacune d'entre elles peut sembler difficile au premier abord, mais n'est pourtant pas hors de portée. Néanmoins, certaines questions demeurent:

- Quel type d'algorithme ou de fonction de perte est le mieux adapté à une tâche donnée, comme indiqué au chapitre 2. Un exemple récent est la génération de lumière à large bande, qui a été démontrée en utilisant des stratégies d'apprentissage évolutionnaire, ainsi que des stratégies d'apprentissage profond [42, 43]. Pourtant, une comparaison i) des performances et ii) de l'efficacité (par exemple, en termes de vitesse de convergence ou de ressources informatiques utilisées) fait défaut. Le choix d'une approche uniquement basée sur la littérature existante (par exemple, "*cela a toujours été fait comme ça, c'est pourquoi nous l'utilisons*") pourrait ne pas être bénéfique s'il n'existe aucune comparaison entre les différentes mises en œuvre.
- La mise en œuvre physique offre des plates-formes uniques qui permettent de rompre avec les conventions traditionnelles de l'informatique, permettant la réalisation d'approches novatrices et transformatrices. Mais quel type de complexité est requis pour les systèmes neuromorphiques photoniques tels que décrits dans le chapitre 3? Comment mettre en œuvre un système qui peut être très linéaire ou très non linéaire en fonction des différentes tâches à résoudre ? Comment former un système aussi complexe?

La résolution de ces questions constituera sans aucun doute une étape cruciale pour la communauté de la photonique intelligente et suscitera certainement une pléthore de nouvelles approches et mises en œuvre.