

COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author's moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author's reputation

For further information contact the University's Director of Copyright Services

sydney.edu.au/copyright

NOVEL LINEAR AND NONLINEAR OPTICAL SIGNAL PROCESSING FOR ULTRA-HIGH BANDWIDTH COMMUNICATIONS

By

Yvan Paquot

A THESIS SUBMITTED TO THE UNIVERSITY OF SYDNEY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Centre for Ultrahigh bandwidth Devices for Optical Systems School of Physics October 2014



Typeset in $\mathbb{E}_{E} X 2_{\varepsilon}$.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Yvan Paquot

Acknowledgements

To begin, I would like to acknowledge the whole Australian continent and, in particular, Sydney for being such an amazing place. A combination of a tech-savvy society, world-class scientific community, down-to-earth people and a vibrant ecosystem make it doubtlessly the best location on the globe for accomplishing a PhD in (light)wave physics.

As part of this scientific community, Ben Eggleton is the person who gave me the opportunity to join the CUDOS family. Living 100 milliseconds of fibre network latency away from home – or 25 hours by plane – is an everyday challenge that can only be taken up with the support of an adoptive family which I partly found in the friendly team at the School of Physics. Among them, Ben Eggleton and Jochen Schröder, my advisors, are the ones who guided me at light speed through this scientific journey. Their attitude, sometimes strict, sometimes open, pushed me to do my best on my current projects while keeping a place for imagination and new ideas.

The CUDOS group values and encourages the creation of links between individuals and teams by numerous social events. These keep the cohesion of the group as a whole, mixing talents, knowledge and human contacts. I particularly remember my excellent connections with the quantum, microwave photonics and mid-infrared teams of Chunle, David, Alex, Blair, Thomas, Tom, Matt and Shayan, who inspired me and were always ready to lend critical laboratory equipment. It is also an environment where students easily meet the leaders. Ben, Simon, Martijn, Ross, Chris, Boris, Emily, Shelley, Vera, Eric and Enbang are people who I greatly appreciated to talk to, be it in a professional or casual context.

Finisar Australia and its main contact Michael Roelens, who is also one of my rare Belgian fellows in Sydney, can be counted as a valuable part of the support related to the use of Fourier domain programmable optical processors. On several occasions, his quick repairs and lending of materials were the key to the success of my experiments.

Trung Vo has been a model throughout my PhD, setting the bar very high

regarding what a thesis should look like. He was one of my first contacts and friends in CUDOS. In fact, my very first connection in the faculty, even before Prof. Eggleton, was "Mi Amigo", also known as Alvaro Casas-Bedoya. Alvaro has been a wonderful friend all the way from my first informal e-mail inquiry about the Centre until the submission of my thesis.

Dr. Evarist Palushani, from the Technical University of Denmark, deserves special thanks for sharing the tropical terabit lab while carrying high precision experiments without air-conditioning during the heatwave of summer 2012. Prof. Leif Oxenløwe made possible my collaboration with Evarist and, himself, helped with troubleshooting our terabit transmission experiment. Among the other people who have contributed to the realization of the terabit test bed, Simon Lefrancois was remarkable for being the only French speaking person in Sydney to agree to talk exclusively in English. He is also the one who significantly improved the performance of the OTDM transmitter and receiver.

I am indebted to Jochen, Ben, Chad, Simon, Mark and Alex for supporting me in launching my project proposal to develop optical neural networks. Diving into a novel research topic is at the same time scary and twice as motivating. They played a major role in the enjoyment I experienced from doing research. I cannot forget Marc Haelterman and Serge Massar, who initiated me to the fascinating world of neural networks.

I also thank warmly my colleague and friend Thomas Büttner and the dolphins of Bondi for sharing unique moments of morning surfs that were doubtlessly the best way to start a day of intense research. Many more physics staff and students played a great role during my PhD, directly or indirectly. To avoid naming the whole School of Physics, I will only mention my close office mates Björn, Felix, Young, Parry, Iman, Steve, Jiakun, Neetesh, Fernando, Stephanie, Alessandro, Scott, Caitlin, Michael and Sahand, as well as Eve, Alexis, Chindy and Shauna from the Student Service Office.

All of this has been enabled by the Australian Research Council (ARC) government grant that funds CUDOS and, in particular, the biggest part of my research. Experimental research often relies on expensive equipment, which can quickly become the major limiting factor of a project. The ARC funding was managed wisely and carefully by the head of the Centre and has never been a worry at any point in my PhD.

Besides academic life, the other part of my adoptive family played an enormous role in these three years. Antoine had no clue of the trick that destiny was playing when he decided, on that windy day of 2010, to cycle 10 km carrying his gear to go kitesurfing at Ramsgate. A future book will have to narrate the adventures that followed. Quickly completed with Nicolas, Chris, Louisa, Lise and Daniel, the core crew of the Crazy Kitesurfers family was founded. It is no wonder if the group was later renamed Sydney Uni Crazy Kitesurfers, destined for taking the place of largest club in the state. The heart of my new family was later extended by Marine, Eveline, Mikko, Hongi and Kaley, as well as Elg, Ed, the Toms, Greta, Dirk, Marvin and so many others who made my time in Sydney so amazing.

Finally, my family and friends in Belgium, although distant, were in my heart every day. The feeling of knowing that my relatives were ready to welcome me back every time I was returning home was reassuring and supporting. My parents and close family were a constant source of encouragements and my regular contact with the Old Continent. All my friends, from my swimming clubs, the Universit Libre de Bruxelles, Jeunes et Nature and other social circles were also irreplaceable. I thank particularly my parents, Vincent, Richard, Ivan, Sophie, Fabrizzio, Jessy and Tuan for visiting me in Sydney during these three years. Welcoming a fellow Belgian felt like bringing a bit of my country and origins Down Under.

After so many names and so many thanks, I am afraid of having forgotten more unforgettable colleagues or friends. Surely I have. To avoid this awkward situation, I invite the reader to wait no longer, and discover the topic that occupied me during my PhD.

> Yvan Paquot University of Sydney December 2013

List of PhD Publications

Journal papers

- Y. Paquot, J. Schröder, J. Van Erps, T. D. Vo, M. D. Pelusi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Single parameter optimization for simultaneous automatic compensation of multiple orders of dispersion for a 1.28 Tbaud signal", **Optics Express**, Vol. 19, Issue 25, pp. 25512-25520 (2011).
- J. Schröder, T. D. Vo, Y. Paquot, and B. J. Eggleton, "Breaking the Tbit/s Barrier: Higher Bandwidth Optical Processing", **Optics and Photonics News**, Vol. 23, Issue 3, pp. 32-37 (2012).
- Y. Paquot, J. Schröder, E. Palushani, R. Neo, L. K. Oxenløwe, S. Madden, D.-Y. Choi, B. Luther-Davies, M. D. Pelusi, and B. J. Eggleton, "Automatic DGD and GVD compensation at 640 Gb/s based on scalar radio-frequency spectrum measurement", Applied Optics, Vol. 52, Issue 9, pp. 1919-1927 (2013).
- R. Neo, J. Schröder, <u>Y. Paquot</u>, D.-Y. Choi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Phase-sensitive amplification of light in a χ⁽³⁾ photonic chip using a dispersion engineered chalcogenide ridge waveguide", **Optics Express**, Vol. 21, Issue 7, pp. 7926-7933 (2013).
- Y. Paquot, J. Schröder, M. D. Pelusi, and B. J. Eggleton, "All-optical hash code generation and verification for low latency communications", **Optics Express**, Vol. 21, Issue 20, pp. 23873-23884 (2013).
- Y. Paquot, J. Schröder, and B. J. Eggleton, "Reconfigurable linear combination of phase-and-amplitude coded optical signals", **Optics Express**, Vol. 22, Issue 3, pp. 2609-2619 (2014).

Post-deadline conference papers

- Y. Paquot, J. Schröder, J. Van Erps, T. D. Vo, M. D. Pelusi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Automatic higher-order dispersion measurement and compensation of a 1.28 Tbaud signal", Australian Conference on Optical Fibre Technology (ACOFT) PD, Melbourne, Australia, 2010.
- Y. Paquot, J. Schröder, M. D. Pelusi, and B. J. Eggleton, "On-chip all optical error detection for ultra-low latency communications", **Frontiers in Optics** PD, Orlando, USA, 2013.

Contributed conference papers

- Y. Paquot, J. Schröder, J. Van Erps, T. D. Vo, M. D. Pelusi, S. Madden, D.-Y. Choi, D. A. Bulla, B. Luther-Davies, and B. J. Eggleton, "Automatic higher-order dispersion measurement and compensation of a 1.28 Tbaud signal", **Conference on Lasers and Electro-Optics** (CLEO), Baltimore, USA, 2011. [UPGRADED TO INVITED PRESENTATION]
- Y. Paquot, J. Schröder, J. Van Erps, T. D. Vo, M. D. Pelusi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Multi-order, automatic dispersion compensation for 1.28 Terabaud signals", SPIE Photonics Europe, 84340I-8, Brussels, Belgium, 2012.
 [BEST STUDENT PAPER PRESENTATION AWARD]
- J. Schröder, R. Neo, <u>Y. Paquot</u>, D.-Y. Choi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Phase-sensitive amplification in a χ⁽³⁾ photonic chip", Conference on Lasers and Electro-Optics (CLEO) Europe, Munich, Germany, 2013.
- Y. Paquot, J. Schröder, and B. J. Eggleton, "Linear Crosstalk Compensation by All-Optical Dot-Product Operation", **Frontiers in Optics**, Orlando, USA, 2013. [OUTSTANDING STUDENT PAPER COMPETITION FINALIST]
- Y. Paquot, J. Schröder, M. D. Pelusi, and B. J. Eggleton, "Format Transparent All-Optical Hash Code Generator And Comparator", Australia and New Zealand Conference on Optics and Photonics (ANZ-COP), Perth, Australia, 2013.

Provisional patents

- Y. Paquot, J. Schröder, M. D. Pelusi, and B. J. Eggleton, "A DEVICE AND A METHOD FOR GENERATING AN OPTICAL HASH SIGNAL FOR AN OPTICAL DATA PACKAGE", Provisional patent, 2013.
- Y. Paquot, J. Schröder, M. D. Pelusi, and B. J. Eggleton, "A METHOD AND A SYSTEM FOR COMPARING A FIRST OPTICAL SIGNAL AND A SECOND OPTICAL SIGNAL", Provisional patent, 2013.

Abstract

The thesis is articulated around the theme of ultra-wide bandwidth single channel signals. It focuses on the two main topics of transmission and processing of information by techniques compatible with high baudrates. The processing schemes introduced combine new linear and nonlinear optical platforms such as Fourier-domain programmable optical processors and chalcogenide chip waveguides, as well as the concept of neural network.

Transmission of data is considered in the context of medium distance links of Optical Time Division Multiplexed (OTDM) data subject to environmental fluctuations. Solutions are developed towards real time adaptation of parameters to optimize the signal quality. Temperature drifts, changes in strain and vibrations can cause the properties of optical fibres to randomly oscillate. Based on a single quality indicator of the signal, we show that a multivariate optimization algorithm can concurrently control multiple sources of impairments. We experimentally demonstrate simultaneous compensation of differential group delay and multiple orders of dispersion at symbol rates of 640 Gbaud and 1.28 Tbaud. The high speed monitoring of the signals is performed via an integrated all-optical RF-spectrum analyzer implemented on a dispersion-engineered chalcogenide ridge waveguide.

Signal processing at high bandwidth is envisaged both in the case of elementary post-transmission analog error mitigation and in the broader field of optical computing for high level operations (optical processor). A key innovation is the introduction of a novel four-wave mixing scheme implementing a dot-product operation between wavelength multiplexed channels. The scheme is first used as a linear distortion compensator to demonstrate error-free operation. It is then proved as a format transparent hash-key comparator for all-optical error detection in links encoded with advanced modulation formats. The complete experimental setup also includes a novel reconfigurable optical hash-key generator based on Fourier domain processing.

Finally, the work presents groundbreaking concepts for compact implementation of an optical neural network as a programmable multi-purpose processor. The proposed architecture leverages the parallelism of frequency division multiplexing and the reconfigurability of Fourier domain optical processors to implement neural networks with several nodes on a single optical nonlinear transfer function. Reconfigurable AND and XOR gate functions are demonstrated with a simplified version of the system. We also show experimentally analog-to-digital conversion with a full-featured all-optical neural network. We then analyze applications that could be supported by a more advanced although feasible implementation of the concept, such as analogue-to-digital conversion with higher resolution, nonlinear distorsion compensation and multilevel logic operations (MIN,MAX).

All nonlinear processes involved in this work employ third-order nonlinear effects in Kerr media. Four-wave mixing is the base of all switching and dotproduct operations. Cross-phase modulation is used for RF-spectrum monitoring and self-phase modulation enables pulse compression for generation of ultra dense OTDM channels. The applications requiring extremely wide bandwidth or low latency use dispersion engineered chalcogenide waveguides, while demonstrations of processing schemes at lower baudrate are made on highly nonlinear fibre platform. However, all the principles can be scaled to higher baudrates by mean of exploiting chip based nonlinearities.

The particularity of the thesis is the new approaches to optical signal processing that potentially enable high level operations using simple optical hardware and limited cascading of components. Ultimately, the thesis concludes with a vision on the role that all-optical signal processing could take in-between passive data transmission and large scale electronic computing in the future.

Contents

A	Acknowledgements v				
List of PhD Publications ix				ix	
\mathbf{A}	Abstract xiii				iii
Li	st of	Figure	es	Х	ix
1	Lig 1.1	h t as v Summ	rector of information		$ 1 \\ 2 $
2	Hist	tory ai	nd background		5
	2.1	Histor	v of telecommunication		5
		2.1.1	The roots of communications		5
		2.1.2	"Modern" communication systems		6
		2.1.3	From analog to digital		8
		2.1.4	Optical fibre communications		9
	2.2	Histor	y of information processing		12
		2.2.1	Development of contemporary computing methods		12
		2.2.2	The success of digital electronics		13
		2.2.3	Optical computing technology	•	13
		2.2.4	Which place for optics in processing?		14
	2.3	Eleme	ents of information theory		16
		2.3.1	Entropy		16
		2.3.2	Channel capacity		17
		2.3.3	Encoding formats		19
		2.3.4	Nonlinear Shannon limit		20
	2.4	Lateno	cy		21
3	Nor	linear	ity in optics	(23
	3.1	Linear	coptics		23
		3.1.1	Equations of propagation	•	24
	3.2	Nonlin	near optics		25
		3.2.1	Intuitive model of nonlinearity		25

		3.2.2	Mathematical description of a nonlinear system	26
		3.2.3	Simplifications	28
	3.3	Wave	equation	30
		3.3.1	Linear wave equation	30
		3.3.2	Modes of propagation	31
		3.3.3	Nonlinear polarization	31
		3.3.4	Second order term	32
		3.3.5	Third order term	32
	3.4	Propa	gation in waveguides	34
		3.4.1	Linear loss	34
		3.4.2	Dispersion	34
		3.4.3	Polarization mode dispersion (PMD)	36
		3.4.4	Self phase modulation	37
		3.4.5	Cross phase modulation	38
		3.4.6	Four-wave mixing	39
	3.5	Nonlir	near optical processing platforms	40
		3.5.1	Highly nonlinear fibre	41
		3.5.2	Integrated chalcogenide platforms	42
		3.5.3	Other common nonlinear platforms	44
			_	4 5
	3.6	Nonlir	near optics at work	45
	3.6	Nonlir	near optics at work	45
4	3.6 Ulti	Nonlir ra high	baudrate signals and automatic compensation meth-	45
4	3.6 Ultrods	Nonlir ra high	baudrate signals and automatic compensation meth-	45 47
4	3.6 Ultr ods 4.1	Nonlir ra high Gener	a baudrate signals and automatic compensation meth-	 45 47 48 40
4	3.6 Ultr ods 4.1	Nonlir ra high Gener 4.1.1	a baudrate signals and automatic compensation meth- ation and detection of OTDM data	45 47 48 49
4	3.6 Ultr ods 4.1	Nonlir ra high Gener 4.1.1 4.1.2	hear optics at work a baudrate signals and automatic compensation meth- ation and detection of OTDM data Interleaving stages Bandwidth requirements OTDM demultipleving	 45 47 48 49 50 51
4	3.6 Ultr ods 4.1	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN	hear optics at work	 45 47 48 49 50 51 54
4	 3.6 Ultrained 4.1 4.2 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1	hear optics at work	 45 47 48 49 50 51 54 55
4	 3.6 Ultradia 4.1 4.2 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1	hear optics at work	45 47 48 49 50 51 54 55 57
4	 3.6 Ultriods 4.1 4.2 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comm	hear optics at work	45 47 48 49 50 51 54 55 57 57
4	 3.6 Ultrods 4.1 4.2 4.3 4.4 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp	hear optics at work	45 47 48 49 50 51 54 55 57 58 60
4	 3.6 Ultransformed and the second second	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDM 4.2.1 4.2.2 Comp Comp	hear optics at work	45 47 48 49 50 51 54 55 57 58 60 62
4	 3.6 Ultrods 4.1 4.2 4.3 4.4 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1	hear optics at work	45 47 48 49 50 51 54 55 57 58 60 62 62
4	 3.6 Ultransformed and the second second	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDM 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62
4	 3.6 Ultrods 4.1 4.2 4.3 4.4 4.5 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2 Auton 4.5 1	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 64 64
4	 3.6 Ultion 0ds 4.1 4.2 4.3 4.4 4.5 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp 4.4.1 4.4.2 Auton 4.5.1 4.5.2	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62 64 64
4	 3.6 Ultrodds 4.1 4.2 4.3 4.4 4.5 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2 Auton 4.5.1 4.5.2	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62 64 64 64
4	 3.6 Ultrodds 4.1 4.2 4.3 4.4 4.5 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2 Auton 4.5.1 4.5.2	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62 64 64 67 70
4	 3.6 Ultrodds 4.1 4.2 4.3 4.4 4.5 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2 Auton 4.5.1 4.5.2 4.5.3 Discut	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62 64 64 67 70 75
4	 3.6 Ultrodds 4.1 4.2 4.3 4.4 4.5 4.6 4.7 	Nonlir ra high Gener 4.1.1 4.1.2 4.1.3 OTDN 4.2.1 4.2.2 Comp Comp 4.4.1 4.4.2 Auton 4.5.1 4.5.2 4.5.3 Discus Conch	hear optics at work	43 47 48 49 50 51 54 55 57 58 60 62 62 62 64 64 64 70 75 75 76

5	Nor	nlinear	optical processing in communication	79
	5.1	You c	an buy bandwidth, not latency	80
	5.2	Energ	y concerns	80
	5.3	Existi	ng all-optical operations	81
		5.3.1	Regeneration	81
		5.3.2	Nonlinearity compensation	82
		5.3.3	Routing	83
		5.3.4	Logic operations	83
	5.4	Multi	path interferometric circuit	84
	5.5	Optica	al dot-product	85
		5.5.1	Principle	86
		5.5.2	Experimental setup with 3 signals	89
		5.5.3	Qualitative results of dot-product between 3 signals	91
	5.6	Linear	distortion compensation	91
		5.6.1	Principle	93
		5.6.2	Results	94
		5.6.3	Discussion	95
	5.7	All-op	tical error detection by hash code comparison \ldots \ldots \ldots	97
		5.7.1	Principle	98
		5.7.2	Implementation of the hash generator	100
		5.7.3	Implementation of the hash comparator	101
		5.7.4	Experiment and results	102
		5.7.5	Integrated hash comparator with ultra low latency	107
		5.7.6	Discussion	109
	5.8	Concl	$usion \ldots \ldots$	112
6	Opt	cical ne	eural networks	115
	6.1	Neura	l network processors	115
		6.1.1	History and intuitive vision of neural networks	115
		6.1.2	Terminology and classes of neural networks	118
		6.1.3	Mathematical model of an artificial neural network	119
		6.1.4	Optimization of neural networks	120
		6.1.5	A simple case: a 2-bit ADC	122
		6.1.6	Applications and potential of neural networks	125
		6.1.7	Hardware neural networks	125
		6.1.8	Existing optical neural network schemes	126
	6.2	Frequ	ency multiplexed neural network	127
		6.2.1	General principle	128
		6.2.2	Advantages of frequency multiplexing	128
	6.3	Imple	mentation	129
		6.3.1	Signal generation	129
		6.3.2	Nonlinear transfer function	130
		6.3.3	Bias summation	131

		6.3.4 Dot-product	. 132
	6.4	Optoelectronic reconfigurable logic gates	. 132
		6.4.1 The easy way: power adder	. 132
		6.4.2 Logic gates	. 134
	6.5	All-optical frequency parallel neural network	. 138
		6.5.1 Combining nonlinearity and dot-product	. 138
		6.5.2 Experimental demonstration of the ONN	. 142
		6.5.3 Results for a 2-bit analog-to-digital converter	. 143
	6.6	Practical functions	. 148
		6.6.1 Complexity limits	. 148
		6.6.2 Training	. 149
		6.6.3 Possible applications of a simple ONN	. 150
	6.7	Future of neural network in optics	. 150
7	Con	clusion and outlook	153
	7.1	Conclusion on high baudrate telecommunications	. 154
	7.2	Conclusion on analog optical computing	. 155
	7.3	General outlook on optics in information science	. 157
A	Fou	rier domain programmable optical processors	159
в	ΟΤΙ	OM demultiplexing troubleshooting	161
	B.1	Timing drift	. 162
\mathbf{C}	OTI	DM setup evolution	165
	C.1	40 GHz demultiplexing pump	. 165
	C.2	10 GHz demultiplexing pump from semiconductor laser source	. 165
	C.3	Chip sample changes and damages in chalcogenide waveguides	. 166
D	Glit	ches	169
\mathbf{E}	Trai	nsfer function of the difference of two neurons	171
Ac	crony	vms	173
R/	foro	1065	175
TUC			тю

List of Figures

1.1	Tree view of some disciplines related to photonics	3
2.1	Information technology throughout the ages	6
2.2	Replacement of copper wires by optical fibres	$\overline{7}$
2.3	Evolution of the Bandwidth-Distance product.	9
2.4	Pushing the limits of baudrate.	11
2.5	Evolution of electronic and photonic technologies.	15
2.6	Evolution of the entropy of one binary symbol	17
2.7	Shannon channel with additive noise and mutual information.	18
2.8	Maximum spectral efficiency defined by the Shannon limit	19
2.9	Advanced modulation format constellations	20
2.10	Nonlinear Shannon limit	21
3.1	Intuitive representation of optical nonlinearity	26
3.2	General nonlinear system.	27
3.3	Effect of anomalous chromatic dispersion	35
3.4	Effect of DGD on an optical pulse.	37
3.5	Chalcogenide chip.	43
3.6	Dispersion profile of a chalcogenide chip waveguide	43
4.1	Principle of OTDM time-interleaving	49
4.2	Generation of 640 Gb/s and 1.28 Tb/s PRBS signals	50
4.3	OTDM signal demultiplexing by nonlinear sampling	51
4.4	Experimental setup for terabaud demultiplexing	52
4.5	Spectral reshaping of the OTDM signal	53
4.6	Error-free demultiplexing of a 640 Gbaud signal	53
4.7	Principle of the optical RF-spectrum analyzer	55
4.8	Implementation of optical RF-spectrum analyzer	57
4.9	Evolution of the RF spectrum vs. dispersion	58
4.10	Signal quality vs. dispersion	59
4.11	Dispersion emulation with a FD-POP.	59
4.12	DGD compensator	61
4.13	Spectral interference patterns	62
4.14	Resolution limit of the programmable delay line	63

4.15	Flow diagram of the optimization algorithm
4.16	Tone power monitoring of sharp DGD changes
4.17	Automatic dispersion compensation scheme
4.18	Automatic dispersion compensation test-bed
4.19	Tone power vs. dispersion
4.20	Automatic dispersion compensation time-line
4.21	Principle of simultaneous compensation of GVD and DGD 71
4.22	Setup for simultaneous compensation of GVD and DGD 72
4.23	Tone power as a function of DGD
4.24	Results of automatic DGD compensation
4.25	Results of automatic GVD and DGD compensation
5.1	Multipath interferometric circuit
5.2	Principle of the all-optical dot-product
5.3	Experimental diagram of an optical dot-product 90
5.4	Optical spectrum after the HNLF
5.5	Sampling scope traces of dot-product output
5.6	Dot-product to revert linear distortions
5.7	FD-POP to apply channel distortion
5.8	Optical spectrum after the HNLF
5.9	Bit-error rate testing of the compensator
5.10	Principle of the all-optical hash code verified link
5.11	Functional diagram of hash code system
5.12	Hash code generation by Fourier domain optical processing 100
5.13	Experimental intensity plots of the hash codes
5.14	Principle of the all optical coherent signal comparator 102
5.15	Experimental setup of optical hash code verification 103
5.16	Experimental optical spectrum at the HNLF output 103
5.17	Time traces of the error signal
5.18	Signal generator to create localized errors
5.19	Time traces of the error signal with a single error
5.20	Hash code verified link of 20.72 km
5.21	On-chip hash comparison system
5.22	Output spectrum and extracted error signal
5.23	Inaccuracy in the hash comparison
61	Natural and artificial neural networks
6.2	Biological and artificial neurons
6.2	Analogy with a notwork of massos
0.0 6 4	Noural notwork architecture
0.4 6 5	Iterative optimization of hardware neural networks
0.0 6.6	Construction of binary code and Cray code
0.0	Evolution of temp transfer for stice
0.7	Evolution of <i>tann</i> transfer function

6.8	Spatially multiplexed implementation of an ANN
6.9	Building blocks of a frequency multiplexed ONN
6.10	Intensity detector as channel adder
6.11	Balanced photodetector as adder and subtracter
6.12	Diagram of an optoelectronic WDM neural network
6.13	Flow chart of an XOR gate
6.14	Measurements of XOR operation
6.15	Flow chart of an AND gate
6.16	Measurements of AND operation
6.17	Comparison of experimental transfer function with $tanh(x)$ 139
6.18	Experimental setup for nonlinear transfer measurement 140
6.19	Transfer function with two neurons
6.20	Experimental diagram for transfer characterization
6.21	Neural network input signal
6.22	Experimental neural network output
6.23	Neural network response vs. input gain
A.1	Fourier domain programmable optical processor
B.1	Timing jitter measurement
B.2	Demultiplexing of terabaud signals
C.1 C.2	RMS jitter of semiconductor laser
D.1	Apparition of glitches in an XOR gate
E.1	Transfer function with two neurons for subtraction

]

Light as vector of information

The use of light for signalling purposes dates back to antiquity [1, 2]. In the three millennia after the legendary king Agamemnon informed his homeland of the capture of Troy with beacons in Ancient Greece [3], naval navigation and armies developed lighthouses, semaphores and light codes as strategic ways to communicate [4].

With the discovery of methods to control electricity in the early 19th century and, later, the discovery of electromagnetic waves, wired and radio telegraphs revolutionized the world of communication [5, 6]. Visual telegraphy systems were quickly outperformed and light was no longer seen as an efficient method for information. Within a century, the growth of the demand for voice and data transmission pushed technology to hit limits in terms of speed, bulk and amount of data carried [7]. The development of low loss guided optics in the seventies opened up a field of progress to keep up with the demand, returning to the use of light as vector of information [8, 9]. Fibre optic cables gave a new breath to telecommunication systems by offering long haul propagation, high bandwidth, interference free and highly multiplexable data channels. Although fundamental bounds exist to the maximum information carried by an optical fibre, these were far out-of-sight and did not appear for decades. This illusory limitless information capacity did not last, as technological progresses in dense optical signals got closer to the theoretical boundary of the nonlinear Shannon limit which states that not only noise, but also the nonlinearity of optical fibres limits the amount of data that can be carried [10, 11].

In parallel with data transmission, the domain of information processing has evolved along a similar pattern. The human brain, eventually helped by rudimentary mechanical tools, was for long the only available way to compute [12, 13]. The advent of automated calculus based on electronics revolutionized the field [14]. Again, recent needs for more abstract operations and human-like machine behaviours led scientists to go back to the basics and reverse-engineer the principles acting in the brain itself.

This process triggered the rise of neural networks as a new paradigm for information processing. The race for clock rates in computation also initiated numerous projects to replace electronic building blocks by optical equivalents. For decades, optical signal processing has nurtured big hopes as a faster and more power efficient platform to carry operations than electronics. However, in the last ten years this proposition has had more and more sceptics inside the optical community itself. Based on a systematic comparison of existing digital and optical technologies, it has even been found to be less powerful and space efficient than digital signal processing [15, 16] for complex functions requiring cascading of multiple logic gates. However, new materials and techniques are now making photonics a competitive solution for more and more applications. Another central argument in favour of all-optical methods is latency. Optics provides an alternative to optoelectronic conversion and electronic processing at low clock-rate.

This thesis is centred on the theme of ultra-high bandwidth optical signal processing, to push both communication and computation beyond their present limits. In this context, optical signal transmission consists of transporting information encoded on optical signals from one point in space to another with minimum distortion to the data. Transmission distances can range from thousands of kilometres for intercontinental links or even satellite communications to millimetres for on-chip photonic busses. This area benefits from optical signal processing all the way from improving the transmitters and receivers to monitoring and reducing distortions caused by the link. On the other hand, all-optical computation focuses on transforming optical data to perform logic operations and calculus. More particularly, we focus on the techniques that involve converting, filtering and carrying operations with optical signals, relating to both the aspects of communication and computation.

1.1 Summary and outline

The reader may have noticed that, in this introduction, we have presented two main aspects of optical information science: communication and computing. Although these two topics might seem qualitatively different, they are complementary and intimately connected in applications (transmission requires processing to condition the information and computing often involves transport of data). In the particular topic of the thesis, it is natural to approach both aspects together given that similar techniques and devices are employed



in the respective experiments achieved.

FIGURE 1.1: Tree view of some disciplines related to photonics. Topics in solid black are part on this work.

The next chapter will expand on the history of telecommunications and information processing before discussing the great breakthroughs that photonics promises at many levels. We will then introduce the key concepts essential to put our work into a modern perspective, such as information theory, the nonlinear Shannon limit, transmission formats and linear and nonlinear optical signal processing.

The third chapter will lay the theoretical basis required for the understanding of our work. Classical optical field propagation and the fundamentals of nonlinear light-matter interactions will be covered first. The usual nonlinear optical media employed for signal processing will be reviewed to justify the choices in our experimental demonstrations and emphasize the challenges related to their use in communications.

We will begin reporting on our core results in chapter 4 with the transmission of data at ultra high baudrates using optical time division multiplexing (OTDM). Existing and novel methods to generate and detect OTDM signals as well as measure and undo their degradations will be covered. In particular, we will focus on mitigation of differential group delay (DGD) and higher-order dispersion fluctuations. We will introduce a method based on Fourier domain optical processing that allows for reconfigurable control of these parameters.

Regarding optical information processing, the two following chapters will investigate methods to leverage the high speed and multiplexing abilities of optics in order to perform mathematical calculations. The common idea behind our architectures is the encoding of the variables involved in the operations onto wavelength division multiplexed (WDM) channels and their simultaneous processing with a single nonlinear optical medium. Ultimately, we aimed at the vision of an all-optical WDM neural network processor made of three main building blocks: signal generation with bias, nonlinear transfer function and dot-product between multiple WDM channels.

These three modules can also have novel applications on their own, which will be reported in chapter 5. At an early stage, the function of our all-optical dot-product has been used as a format transparent and low-latency linear distortion compensator at 40 Gb/s (see section 5.6). The combination of the dot-product with the interferometric setup for bias addition was also employed to demonstrate comparison of hash-code signals for all-optical error detection in communication networks. The generation of the hash codes was done with a Fourier domain programmable optical processor (FD-POP) pre-processing device.

The last chapter is dedicated to novel methods for all-optical computing. These ground breaking principles relate to the paradigm of neural networks and possible optical implementation. A brief introduction to neural network information processors is given for scientists who are familiar with nonlinear systems such as nonlinear optics. Neural networks can be understood as highly nonlinear systems described by a similar formalism. Experimental demonstrations of such systems might suggest that optical neural networks could be a general, reconfigurable computing platform possibly outperforming state-ofthe-art electronics and opening a new field in optical signal processing [17, 18]. The remarkable advantage of the architectures presented is their extreme simplicity in terms of hardware. The use of only one nonlinear optical medium, combined with WDM data encoding, enables complex computation functions that would require a large number of logical gates in classical binary logic. Our experimental results include demonstrations of a rudimentary neural network realizing a reconfigurable AND and XOR gate at 40 Gb/s and a more complete all-optical scheme operating analog-to-digital conversion (ADC) with 2-bits resolution.

One of the key enabling technologies that made these experiments possible was the use of FD-POP. These devices, explained in appendix A, allow for easy wavelength-selective phase and amplitude control on optical signals, as well as a flexible implementation of interferometric circuits.

Green light to the reader

Through the work achieved during this thesis, we hope to have contributed to the development of the fields of optics and information technology. Our discoveries and improvements of existing methods rely on centuries of fascinating science that is still bursting with new domains to explore.

We hope the reader finds our new approach to optical processing as interesting as we did.

2 History and background

2.1 History of telecommunication

Communication is an essential function in human groups since time immemorial. Starting with the appearence of complex oral language, transmission of information has played a major role in the development of civilization. The decisive commercial, military and social advantages of societies with improved communication have continued the drive towards ever more sophisticated methods.

2.1.1 The roots of communications

Communicating over long distances seems natural during the era of internet and mobile phones. However, people as close as our grandparents can attest of the incredible change that the generalization of telephones, radios and increasingly sophisticated media brought to society within just two generations.

Less than two centuries ago, few tools were available for remote communication. Postriders and line-of-sight visual signals were the few methods available. The first of these allowed for large amounts of data to be sent over long distances, but was slow due to the time needed for the postman to travel. Visual systems, however, allowed for virtually instantaneous transmission, but at a low bandwidth and on spans limited by the performance of the human eye and weather conditions.

Since the origins of time until the 19th century, these communication methods had seen only a very slow evolution. Word of mouth, messengers and traditional stories were the most common ways to connect people in antiquity. The use of written messages was one notable progress that allowed for reliable transmission of more instructions. For particular applications where succinct messages had to be sent quickly, such as military orders or naval warnings, light, smoke and other visual signals were commonly used. Among noteworthy strategies used in history are the already mentioned chain of beacons lit up to announce a victory or an attack, the use of a cloth to modulate the intensity of the smoke in order to send messages faster, carrier pigeons or the combined use of torches with a timing device by antique Roman armies. Lighthouses with various colour codes and flashing patterns to locate coastal positions and warn of a danger are still in use in present naval navigation.

The highest degree of evolution of ancient visual transmissions probably comes to the Chappe telegraph deployed across France and Europe from 1792. Made of a network of semaphore lines, it could transmit messages at a rate up to about two words per minute. Relay towers separated by 30 km repeated up to the destination node of the chain a series of pre-arranged visual signals formed by an articulated assembly of three wooden planks.



FIGURE 2.1: Information technology throughout the ages. Technologies boxed in pink are the object of this work.

2.1.2 "Modern" communication systems

We decide to set the start of the modern era of communication from the use of electricity for transmitting information. This arbitrary choice is justified by the fact that electricity is the enabling technology that triggered the exponential growth of communication between people. It is also the starting point of automated transmission lines.

The discovery of electricity as a vector of information enabled multiple

breakthroughs in the early 19th century. Electrical telegraphy was first experimentally demonstrated in an electrochemical setup where the receiver was observing electrolysing of an acid solution as the sender was closing a circuit [6]. Real world applications started with the realization of the first commercial electromagnetic telegraph capable of receiving telegrams with sub-second modulation, independently patented in 1837 by the duo Cooke and Wheatstone and by Samuel Morse.

The joint success of the telegraph and telephone (patented in 1876) led to the rapid development of an increasingly complex network of copper wire lines in many countries. This analog communication grid was used as a first backbone for modern datacom when automated coded messaging devices were introduced. The Telex (Teletypewriter eXchange, 1926) and Minitel (1978) paved the way for the generalization of the transfer of binary data. Since the 1980's, the rise of modern internet revolutionized again the exchange of information and amplified the boom in the amount of data transmitted.



FIGURE 2.2: Left: management of copper phone lines becomes a mess in large networks (Alyson Shontell, Business Insider Australia). Right: replacement of copper wires by optical fibres (Gary Valentine, IRWA). One optical fibre (bottom) carries the same data as a bundle of copper wires (top).

The public access to these technologies and the ascent of data intensive applications quickly saturated the public switched telephone network (PSTN) originally designed for telephony. Although significant improvements could be made to local connections of the user to a nearby hub (ISDN, DSL, coaxial cable), copper cables are not suited for long haul transmissions at high bandwidths. One of the main limitations is the fact that a copper wire acts like an antenna and the radiated electromagnetic field increases with the frequency carried. At high bit rates, this causes losses and interference between neighbouring channels. The emergence of optical fibre technology gave the internet backbone the potential to support much higher bandwidths.

2.1.3 From analog to digital

As rudimentary as they were, the archaic systems of light signals introduced in section 2.1.1 already featured general characteristics of communication systems and can be used as intuitive examples to illustrate some abstract notions.

Natural language between humans is essentially analog in the sense that there is no clear quantifiable limit between two signs with different meanings. The example of spoken language contains an immense number of codes (words) that can only be distinguished and identified by a trained ear. The physical carrier of the voice – the sound wave – presents continuous variations in air pressure. Transport of such signals over long distances is challenging as they cannot be regenerated to remove noise added during propagation, for example through an original plain old telephone service (POTS) phone line. This is also the case for classical analog radio and TV broadcasting that permanently degrade over distance.

Antique communication systems by visual signals could hardly be used in the analog regime. For example giving various meanings to slightly different degrees of brightness of a fire is likely to be confusing for the observer. Instead, two significantly different states like fire lighted or extinguished are easy to identify from distance and can code simple binary messages such as "danger/no danger", "battle won/not won", "dangerous submerged rocks/safe sea". Encoding of the message onto a finite number of quantized states instead of an analog continuum allows for clear distinction and eventually repeating of the information. This is the basic principle of digital communications.

More complex messages require either repetition of this operation at a fixed rate (called baudrate or symbol rate) or the use of signs with multiple levels (such as various durations of the lighting or positions of the semaphore) or a combination of both. An elementary block of information corresponding to the time window between two clock ticks is called a symbol and the type of coding of the symbol (on-off, multilevel, duration, etc.) is commonly called modulation format, which is similar to the alphabet used in a written message.

In modern language, a multilevel coding of visual signals would be called "advanced modulation format" (AMF), in contrast to a simple on-off method called on-off keying (OOK). As the distance between the sender and recipient increases, reading of the message becomes more difficult due to attenuation by spatial spreading of the light beam, meteorological conditions and confusion with nearby light sources. Ultimately, errors can occur or the communication becomes impossible. The ability to distinguish the signal from the background noise is quantified by the signal-to-noise ratio (SNR), which in the optical domain becomes the optical signal-to-noise ratio (OSNR).

2.1.4 Optical fibre communications

The realization of optical cables with loss lower than 20 dB/km, predicted by Charles K. Kao and demonstrated in 1970 by the team of Robert D. Maurer, opened up their use for communication. Present fibre technology allows for propagation losses as small as 0.2 dB/km. The development of high speed light modulators and detectors allowed the translation of electronic data into optical signals, which could be transmitted through an optical link and then converted back to the electrical domain at the receiver.



FIGURE 2.3: Evolution of the Bandwidth-Distance product offered by optical fibre communications. Diamonds: single wavelength; open diamonds: OTDM; triangles: WDM, filled circles: OFDM; open circles: coherent detection (AMF). Figure adapted from [11].

Increasing the baudrate

The first idea used to transmit more data through a single fibre was increasing the baudrate with faster electronic systems and opto-electronic conversion. This trend had already been initiated in copper wire networks working at increasing frequency. Traditionally, the growth of the symbol-rate has been associated with a dramatic reduction of the cost-per-bit of telecommunications. Electronic time division multiplexing (ETDM) allowed driving of the optoelectronic transducers at very high frequencies. Figure 2.3 plots the bandwidth-distance product achieved experimentally over the years. Filled diamonds correspond to ETDM based systems.

Early optical fibres were the source of numerous types of impairments limiting the baudrate, such as polarization mode dispersion (PMD), that will be explained later in chapter 3. Also, electronic systems generating the signals and optoelectronic repeaters placed regularly on the links to regenerate and amplify them were limited in bandwidth. AMF such as Quadrature Phase Shift Keying (QPSK) or Quadrature and Amplitude Modulation (QAM) increase the amount of information encoded on every symbol by using multilevel amplitude and phase alphabets [19]. However, the signals generated with this method are more sensitive to distortions and require coherent detection at the receiver.

Parallel transmission

Numerous innovations were made to increase the throughput of optical links even further. The invention of Erbium doped fibre amplifiers (EDFA) was a major technological advance that made possible re-amplification of optical signals without the need for optoelectronic repeaters. It enabled the use of wavelength division multiplexing (WDM) to copropagate multiple data channels at different wavelengths and multiply the total capacity without further increase in the baudrate. Thanks to EDFAs, WDM is relatively simple to implement and allows keeping the channels independent over long distances, making it popular for long haul transmissions. Optical frequency division multiplexing (OFDM) is also based on partitioning the optical spectrum and offers an improved spectral efficiency compared to WDM, at the price of a higher degree of complexity [20]. Polarization multiplexing doubles the bitrate by using the two orthogonal polarization states available in the waveguide. As these techniques are being pushed closer and closer to their maximum performances, other schemes are also investigated to increase the number of channels available by using multimode [21] or multicore fibres [22] and by occupying new wavelength ranges in the Mid-Infrared [23, 24]. Hollow core fibres are also proposed as a way to reduce latency and nonlinear distortions [25]. The evolution of WDM, AMF and OFDM is represented in figure 2.3, respectively by triangles, open circles and filled circles.

Pushing the limits of baudrate

Following the trend of earlier ETDM systems, optical methods have been developed to push the baudrate even further. Optical Time Division Multiplexing (OTDM) allows for generating a channel at a higher symbol rate by time interleaving lower baudrate channels. The realization of OTDM links at extreme baudrates was pionneered by M. Nakazawa at 640 Gbaud in 1998 [26]. Since then, several research groups around the world have passed the barrier of the Terabaud [27]. These advances in raw baudrate have been combined with advanced modulation formats, signal processing and monitoring methods compatible with such bandwidths [28]. Figure 2.4 presents the evolution of OTDM technology over the last 15 years.

We acknowledge the limitations of OTDM methods mainly due to the high sensitivity of the resulting signals. Although extreme baudrates are unlikely



FIGURE 2.4: Research was first focused on increasing the bit rate of 640 Gbaud signals via polarization multiplexing and high efficiency modulation formats. The current record for single channel bitrate is a remarkable 10.2 Tbit/s for a polarization multiplexed 1.28 Tbaud quadrature amplitude modulation signal.

to ever be implemented for long-haul transmissions, they could provide fundamental advantages in short distance systems including signal processing. As we will develop later in the text, some processing operations can be performed all-optically with notable latency and power efficiency advantages over electronics. OTDM signals may allow serial processing of a much higher data flow without requiring multiplication of the number of optical processing units for every separate channel of spectrally multiplexed systems such as WDM and OFDM.

Moreover, the energy cost of a serial operation would be similar for a signal at 40 Gbaud or at 1.28 Tbaud, showing a direct advantage of time multiplexing. Indeed, the efficiency of processing based on nonlinear optics mostly depends on the peak power of the signals, independently of the baudrate. As a consequence, using shorter pulses to squeeze more symbols with same duty cycle onto a single channel would change neither the average power nor the peak power of the signal, allowing processing at a higher bit-rate without increasing the power consumption.

Rather than using a single technology, next generation optical communication systems will likely be combinations of the parallel and serial techniques presented.

2.2 History of information processing

Concurrently with the development of communication systems, most civilizations have been looking for techniques to help the human mind with classifying, analyzing and calculating information. Computing may refer to the extremely general discipline of handling numbers. Here we restrict our scope to systems for automated calculation and quickly lay the historical background that led to present computers.

2.2.1 Development of contemporary computing methods

Similar to the development of telecommunications, automatic computing methods remained extremely basic until electric switching devices were introduced, then followed an explosive growth which continues today.

Some rare prehistoric evidence of calculation methods have been found, such as the Ishango bone used for elementary arithmetic as early as 18,000 BC. The Babylonian abacus was introduced around 2400 BC and programmable automates were realized from 850 AD [13]. Numerous automata, clocks and geared astrolabes were later developed. During the 17th century, mechanical calculators appeared, followed a century later by the ambitious - and never finished - project of a programmable mechanical computer by Charles Babbage [12]. It is only in the mid of the 19th century that companies like IBM or Bell Labs and the WW2 Nazi and Allied research programs initiated the production of computing machines for industrial and military purpose, that offered significant advantages over brain-powered calculus [14].

Since then, the need for processing power has increased relentlessly and it is only thanks to regular technological breakthroughs that one could keep up with this pace. Moore's law of exponential growth of the number of transistors integrated on a processor (Fig. 2.5, top) has remained valid since 1958, the year of the creation of integrated circuits. However, although the clock speed followed a similar trend until a couple of years ago, it now seems to be running out of steam (Fig. 2.5, bottom). It is the clock speed that ultimately limits the frequency of sequential (serial) instruction processing.

It is interesting to note the radical difference in the evolution of domains such as construction or agriculture and information science. Construction, for example, has had a relatively linear evolution across Human history, where extremely complex building techniques were already found back in the Antiquity. In information science, however, the methods remained very rudimentary until very late in history and, from then on, experienced an exponential boom. The mastering of electricity appears as a key enabling technology in both transmission and processing of information.

2.2.2 The success of digital electronics

Present information processors, mainly embodied by CMOS electronics, rely on binary logic for manipulating data deterministically. Binary representation meets the fundamental need for conservation and regeneration of logical states. By encoding data onto a finite number of physical levels separated by a guard band, the states can be conserved from one computational node to another, independent of the presence of noise, distorsions or imperfect response of the components. Given that the outputs can only adopt two discrete levels, every step in the processing circuit is also regenerative.

Binary representation of information also permits construction of any algebraic operation from the simple rules of Boolean logic. The fact that elementary Boolean functions can be realized on simple physical systems facilitated the implementation of binary computing.

Although some computing operations can be achieved with linear systems, most high level functions require a nonlinear transformation of the information, or switching. The elementary nonlinear component in CMOS processors is the transistor, which has now reached a high level of maturity in terms of cost, integration, cascadability and reliability.

2.2.3 Optical computing technology

The qualities of light as a vector of information are also attractive to support computation and an optical equivalent of the transistor is sought for realizing the vision of an all-optical computer. The idea of using light for computing has been around for a long time and was even expected to supplant electronics in logic circuits [29]. Many schemes of optical processors have been proposed for problems such as matrix operations [30] or general digital arithmetics [29].

Today, some disillusion has replaced the early enthusiasm. As seen in the previous section, most logic operations require nonlinear transfer functions. Optics is just the opposite, it is a field of very linear behaviour. Due to their bosonic nature, photons have extremely low interaction with matter. The inherent weakness of the interactions and the difficulty of cascading nonlinear optical nodes make most complex architectures impractical. This has constituted a big obstacle to the use of light for information processing and it was not until the late 80's that we saw logic operations based on nonlinear optics becoming a domain of active applied research. Recent advances in nonlinear platforms detailed in section 3.5 now enable elementary operations to be carried all-optically and new and faster applications are demonstrated every year.

Since then, the field of optical processing has taken the path of carrying
simple analog operations at very high bandwidth. Devices such as 3R regenerators [31] and impairments compensators [32] have been successfully demonstrated. Various logic gates and simple binary functions exploiting nonlinear optical effects have also been implemented [33].

2.2.4 Which place for optics in processing?

A closer look at figure 2.5 reveals a notable point. The degree of integration and cascadability of optical logic gates is still at an embryonic stage, while billions of transistors can be embedded on a CMOS chip, as seen on the top of the figure. A fundamental limitation exists to the integration of optical components. The smallest possible dimension of an elementary optical module is bound by the wavelength of light which is of the order of 1 μ m¹. In comparison, the wavelength of a relativistic electron of mass m_0 , charge e and at a potential U is given by

$$\lambda = \frac{h}{\sqrt{2m_0 eU}} \frac{1}{\sqrt{\left(1 + \frac{eU}{2m_0 c^2}\right)}} \tag{2.1}$$

where c is the speed of light in vacuum. The wavelength of an electron at a potential of 1 V, which is a typical value in electronic processors, is 1.226 nm. As a consequence, the minimum footprint of electronics is much smaller than any optical device.

Conversely, the available clock rate of optical components massively outperforms the state of the art in electronics. In the bottom plot, the clock rate of optical logic and signal processing devices (unlike the bandwidth of transmission systems) is compared with the rate of CMOS processors.

This may suggest that, at least in the near future, optics will not supplant electronics in mass computing. Nonetheless, it is conceivable that it will be used to perform very elementary operations at an extremely high throughput, at the manner of a coprocessor [35]. A present example of symbiosis between electronics and photonics, although it does not involve nonlinear processing, is the active research and development effort carried towards realization of photonic busses to transport data inside a classical CPU [36].

At this stage, the reader can still wonder what interest motivates such active research in optical processing, considering in particular the difficult-tobeat efficiency of digital signal processing [16]. The statement that optics is less efficient than electronics is based on the assumption that optical systems should mimic the operation of digital processors made of transistors. One can envision that new methods based on other principles than the usual binary logic

¹Although the field of plasmonics permits applications at scales smaller than the wavelength with structures confining light [34], it is unlikely that future systems will integrate optical elements significantly below this limit.



FIGURE 2.5: Evolution of electronic and photonic technologies.

would be able to exploit the richness of optical effects. Chapter 6 develops the idea that we would need to re-think or approach to automated calculus.

Finally, latency (defined later in this chapter) is another reason that motivates the use of light to carry functions that already exist in electronic systems. In chapter 5 we will show that optics can be used favourably in order to reduce the lag time due to the processing of operations.

2.3 Elements of information theory

In our historical introduction to telecommunications, we mentioned the existence of fundamental limits to the capacity of a transmission channel. This notion is essential in the quest for an ever growing bandwidth, as it sets a benchmark and an ultimate goal for the efficiency of the systems. The developments presented here are completely general and are not restricted to the case of optical transmissions [37].

Some preliminary theoretical study of telegraphic systems was initiated by Harry Nyquist in 1928 [38]. Nevertheless, the fundamental principles of communications had not been rigorously investigated until Claude Shannon formalized the notion of information in 1948 [39]. His systematic approach of information set a mathematical background to intuitive concepts such as data, encoding format, compression and noise. More importantly, he introduced a theoretical upper limit to the amount of information that can be carried by a noisy channel, which has had an immense impact on the area of telecommunications.

A good theoretical introduction to the field can be found in [40], including practical examples and demonstrations of most of the material summarized in this section.

2.3.1 Entropy

A central idea to the theory is the notion of *entropy* [39] as a measure of the number of binary units required to encode one symbol of a message. It is linked to the degree of randomness in the encoding of data. In other words, X being a variable with discrete values, its entropy H measures the uncertainty associated with the value of X. Supposing a message encoded with binary symbols, the entropy linked to one symbol depends on its likelihood to be a "0" or a "1". It is natural to set the entropy to zero if the value of the symbol is known in advance (no transmission of information is required) and to "1 bit" if the symbol has equal probability of being in both states. The general logarithmic definition of the entropy [40] for a message $x \in X$ that has

probability p(x) to occur with the given set of symbols is

$$H(X) = E_X [I(x)] = -\sum_{x \in X} p(x) \log [p(x)]$$
(2.2)

The entropy of a message constituted of N symbols is maximum when all the possible messages encodable on these N symbols are equiprobable. In the particular case of a message made of one single binary symbol with probability p(x = 0) = p to have value "0", the entropy in base 2 is

$$H(X) = -p(x = 0) \log_2 [p(x = 0)] - p(x = 1) \log_2 [p(x = 1)]$$

= -p \log_2 [p] - (1 - p) \log_2 [1 - p] (2.3)



FIGURE 2.6: Evolution of the entropy of one binary symbol with probability p of taking value 0.

The concept of entropy of a variable will be used in the next section to define the mutual information of two variables. As represented in figure 2.7, this quantity corresponds to the intersection of the conditional entropies of these variables.

2.3.2 Channel capacity

Our main motivation for briefly developing information theory is to introduce the maximum amount of information allowed through a communication channel in the ideal case. In practical terms, it sets a fundamental boundary to the highest possible data rate with presence of noise. In this section, we define the channel capacity and give the fundamental Shannon limit relation for a channel with additive Gaussian noise [40].

In the following development, X and Y are respectively the message coded before transmission through the channel and the received message. $p_X(x)$ represents the marginal probability distribution that a variable X takes value x. The joint probability distribution $p_{X,Y}(x,y)$ is the probability that X =x and Y = y simultaneously. If X and Y are decorrelated, $p_{X,Y}(x,y) =$



FIGURE 2.7: Left: diagram of a Shannon channel with additive noise. Right: Visualization of mutual information I(X, Y) as the intersection of the sets of states of variable X given a certain value of Y and vice versa.

 $p_X(x).p_Y(y)$. From there, the mutual information of X and Y is the mutual dependence of the two variables and is defined in base 2 by

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p_{X,Y}(x,y) \log_2\left(\frac{p_{X,Y}(x,y)}{p_X(x).p_Y(y)}\right)$$
(2.4)

It represents the information that both variables have in common, or the knowledge about one variable brought by knowing another variable. It is interesting to note that for decorrelated variables, the argument of the logarithm is 1 and the mutual information cancels.

The formal definition of the capacity C of a channel is [40]

$$C = \sup_{p_X(x)} I(X, Y) \tag{2.5}$$

where $sup_u V$ is the maximum value that the quantity V can take over a range of parameters u.

The noisy channel coding theorem applies to a channel of capacity C working at a data rate R with an error probability e at the receiver and states the existence of an encoding algorithm \mathcal{E} and a decoding algorithm \mathcal{D} : $\forall \epsilon > 0, R < C, \exists \mathcal{E}, \mathcal{D} \quad s.a. \quad e < \epsilon.$

In practical terms, the theorem asserts that a message coded as X sent by the emitter through a channel of capacity C can be received with an arbitrary precision as Y by the receiver, provided that the data rate R is lower than or equal to the capacity. The capacity C of the channel guarantees that the correlation between the message sent X and the message received Y is sufficient to allow arbitrary low error rate at a valid data rate.

In the particular case of a channel with additive white Gaussian noise, the capacity is expressed by equation 2.6, as derived by Claude Shannon [39, 40].

$$C = B.log_2\left(1 + \frac{S}{N}\right) \tag{2.6}$$

where B is the bandwidth (in Hz) of the channel and S/N is the Signal-to-Noise Ratio (SNR) with S and N being the average power of the signal and the noise (in W).



FIGURE 2.8: Maximum spectral efficiency defined by the Shannon limit, as a function of signal-to-noise ratio. The dashed line shows the asymptotical behaviour. Greyed area is not physically allowed.

For example, a channel with a SNR of 20 dB and a bandwidth of 40 GHz can, in theory, carry up to 266.3 Gbit/s of data. In the description of our experiments, we regularly use the idea of channel capacity as a measure of the amount of information that can be transmitted through an optical fibre.

2.3.3 Encoding formats

Shannon's theory states that there exists optimal transmitter coding and channel coding for maximizing the channel capacity. One major technique to increase data rates and push communication systems closer to the Shannon limit [41] is the use of advanced modulation formats [42] and check codes maximizing the information carried by every symbol [40].

Although widely used in many present applications, binary data encoding does not exploit the full capacity of a channel. A first immediate improvement would be the addition of multiple levels. For example, amplitude shift keying (ASK) uses multiple intensity levels. On top of an intensity modulation, optical carriers can also be encoded in phase using a phase shift keying (PSK) method, such as binary phase shift keying (BPSK) and quadrature phase shift keying (QPSK) [43]. Combination of both ASK and PSK leads to formats mixing phase and amplitude modulation, that fall under the generic term of "advanced modulation formats" [44]. In most cases, the various values of phase and amplitude that the signal can take are a finite number of states in the complex space arranged following a regular grid or pattern. The ensemble of the states that the signal is set to take is called "constellation". Figure 2.9 illustrates some constellations used in information systems.



FIGURE 2.9: Various types of advanced modulation format constellations. Graphs from [44].

However, the number of distinguishable states per symbols is limited by the SNR (or OSNR in optics). On top of the OSNR degradation due to the noise introduced by the channel, optical communications suffer from an additional limitation imposed by the nonlinearity of the optical media. This concern is raised in the next section.

2.3.4 Nonlinear Shannon limit

The Shannon limit derived in section 2.3.2 is a very general consideration that does not include the properties of a practical implementation. The hypothetical channel is supposed to introduce an additive noise in the transmission, but not to distort the signals [40]. The asymptotical behaviour shown by the dashed trace on figure 2.8 supposes a system able to reach infinite SNR. Given that a certain level of noise is unavoidable in real devices, this implies that the channel can stand arbitrarily high power levels with constant properties. In other words, the transfer function of the channel is purely linear. However, in the case of optical communications, increase in the optical power injected into an optical fibre leads to additional distortions due to optical nonlinearities [45]. These effects will be detailed in chapter 3. Over a certain threshold, the SNR improvement due to a further power increase does not counter-balance the detrimental effect of nonlinearity and the total capacity drops [46], as shown on figure 2.10 for various types of modulation formats.

This phenomenon was first pointed out by Partha P. Mitra and Jason B. Stark in 2001 (Bell Labs) [10]. They named it *nonlinear Shannon limit* for being an extension of the original linear theory.

Recently, with the growth of the capacity demand, the nonlinear Shannon



FIGURE 2.10: Nonlinear Shannon limit. Left: Predicted linear (black) and nonlinear (blue) Shannon limits and experimentally achieved spectral efficiencies (data points). Right: simulated spectral efficiencies for various types of encoding constellations. Graphs from [44].

limit has begun to pose an actual problem by limiting the bandwidth-distance product in long haul transmissions [47]. In optical networks, the attenuation due to fibres is usually compensated with erbium doped fibre amplifiers (EDFA) that introduce noise by amplified spontaneous emission (ASE). The use of higher-order advanced modulation formats requires higher OSNR to maintain an acceptable error-rate [11]. As the noise accumulated during transmission is determined by the channel, the power of the signal must be increased to keep the OSNR acceptable for longer distances, which enhances nonlinear effects. On the other hand, the addition of spans of optical fibre increases the interaction length and nonlinear distortions are even accentuated.

2.4 Latency

Bandwidth is usually the main parameter considered in a telecommunication network, be it for the choice of an ADSL provider by an end-user or the design of the internet backbone. New applications of datacom have driven the exponential growth of the amount of data exchanged as described in the introduction to this thesis. The good news is that, given sufficient resources, the total available throughput can virtually be increased arbitrarily by using techniques such as advanced modulation formats, OFDM, WDM, OTDM and by multiplying the number of fibre cores (spatial multiplexing). The recent transmission of an aggregate data rate of one petabit through a multicore fibre [48] demonstrates a link whose capacity would presently hardly find a use, even on the busiest sections of the internet backbone [49]. As a last resort, the bandwidth can simply be increased by multiplying the number of physical cables laid.

The issue, however, is more subtle in the case of latency. Latency is the time required for information to transit to the receiver and is mostly fixed. It is ultimately limited by the speed of light in vacuum [50]. Present silica optical fibres transport light at 68.6% of the speed of light in vacuum, where the reduction of the speed comes from the refractive index of their constituting material ($n_{\rm silica} = 1.44$ at 1550 nm). This allows at best for an improvement of 31.4% of the latency inherent to the fibres (for example by using hollow core fibres [25]), no matter the number of cables and the resources invested. Hence, the few other variables left for optimizing the time of flight of information are the minimization of the path length and the reduction of the processing time at each node.

Some niche applications such as high frequency trading between financial markets are already looking for even faster communication systems [50, 51]. Specialized financial networks have been designed to minimize electronic processing operations². They also consider seriously the technology of hollow core fibres as a way to beat the additional delay due to the refractive index of silica [52]. At a broader scale, the current evolution of computing tends to a more and more scattered distribution of information (the "Cloud") and an increasing distribution and parallelism in the processing of information (cloud computing). General audience products such as some mobile applications rely on algorithms running in distant computing centres (mobile voice recognition, songs identification, search engine,...). Cluster supercomputers and distributed computing has been used for years by the scientific community to exploit multiple workstations or centres for a single arduous task. Less parallelizable problems could benefit from faster exchanges between the tributaries [53]. More generally, the execution time of a process requiring pieces of data one after the other is directly impacted by the time required to transfer them.

This is why any reduction in latency would be beneficial to some present and future applications. In optical communications, optoelectronic conversion and digital electronic processing add delays that affect the overall transfer time. Some optical processing schemes exhibit significantly lower latency and could replace them in applications where time is critical.

 $^{^2}$ "We have literally done every optimization you can imagine." [50]

3 Nonlinearity in optics

Most optical phenomena observed in everyday life fall under the category of linear optics which explains common effects such as diffraction, refraction and interference. However, at high intensities, complex interactions can occur, which are described by the theory of nonlinear optics [54].

This chapter describes the theoretical background required for understanding the experiments and results presented in the rest of the thesis. A classical linear theory of light is first presented, then extended to describe propagation in nonlinear media. The central role of nonlinear optics in optical signal processing is justified by a section on general information processing principles, which state that nonlinearity is required for most logic operations.

3.1 Linear optics

The nature of light has been the subject of many historical debates. In 1845, Michael Faraday was the first to provide evidence of the relationship between light and electromagnetism. Two decades later, James Clerk Maxwell derived a mathematical theory of electromagnetism and concluded that light was electromagnetic radiation. The founding relations of electromagnetism still hold the name of Maxwell's equations. They describe the evolution of the electric and magnetic fields in vacuum and can be extended to the propagation in dielectric media by introducing the relative permittivity ϵ_r and relative permeability μ_r .

3.1.1 Equations of propagation

Light is an electromagnetic wave made of an electric and a magnetic field that are solutions of the Maxwell's equations. They are mathematically described by vector fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ that, at any location $\mathbf{r} = (x, y, z)$ in space and at any time t associate a vector of given magnitude and direction. Magnitude and direction are defined by the action of the field on a test charge q=1 C.

Maxwell's equations provide a complete description of the propagation of electromagnetic waves in vacuum. In their differential form, also called microscopic form, they come as a set of four coupled differential equations of **E** and **B**. Propagation in matter can be described by adding the constitutive equations defining the response of the medium. In the equations below, **J** is the current density, ρ the charge density, **D** the displacement field and **H** the magnetizing field.

$$\nabla \cdot \mathbf{D} = \rho \qquad (3.1) \qquad \nabla \cdot \mathbf{B} = 0 \qquad (3.3)$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (3.2) \qquad \nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \qquad (3.4)$$

These were first considered by James Clerk Maxwell (1831-1879) as the fundamental equations governing electromagnetism. They also respond to the common names of Gauss law (3.1), Faraday's law (3.2) and Ampere's law (3.4) from the names of their respective inventors.

In the following sections, we will consider dielectric materials with no macroscopic charge distribution $\rho = 0$ and equations for the electric and magnetic flux densities **D** and **B**.

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \qquad (3.5) \qquad \mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M} \qquad (3.6)$$

The materials studied, mainly silica and chalcogenide glasses, are non magnetic and contain no current and no electromagnetic field source, so that \mathbf{J} , ρ and the magnetic polarization \mathbf{M} are zero. Although a dielectric medium has a constant zero distribution of free charges ρ , the electric polarization \mathbf{P} represents the separation of bound charges that occurs in presence of an electromagnetic field.

For small optical powers, the polarization due to an external field is well modelled by a linear dependence of the electric field. This is valid as long as the excitation frequency remains out of the range of any material resonance, which is the case whithin the usual telecommunication transmission bands for silica and chalcogenide.

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \tag{3.7}$$

3.2 Nonlinear optics

In a propagation medium, light can exhibit more complex behaviours that must be understood as an interaction between the optical field and the atoms and molecules constituting the medium. Materials are made up of atoms comprising charged particles (i.e. protons and electrons) that are affected by the electric and magnetic fields. Although dielectric materials have a neutral charge distribution at a macroscopic scale, they comprise an equal number of positive and negative charges at the atomic scale. An incoming oscillating electric field \mathbf{E} exerts an action on the charged particles that are accelerated proportionally to its amplitude. The moving charges then radiate back an electric and magnetic field. As long as the perturbation of the initial trajectory of the charges is small, the displacement induced is directly proportional to the amplitude of the excitation. However, stronger deformations of the charge distributions may lead to break this proportional relation.

3.2.1 Intuitive model of nonlinearity

Figure 3.1 provides a naive representation of the nonlinear response due to the medium. In the Lorentz model, the electrostatic force maintaining the electrons on their orbit around the nucleus is modeled by a spring linking two charges. Application of an external electric field has opposite actions on the positive and negative charges. The resulting force F_e is canceled by the action of the deformed spring F_s . For small displacements x, the restoring force is proportional to the force applied $F_s = -\kappa x$.

The model is used to describe simple linear light-matter interactions. However, intuitively, a spring has only a limited range of deformation, ultimately limited by its total unwinded length. As a consequence, harder driving forces may push the spring away from its linear behaviour. The model can be extended to this case by adding nonlinear terms in the mathematical expression of F_s .

The power density required to stimulate nonlinear effects in optical materials is so high that they can usually not be observed in everyday life phenomena. Experimental display of nonlinear optics started from 1961 [55] with the use of intense coherent light sources and specific media. Later, strong spatial and temporal confinement of light and the development of materials with high non-linear response or long interaction length (presented in section 3.5) brought down the required threshold intensity. Thanks to these advances, even inefficient nonlinear mechanisms can now be exploited inside relatively compact and low power devices.

Some nonlinear interactions, such as harmonic generation or multiple wave mixing in second order $(\chi^{(2)})$ and Kerr media, have a timescale typically of the order of a femtosecond, determined by the atomic dynamics. In the context of



FIGURE 3.1: Intuitive representation of the origin of matterinduced optical nonlinearity. The force linking two opposite charges in equilibrium (e.g. an electron orbiting around a proton) is modelled by a spring. Application of an intense external force on the system strongly deforms the spring that can get off the linear regime.

our work, this ultrafast reaction of the medium, roughly two orders of magnitude faster than the switching speed of classical CMOS transistors, is one of the main arguments for using light in information technology.

3.2.2 Mathematical description of a nonlinear system

Before considering the particular case of optics, let us derive some general relations describing a nonlinear system. A simple look at the equations that we will write will underline the richness and complexity of nonlinearities. In their general form, they are often intractable and it could seem hopeless to even consider applying them to a practical problem. Strong assumptions will be made on the nature of the materials and optical fields with a view to simplify their expression. However the exercise of developing the global theory explains the wide variety of effects that can take place in a nonlinear optical material and suggests that care must be taken when approaching nonlinear optics theory. Also, the assumptions made to simplify the expressions may not remain valid in extreme cases of short light pulses, high powers and ultra fast phenomena, so that optical physicists may be conscious that explaining odd new effects might eventually require them to go back to the basic theory. For example, understanding the complex nonlinear interactions in dense WDM networks may need to consider a more detailed theory than the effects usually considered [56].

Another reason that motivates us to derive this relatively abstract mathematical background is that it is strongly related to the matter of section 6.1 and can be taken as a base for the theory of neural networks, which are another example of nonlinear systems. Although the introduction to neural networks given in this thesis is meant to summarize mostly qualitatively the elementary concepts with a minimum amount of mathematical expressions, they must be considered in the framework of generalized nonlinear systems for more advanced structures. Some types of continuous recurrent neural networks have even been implemented on optical and other physical systems, requiring the following complete theory to be understood. As a consequence, we find interesting to develop these very general considerations on nonlinear physics.

Our reasoning begins with a given nonlinear system S sketched in figure 3.2, having an input X and an output Y, not limited to optics. Before introducing



FIGURE 3.2: General nonlinear system with an input and an output.

the notion of nonlinearity, we will initially consider that S is linear. As a simple exercise of notation, we can express the output at time t as a function of the input.

$$Y(t) = \tilde{y} + \int_{-\infty}^{\infty} K^{(1)}(t, t') X(t') dt'$$
(3.8)

Here, we have not specified yet the nature (scalar, tensor, real,...) of X and Y. The kernel $K^{(1)}$ defines the first order temporal response of \mathcal{S} to the external excitation, which is sufficient for a linear system. The integration over the infinity depicts a system with memory, whose response depends on the history of its inputs. Note that the integration over times t' > t would suggest that \mathcal{S} can foresee the future. Therefore $K^{(1)}$ is zero for t' > t for a physically meaningful system.

If $X = (X_1, ..., X_M)$ and $Y = (Y_1, ..., Y_N)$ are vectors with respectively M and N components, the expression becomes

$$Y_n(t) = \tilde{y}_n + \sum_{m=1}^M \int_{-\infty}^{\infty} K_{nm}^{(1)}(t, t') X_m(t') dt' \quad (n = 1...N)$$
(3.9)

Let us now consider that the response of the system is nonlinear. The statement that the argument in the integral is a product of the kernel with the input is no longer valid, provided that its dependence in X is nonlinear. We have to express it as some function f of the stimulation.

$$Y_n(t) = \tilde{y}_n + \int_{-\infty}^{\infty} f_n(t, X(t')) dt' \quad (n = 1...N)$$
(3.10)

A very common approach to mathematical modelling of continuous nonlinear functions is the expansion in *Taylor-Maclaurin series*. Supposing that the series converges and that all the terms exist, a continuous function $f : \mathbb{R}^M \to \mathbb{R}^N : x \to y = f(x)$ is developed on a polynomial basis around \tilde{x} .

$$f_n(x) = \sum_{p_1=0}^{\infty} \cdots \sum_{p_M=0}^{\infty} \left(\frac{\partial^{p_1 + \dots + p_M} f_n}{\partial x_1^{p_1} \cdots \partial x_M^{p_M}} \right) \Big|_{\tilde{x}} \frac{(x_1 - \tilde{x}_1)^{p_1} \cdots (x_M - \tilde{x}_M)^{p_M}}{p_1! \cdots p_M!} \quad (n = 1...N)$$
(3.11)

The function is approximated at the order P by truncating the series at the P^{th} term. If the series converges, the higher order terms vanish for $P \to \infty$ and the quality of the approximation improves as P increases.

For better readability let us note

$$K_{n,p_1\cdots p_M} := \frac{1}{p_1!\cdots p_M!} \left(\frac{\partial^{p_1+\cdots+p_M} f_n}{\partial x_1^{p_1}\cdots \partial x_M^{p_M}} \right) \Big|_{\tilde{x}}$$
(3.12)

The expanded model of the response is obtained by substituting equation 3.11 in equation 3.10 and adding a time dependence to the coefficients.

$$Y_{n}(t) = \tilde{Y}_{n} + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{p_{1}=1}^{\infty} \cdots \sum_{p_{M}=1}^{\infty} K_{n,p_{1}\cdots p_{M}}(t, t', \cdots, t'^{\cdots \prime}) \\ .(X_{1}(t') - \tilde{X}_{1})^{p_{1}} \cdots (X_{M}(t'^{\cdots \prime}) - \tilde{x}_{M})^{p_{M}} dt' \cdots dt'^{\cdots \prime}$$
(3.13)

By permuting sums and integrals and grouping the terms of same differential order $p = p_1 + \cdots + p_M$, the previous equation takes the form

$$Y_{n}(t) = \tilde{Y}_{n} + \sum_{p=1}^{\infty} \left[\sum_{p_{1}+\dots+p_{M}=p} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} K_{n,p_{1}\cdots p_{M}}^{(p)}(t,t',\dots,t'^{\dots'}) \right] \\ \cdot \left(X_{1}(t') - \tilde{X}_{1} \right)^{p_{1}} \dots \left(X_{M}(t'^{\dots'}) - \tilde{X}_{M} \right)^{p_{M}} dt' \dots dt'^{\dots'} \right]$$
(3.14)

This development of the response of the system is called *Volterra series* according to the name of the Italian mathematician Vito Volterra (1860-1940).

3.2.3 Simplifications

We will now make general assumptions that will significantly reduce the complexity of the description obtained in the previous section. The first assumption made is that the system is **permanent**. In other words, the systems response does not change over time and an excitation at time t would give the same output as at time $t + \tau$.

$$Y_n (X(t+\tau)) = \tilde{Y}_n + \int_{-\infty}^{\infty} f_n (t, X(t'+\tau)) dt'$$

= $Y_n (X(t))$ (n = 1...N) (3.15)

This being true for any function, the kernels in equation 3.14 must verify $K_{n,p_1\cdots p_M}^{(p)}(t-\tau,t'-\tau,\cdots,t'^{\cdots'}-\tau) = K_{n,p_1\cdots p_M}^{(p)}(t,t',\cdots,t'^{\cdots'})$. In particular we can set $\tau = t$ to remove the variable t in the kernel.

$$K_{n,p_1\cdots p_M}^{(p)}(t,t',\cdots,t'^{\cdots'}) = K_{n,p_1\cdots p_M}^{(p)}(0,t'-t,\cdots,t'^{\cdots'}-t)$$
(3.16)

For simplicity, we will change the writing of the kernel by defining $\mathcal{K}_{n,p_1\cdots p_M}^{(p)}(t',\cdots,t'^{\cdots'}) = K_{n,p_1\cdots p_M}^{(p)}(0,-t',\cdots,-t'^{\cdots'}).$

The nonlinear optics phenomena studied here involve passive dielectric materials that contain no internal source of energy. Given that the limit of linear behaviour is valid for small optical fields $(\lim_{|X(t)|} \to 0)$, it is natural **to set the origin of the Taylor-Maclaurin expansion to** $\tilde{X} = 0$. Combined with the previous assumption, we can now write

$$Y_{n}(t) = \tilde{Y}_{n} + \sum_{p=1}^{\infty} \left[\sum_{p_{1}+\dots+p_{M}=p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{K}_{n,p_{1}\cdots p_{M}}^{(p)}(t-t',\cdots,t-t'^{\dots'}) \right] .X_{1}(t')^{p_{1}}\cdots X_{M}(t'^{\dots'})^{p_{M}}dt'\cdots dt'^{\dots'}$$
(3.17)

We can now introduce a notation to write 3.17 in a more compact form.

$$\mathcal{K}^{(p)} : X \otimes \cdots \otimes X = \sum_{p_1 + \dots + p_M = p} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathcal{K}^{(p)}_{n, p_1 \cdots p_M} (t - t', \cdots, t - t'^{\dots'})$$
$$.X_1(t')^{p_1} \cdots X_M (t'^{\dots'})^{p_M} dt' \cdots dt'^{\dots'}$$
(3.18)

The triple point operator : emphasizes the fact that each convolution is applied to a different variable of the kernel $\mathcal{K}_{n,p_1\cdots p_M}^{(p)}(t-t',\cdots,t-t'^{\cdots'})$ and symbolises the sum over all the combinations of p_1,\cdots,p_M such that $p_1+\cdots+p_M=p$.

The Volterra series can finally be written in its compact form.

$$Y_n(t) = \tilde{Y}_n + \sum_{p=1}^{\infty} \mathcal{K}^{(p)} \vdots X \otimes \dots \otimes X$$
(3.19)

In this thesis, we mostly consider that the nonlinear response of the material can be approximated by taking into account only the **third order of the** series. As a last notable simplification, we will assume that all **nonlinearities** are instantaneous. The reaction of the matter at a time t depends on the state of the external excitation only at that same time t, which allows to replace the kernel by Dirac distributions and remove the integrals over the time. This assumption must be made with care when dealing with ultrafast phenomena as in the experiments detailed in the following chapters. The OTDM setups involve light pulses as short as 300 fs and many nonlinear optical devices such as Semiconductor Optical Amplifiers (SOA) and electro-optical nonlinearities are far from having an instantaneous response at these timescales. The Kerr media employed in our schemes have ultra fast dynamics that allow for using the approximation of instantaneity with short pulses. However, this assumption of instantaneous response of the material is actually not entirely valid in the extreme case of the sub-picosecond pulses used in OTDM experiments. The timescale of the molecular vibrations in silica glass of the order of 100 fs is no longer negligible and leads to the Raman effect. Nevertheless, given that it does not have a significant impact on our setups, we do not develop its theory any further, but keep in mind that a rigorous approach must include Raman effect.

Finally, we would like to note that in all the previous equations, we have considered a **local response** of the material, i.e. the field radiated by the matter in one point of space depends on the incoming field only in that point. Introducing nonlocality would require adding integrations over all the spatial variables $\mathbf{r}_1, \ldots, \mathbf{r}_M$, which goes beyond the scope of this work.

3.3 Wave equation

The system of Maxwell's equations (3.1) to (3.4) describes the propagation of light. However, it is not easy to use in practical situations. With a few more assumptions, the theory can be put in the form of a single partial differential equation (PDE).

3.3.1 Linear wave equation

The Maxwell's equations can be combined into a wave propagation equation by taking the curl $\nabla \times$ of equation 3.2 and substituting 3.4, 3.5 and 3.6.

$$\nabla \times \nabla \times \mathbf{E} = -\frac{1}{c_0^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}$$
(3.20)

The differential equality $\nabla \times \nabla \times \mathbf{A} = grad \ div \mathbf{A} - \Delta \mathbf{A}$ for a given vector field *A* reduces 3.20 to a simpler wave equation describing the light propagation in a dielectric medium. In this notation, the Laplace operator $\Delta \equiv \nabla . \nabla$ corresponds to the divergence of the gradient of a function on the Euclidean space. The assumption that $\rho = 0$ implies that $div \mathbf{E} = 0$ and leads to equation (3.21). Note that this assumption is no longer valid in the nonlinear case developed in section 3.3.3.

$$\Delta \mathbf{E} = \frac{1}{c_0^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu_0 \frac{\partial^2 \mathbf{P}}{\partial t^2}$$
(3.21)

With the definition of the relative permittivity $\epsilon_r = 1 + \chi$, one can rewrite the wave equation in a dielectric medium as equation (3.22).

$$\Delta \mathbf{E} = \frac{\epsilon_r}{c_0^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{3.22}$$

3.3.2 Modes of propagation

The electric field of amplitude $A(\mathbf{r}, t)$ and average angular frequency ω_0 is given in equation (3.23) where $\phi(t)$ contains the deviations of the phase from the linear phase $\omega_0 t$.

$$\mathbf{E}(\mathbf{r},t) = \mathbf{A}(\mathbf{r},t)e^{i\phi(t)}e^{i\omega_0 t}$$
(3.23)

In this work, we consider that the fields propagate through single mode waveguides such as optical fibres or integrated ridge waveguides. The mode corresponds to the spatial distribution of optical intensity guided by the waveguide and solution of the Maxwell's equations. In the particular case of waveguide designs that only allow propagation of one mode, the mathematical description of the field is simplified by removing the transverse spatial dependence (the field is invariable with the spatial variables x and y). The electric field can then be described by a field depending on a single longitudinal direction of propagation z. Moreover, most of our experiments are well described by signals on a single transverse polarization state, although the general case comprises both components.

The field, written $E_{x,y}$, takes the form (3.24) where $A_{x,y}(\omega)$ is the amplitude, with components only in the transverse plane (x, y).

$$E_{x,y}(z,\omega) = A_{x,y}(\omega)exp(i\beta z) \tag{3.24}$$

The longitudinal propagation is governed by a propagation constant $\beta(\omega)$ or "wave number" that describes the change in spectral phase per unit of distance.

3.3.3 Nonlinear polarization

The model constructed in section 3.2 approaches the nonlinear regime as a deviation from the linear behaviour at high energies. Maxwell's equations remain valid with a term of nonlinear polarization added to the linear constitutive relation 3.7, which only describes the limit for small optical fields. The

linear theory is extended by perturbative method to take the nonlinearity into account.

We have stated in section 3.1.1 that the constitutive relation describes completely the interaction of light with a dielectric medium. Given that the nonlinear component of the field is entirely due to the radiation from matter, it can be included in the expression of the polarization.

Equation 3.7 can be modified to include an anharmonic response. Section 3.2 concluded that a nonlinear behaviour can be approached arbitrarily close by a series expansion that can be truncated for numerical simplicity. Here we develop the expression of the nonlinear polarization as the three first orders of a Volterra series under the assumption of permanence, locality and instantaneity used previously. The instantaneous response allows us to remove the integrals over time, hence write the product sum with only the double (for the second order) or triple (for the third order) point notation.

$$\mathbf{P} = \epsilon_0 \chi \cdot \mathbf{E} + \epsilon_0 \chi^{(2)} : \mathbf{EE} + \epsilon_0 \chi^{(3)} : \mathbf{EEE} + \dots$$
(3.25)

The three terms in equation 3.25 are successively the linear, second order and third order polarizations corresponding to the susceptibility tensors of the first, second and third orders χ , $\chi^{(2)}$ and $\chi^{(3)}$.

3.3.4 Second order term

In general, the magnitude of the quadratic term is significantly bigger than the third order term and $\chi^{(3)}$ can be neglected. The polarization then expresses as

$$\mathbf{P} = \epsilon_0 \chi \cdot \mathbf{E} + \epsilon_0 \chi^{(2)} : \mathbf{E} \mathbf{E}$$
(3.26)

Although the study of $\chi^{(2)}$ effects is an important branch of nonlinear optics, we do not develop this part of the theory, given that we will not make further use of it due to the reasons given below.

3.3.5 Third order term

=

In the particular case of centrosymmetric media, also called Kerr media, such as amorphous silica glass, the quadratic term vanishes. Materials with an isotropic symmetry such as amorphous glasses imply that an inversion of the direction of the incoming field must cause an inversion of the polarization. Because all even orders in the Taylor expansion 3.25 force the polarization to change sign, we can conclude that $\chi^{(2)} = 0$.

$$\mathbf{P^{(2)}} [-\mathbf{E}] = \epsilon_0 \chi^{(2)} : (-\mathbf{E})(-\mathbf{E}) = \epsilon_0 \chi^{(2)} : \mathbf{E}\mathbf{E} = \mathbf{P^{(2)}} [\mathbf{E}]$$
(3.27)

$$-\mathbf{P}^{(2)}[\mathbf{E}] \implies \chi^{(2)} = \mathbf{0}$$
(3.28)

In that case, the first nonzero nonlinear term is $\mathbf{P}^{(3)} = \epsilon_0 \chi^{(3)}$: **EEE** and the polarization comes to

$$\mathbf{P} = \epsilon_0 \chi \cdot \mathbf{E} + \epsilon_0 \chi^{(3)} \cdot \mathbf{E} \mathbf{E} \mathbf{E}$$
(3.29)

Slowly varying envelope approximation

The analytical study of third-order nonlinear effects requires some more approximations. One major assumption is the slowly varying envelope approximation (SVEA). At a given exact angular frequency ω , light is fully characterized by an amplitude, a phase and a harmonically oscillating term. In optical communications, the factor $e^{i\omega t}$ is often called "carrier" as it represents a monochromatic wave at angular frequency ω that can be modulated and modified to encode data. The amplitude and phase term $A(\omega, t)e^{i\phi(\omega,t)}$ is supposed to have slow variations regarding the carrier oscillations. This separation of the fast varying component $e^{i\omega t}$ and slowly varying terms $A(\omega)$ and $\phi(\omega)$ is the basic idea of the SVEA that allows for a simplified description of nonlinear propagation.

Nonlinear pulse propagation

The expression (3.29) of third order nonlinear polarization, combined with the wave equation 3.21, describe the evolution of light in a nonlinear material. Under the assumption of the SVEA, the one-dimensional equation for optical pulses that includes dispersive effects in the medium is called Nonlinear Schrödinger Equation (NLS) for its similarity with its quantum mechanical equivalent.

$$\frac{\partial A}{\partial z} + \beta_1 \frac{\partial A}{\partial t} + \frac{i\beta_2}{2} \frac{\partial^2 A}{\partial t^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial t^3} - \frac{i\beta_4}{24} \frac{\partial^4 A}{\partial t^4} + \frac{\alpha}{2} A = i\gamma |A|^2 A \qquad (3.30)$$

In the equation above, we included the effects of dispersion up to the fourth order with the coefficients β_k (k = 0, 1, 2, ...) called k^{th} order of dispersion. This degree of precision in the model is required when dealing with extremely short pulses below the picosecond regime. Note that the term β_1 can be eliminated by changing the coordinates to the referential of the pulse. The nonlinear coefficient γ translates the strength of the nonlinearity and is defined as

$$\gamma = \frac{n_2 \omega_0}{c A_{\text{eff}}} \quad \left[W^{-1} m^{-1} \right]$$
(3.31)

where A_{eff} is the effective mode area. n_2 is the nonlinear refractive index coefficient and can be deduced from the real part of $\chi^{(3)}$. It is responsible of the intensity dependence of the refractive index.

$$n_2 = \frac{3}{8n} Re\left[\chi_{xxxx}^{(3)}\right] \quad (3.32) \qquad n = n_0 + n_2 \left|E\right|^2 \quad (3.33)$$

A useful notion in nonlinear optics is the definition of nonlinear length L_{NL} as the typical distance that a light wave has to travel inside a given medium to experience significant nonlinearity. The point where the nonlinear phase shift $\Delta\phi_{NL} = 1$ is taken as a reference of strong influence of the nonlinearity. With $\Delta\phi_{NL} = \Delta k \ L = \frac{\omega_0}{c} \ L \ n_2 \ |E|^2$ we obtain

$$\Delta \phi_{NL} = \gamma \ P \ L \tag{3.34}$$

where we have taken the total power in the waveguide section $P = |E|^2 A_{\text{eff}}$. From there, it is natural to define the nonlinear length as

$$L_{NL} = \frac{1}{\gamma P} \tag{3.35}$$

This distance translates the strength of the nonlinear interaction. Comparison with the distance actually travelled by the light determines whether or not the nonlinear effect is significant. It is also interesting to compare L_{NL} with the dispersion length L_D defined in (3.44) to determine whichever of the dispersive or nonlinear effect is dominant along the propagation.

3.4 Propagation in waveguides

Among the wide range of phenomena that can occur in optical media, some mechanisms can be described with relatively simple expressions from the theory developed earlier. Here we only consider the effects directly relevant to our experiments. Sections 3.4.1 to 3.4.3 focus on strictly linear effects, while sections 3.4.4 to 3.4.6 consider nonlinear systems.

3.4.1 Linear loss

Propagation of light through matter usually leads to attenuation of the initial wave. For telecommunication applications, loss is a critical parameter of optical fibres that largely contributes to limiting the maximum distance over which information can be sent without regeneration.

To compensate for losses, optical amplifiers are commonly inserted in long haul links. However, they introduce noise in the system and reduce the OSNR. The OSNR degradation impairs the quality of the signals and, consequently, limits the number of reamplification stages along the link.

3.4.2 Dispersion

The index of refraction n of a material is generally not constant over the whole spectrum, so that waves of various wavelengths can propagate with different group velocities. This effect is particularly relevant in the case of ultra high speed communications where signals extend over broad bandwidths. Short optical pulses are also strongly affected by dispersion as they exhibit a wide spectrum. The spectral components of the pulses travel at different speeds, causing pulse broadening.



FIGURE 3.3: Effect of anomalous chromatic dispersion on a short optical pulse. The wavelength components composing the pulse travel at different speeds, resulting in time domain broadening at the output of the waveguide.

A Taylor development of the mode propagation constant $\beta(\omega) = n(\omega)\frac{\omega}{c}$ allows to take into account the variation of group velocity with the wavelength.

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{\beta_2}{2}(\omega - \omega_0)^2 + \frac{\beta_3}{6}(\omega - \omega_0)^3 + \frac{\beta_4}{24}(\omega - \omega_0)^4 + \dots \quad (3.36)$$

$$\beta_k = \left. \frac{d^k \beta}{d\omega^k} \right|_{\omega = \omega_0} \tag{3.37}$$

The β_k (k = 0, 1, 2, ...) have been defined previously as the successive orders of dispersion. In particular, the second order of dispersion is commonly named group velocity dispersion (GVD) and is often expressed with the parameter D, popular among telecommunication engineers. This parameter is better suited to fibre optics applications as it gives an estimate of the pulse spread as a function of the distance traveled and the bandwidth. Its units are [ps nm⁻¹km⁻¹]. Equation 3.38 gives the link between D and β_2 .

$$D = \frac{d\beta_1}{d\lambda} = -\frac{2\pi c}{\lambda^2}\beta_2 \tag{3.38}$$

Dispersion is commonly separated into two cases: the normal dispersion regime $(\beta_2 > 0)$ where the group velocity decreases for higher angular frequencies and the anomalous dispersion regime $(\beta_2 < 0)$ where the group velocity increases with frequency.

Given that dispersion is a linear effect, it can be totally compensated by applying an opposite amount of dispersion after propagation. Engineered specialty fibres such as dispersion compensation fibre (DCF) and dispersion shifted fibre (DSF) are commonly used. Other dispersion compensation schemes exist and are presented in chapter 4. In this work, the use of sub-picosecond pulses for OTDM data transmission required dispersion management up to the fourth order to maintain the pulse shape upon propagation.

Dispersion of Gaussian pulses

In most cases in this work, the temporal waveform of the pulses transmitted have a Gaussian shape of full width at half maximum (FWHM) T_{FWHM} and peak power P_0 .

$$U(t) = \sqrt{P_0} exp\left(\frac{-t^2}{2T_0^2}\right) \quad (3.39) \qquad \qquad T_0 = \frac{T_{FWHM}}{2\sqrt{ln(2)}} \tag{3.40}$$

The effect of second order dispersion on the pulse shape is given by the linear part of the partial differential equation (PDE) (3.30) restricted to its term in β_2 , that has solution 3.42 in the spectral domain.

$$i\frac{\partial A}{\partial z} = \frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} \qquad (3.41) \ U(z,\omega) = U(0,\omega)exp\left(\frac{i}{2}\beta_2\omega^2 z\right) \tag{3.42}$$

Dispersive propagation adds a quadratic spectral phase to the original electric field. This linear effect is entirely reversible and can be canceled by transmission through a waveguide of opposite amount of dispersion. Also, it appears from 3.42 that any method capable of applying a controlled quadratic phase of opposite sign can compensate for GVD. It is this property, extended to higher orders of dispersion, that will be used later for adaptive compensation of dispersion. The broadened pulse considered previously takes the chirped form

$$U_{\text{dispersed}}(z,t) = \frac{T_0}{\sqrt{T_0^2 - i\beta_2 z}} exp\left(\frac{-t^2}{2\left(T_0^2 - i\beta_2 z\right)}\right)$$
(3.43)

Consequently, a Gaussian pulse broadens proportionally to the distance traveled while maintaining its shape. We can define the dispersion length L_D as the propagation distance after which the pulse has broadened by a factor $\sqrt{2}$. This relative unit of the strength of GVD will be used later in chapter 4 and will be extended to higher orders of dispersion.

$$L_D = \frac{T_0^2}{|\beta_2|} \tag{3.44}$$

3.4.3 Polarization mode dispersion (PMD)

The birefringence of optical fibres and its change with respect to the wavelength introduce another range of possible distortions of temporal waveforms called PMD. Birefringence is, by definition, a dependence of the refractive index on polarization. In most cases, field propagation in optical materials can be separated onto two principal components (the slow and fast axes). These correspond to the eigenvalues of the propagation constant $\beta(\omega)$. The parts of the field propagating along the two axes experience different refractive indices, hence different group velocities [57].



FIGURE 3.4: Effect of DGD on an optical pulse. An incoming pulse with random polarization sees its components along the fast and slow axis (P_{fast} and P_{slow}) traveling at different speeds, resulting in detection of a double pulse at the fibre output.

A randomly polarized optical field sees its components parallel to the slow and fast axis traveling at different speeds. As a consequence, the two components of a pulse would pass by a given point of the medium at different times. After propagation through an optical fibre, the birefringence leads to formation of a double pulse and alters the waveform of the total signal as shown in figure 3.4. In communication systems, when the wavelength dependence of the birefringence can be neglected, this effect is known as differential group delay (DGD) or first order PMD. In the general case, the index difference between principal polarization states also depends on the wavelength; this is called higher-order PMD.

Again the effects of PMD become increasingly important with increasing bandwidth of the signals and the shortness of the pulses. In this work, we limit our study of PMD compensation to the first order as presented in chapter 4.

3.4.4 Self phase modulation

For pulsed optical waves, a first notable effect arising from third order nonlinearity is a change of the refractive index due to the variation of the optical intensity. This produces a time dependent phase shift, leading to spectral broadening. For a rigorous approach of the case of ultra narrow pulses, the NLS equation 3.30 must be solved. However, a simple intuitive development is possible for broader FWHM that allows for neglecting dispersion. Although this assumption does not allow a complete description of our extreme case, we will limit ourselves to a qualitative reasoning, knowing that experimental measurements can diverge from this simplified theory. The intensity induced variation of the refractive index (3.33) suggests that the progressive increase and decrease in optical power linked to a pulse propagating produces an instantaneous time-dependent phase shift. Considering a Gaussian pulse of equation (3.39), the rising and falling edges create a change in refractive index.

$$\frac{d n(|U|^2)}{dt} = n_2 \frac{d |U|^2}{dt} = \frac{-2 t}{T_0^2} n_2 P_0 \exp\left(\frac{-t^2}{T_0^2}\right)$$
(3.45)

Given the definition of the instantaneous phase $\phi(t) = kL - \omega_0 t$ after a travel distance L, the presence of nonlinearity causes a shift in instantaneous angular frequency.

$$\omega(t) = -\frac{d \phi(t)}{dt} = \omega_0 - n_2 \frac{\omega_0}{c} L \frac{dn \left(|U|^2\right)}{dt} = \omega_0 \left(1 + \frac{2 n_2 P_0 L}{c} \frac{t}{T_0^2} exp\left(\frac{-t^2}{T_0^2}\right)\right)$$
(3.46)

This results shows that the front of the pulse, in the region t < 0, is affected by a decrease in instantaneous frequency, while its back sees an increase. SPM occurring in a material with anomalous dispersion leads to the compression in time of the waveform as this dispersion regime causes the lower frequency components to travel slower than the higher frequencies [58]. As a consequence, the tail of the pulse ($\Delta \omega > 0$) tends to catch up with the front ($\Delta \omega < 0$) and the FWHM decreases. This effect is used intensively in our OTDM transmission experiments to compress pulses in time by spectral broadening. It is also the technique employed for obtaining a quasi flat frequency comb spectrum in our logic experiments.

3.4.5 Cross phase modulation

An intense signal coupled with another light wave mutually influence each other when copropagated inside a $\chi^{(3)}$ medium. Intuitively, a strong signal modifies the refractive index of the Kerr medium locally and this change is also seen by another field propagating on the same path. Hence a signal at one frequency can change the phase of a signal at another frequency without energy transfer.

According to the expression of the nonlinear polarization (3.29), two optical fields E_1 and E_2 of identical polarization and respective carrier angular frequencies ω_1 and ω_2 generate a third order component at the original angular frequency ω_1 , written

$$P_{\omega_1}^{(3)} = \frac{3\epsilon_0 \chi_{xxxx}^{(3)}}{4} \left(\underbrace{|E_1|^2}_{SPM} + \underbrace{2|E_2|^2}_{XPM} \right) E_1$$
(3.47)

The part depending on the intensity $|E_1|^2$ corresponds to the self phase modulation effect. The term in $|E_2|^2$ is responsible for cross phase modulation (XPM) and is twice as efficient as SPM. It appears clearly in this expression that the field E_1 is affected by the intensity of the other one $|E_2|^2$ through the nonlinearity of the material.

XPM is the basis of our all-optical radio-frequency spectrum analyzer used in chapter 4 as an impairment monitor for ultra high speed signals.

3.4.6 Four-wave mixing

Up to now we have only analyzed a couple of particular cases of equation (3.30) with optical fields at one or two distinct frequencies. However the form of the third order polarization (3.29) suggests a possible interaction between four waves. Under some conditions, this four-wave interaction can actually occur and requires considering all the combinations of terms in the nonlinear polarization.

In the case where an incoming field with three frequency components is introduced in the expression of the third order polarization, SPM, XPM and FWM terms appear. The fourth wave is generated by the nonlinear interaction and commonly named idler product.

$$P_{\omega_{4}}^{(3)} = \frac{3\epsilon_{0}\chi_{xxxx}^{(3)}}{4} \left(\underbrace{\left(|E_{4}|^{2} + 2|E_{1}|^{2} + 2|E_{2}|^{2} + 2|E_{3}|^{2}\right)}_{SPM} E_{4} + 2E_{1}E_{2}E_{3}e^{i[(k_{1}+k_{2}+k_{3}-k_{4})z-(\omega_{1}+\omega_{2}+\omega_{3}-\omega_{4})t]} \rightarrow FWM 1 + 2E_{1}E_{2}E_{3}^{*}e^{i[(k_{1}+k_{2}-k_{3}-k_{4})z-(\omega_{1}+\omega_{2}-\omega_{3}-\omega_{4})t]} \rightarrow FWM 2 + \cdots \right)$$

$$(3.48)$$

The condition to allow the FWM contributions for building up along propagation is that the arguments of the complex exponentials are null. Otherwise, the fields generated at different coordinates z are not always in phase and cyclic constructive and destructive interferences avoid growth of the product generated. The condition on the wave numbers k is commonly called the *phase-matching condition*.

$$k_4 = k_1 + k_2 + k_3 \to \text{FWM 1}$$
 (3.49)

$$k_1 + k_2 = k_3 + k_4 \quad \to \text{FWM 2} \tag{3.50}$$

The first condition is hardly met in practice and the corresponding FWM term usually vanishes. The second case corresponds to two waves giving energy to two new waves at angular frequencies ω_1 and ω_2 and is easier to verify in fibre and waveguide applications.

Solving of the wave equation by making the approximation of slowly varying envelope leads to a simplified relation of evolution of the amplitude envelope A. The dispersion and propagation terms have been removed for simplicity in the following formal equation.

$$\frac{\partial A_4}{\partial z} = i\gamma \sum_{p_1+p_2+p_3=3} \xi_{p_4p_1p_2p_3} A_1^{p_1} A_2^{p_2} A_3^{p_3}$$
(3.51)

This equation results in a system of four coupled mode equations governing the generation of the various products.

$$\frac{dA_1}{dz} = i\gamma \left(\xi_{11} |A_1|^2 + 2\sum_{j \neq 1} \xi_{1j} |A_j|^2\right) A_1 + 2i\gamma \xi_{1234} A_2^* A_3 A_4 e^{i\Delta k \, z} (3.52)$$

$$\frac{dA_2}{dz} = i\gamma \left(\xi_{22} |A_2|^2 + 2\sum_{j \neq 2} \xi_{2j} |A_j|^2\right) A_2 + 2i\gamma \xi_{2134} A_1^* A_3 A_4 e^{i\Delta k \, z} (3.53)$$

$$\frac{dA_3}{dA_3} = i\gamma \left(\xi_{22} |A_2|^2 + 2\sum_{j \neq 2} \xi_{2j} |A_j|^2\right) A_2 + 2i\gamma \xi_{2443} A_4 A_5 A_5^* e^{-i\Delta k \, z} (3.54)$$

$$\frac{1}{dz} = i\gamma \left(\xi_{33} |A_3|^2 + 2\sum_{j \neq 3} \xi_{3j} |A_j|^2\right) A_3 + 2i\gamma \xi_{3412} A_1 A_2 A_4^* e^{-i\Delta k} z_{(3.54)}^* \frac{dA_4}{dz} = i\gamma \left(\xi_{44} |A_4|^2 + 2\sum_{j \neq 4} \xi_{4j} |A_j|^2\right) A_4 + 2i\gamma \xi_{4312} A_1 A_2 A_3^* e^{-i\Delta k} z_{(3.55)}^*$$

Analysis of the FWM component of the coupled rate equations, for example (3.52), shows that the phase relations between the four waves at maximum gain respects the following equality. The field of amplitude A_1 results from the product of A_2^* , A_3 and A_4 , hence its phase is the sum of their respective phases. The factor *i* in the equation adds $\pi/2$ to the phase¹.

$$\phi_1 + \phi_2 = \phi_3 + \phi_4 + \frac{\pi}{2} \tag{3.56}$$

This phase relation is the founding principle of the optical dot-product scheme presentend in chapter 5.

In practice in this work, we only use a particular case of FWM called degenerate FWM. In this configuration, three initial waves are sent to the medium, with two of them having same frequency $\omega_3 = \omega_4$ (the two pump waves providing the energy). The phase matching condition is then easily verified and the effect, involving only three frequencies, is simpler to put into practice.

3.5 Nonlinear optical processing platforms

The study of nonlinear optics in the framework of information science has two main motivations. As already stated in section 2.2.2, nonlinearities are

¹In most publications, the $\pi/2$ phase offset is omitted because it has no observable effect in most real world configurations. The $\pi/2$ phase advance of the generated wave determines the direction of the energy transfer. Given that it does not influence directly our theory, we will omit it in the following chapters. I thank Thomas Büttner for the interesting discussion about this rarely cited fact.

Kerr	γ	$D \ [ps \ nm^{-1}]$	$\alpha [\mathrm{dB/m}]$	Typ. length
platforms	$[W^{-1}km^{-1}]$	$km^{-1}]$		[m]
HNLF [60]	11	0.035	2.3×10^{-3}	50
As_2S_3 chip	9.9×10^{3}	29 (over	60	0.06
waveg-		50 nm)		
uide [61]				
SOA [60, 62]	N/A	N/A	N/A	8×10^{-4}
Silicon	10^{5}	100	140	6×10^{-3}
chip [63, 64]				
PhC waveg-	9.8×10^6	> -1.92	1000	4.4×10^{-4}
uide [65, 66]		$\times 10^7$ (over		
		6 nm)		

TABLE 3.1: Common nonlinear platforms used in photonic signal processing.

required for creating basic switching elements which are the essential building blocks of a logic circuit. On top of this, nonlinear effects can cause detrimental distortions to long haul fibre optic communication signals. Extensive research has been carried out on the topic of reducing nonlinear impairments [25, 59], however these cannot be totally avoided. Here we do not consider the second case and aim at enhancing the optical nonlinear effects for processing purposes.

Nonlinear optical responses play a key role at many levels of this work so that the choice of optical materials with the right properties is decisive to the success of our experiments. Although all optical media exhibit nonlinear behaviour, only carefully engineered platforms provide sufficient nonlinear response to be used in processing. Table 3.1 summarizes the typical key characteristics of the most common nonlinear photonic platforms detailed in the following sections. The critical parameters to be taken into account when choosing a nonlinear device are its nonlinear coefficient γ , dispersion D, attenuation α , spatial footprint and nonlinear loss such as two photon absorption (TPA) α_{TPA} or free carrier absorption (FCA) α_{FCA} .

3.5.1 Highly nonlinear fibre

Classical optical fibres are typically designed to carry information with lowest possible loss and distortion. Maximum linearity is a key performance indicator of fibres, which is incompatible with the high nonlinearity sought here. However, optical fibre devices have great features that can be desired also for nonlinear systems. Low loss, high power handling, robustness and stable coupling with other devices are advantageous properties for nonlinear mechanisms that typically involve intense light waves, low conversion efficiency and complex setups.

Third order effects can be enhanced by increasing the interaction length, but long fibres are unwanted in integrated or high speed systems for reasons of increased loss, dispersion, latency and bulk . Special designs of waveguides have been developed to confine light in a smaller area of the core with a view to maximise power density. These are known as highly nonlinear fibres (HNLF) and are the most common platform for commercial products using optical nonlinearities. Depending on the purpose, manufacturers can engineer critical properties such as the zero dispersion wavelength (ZDW) and the dispersion slope.

Other specialty fibres use materials with high nonlinear coefficients such as chalcogenide compounds instead of pure silica. Although HNLF are well suited for many applications, they can become limiting for some applications requiring on-chip integration or very low dispersion. In the case of our 640 Gb/s and 1.28 Tbit/s experiments, the dispersion accumulated in even a couple of tens of metres of HNLF causes unacceptable walk-off between OTDM signals occupying a wide bandwidth.

3.5.2 Integrated chalcogenide platforms

A relatively recent platform on the scene of nonlinear optics is an integrated chalcogenide waveguide [61]. Current chips synthesized at the Australian National University (ANU) and used in our high bandwidth experiments are made of a silica substrate covered with an Arsenic trisulfide (As_2S_3) layer etched to form a ridge waveguide.

Chalcogenides (ChG) are molecules containing chemical elements from the Ist and VIth families of the periodical table. They have interesting optical properties such as extremely high optical nonlinearity and a transmission spectrum ranging from the near-infrared to the mid-infrared. Samples must be handled with care due to potential toxicity of Arsenic compounds present. They also suffer from limited power handling and photosensitivity that degrades their performances after prolonged exposure to visible and infrared light. Damages due to photosensitivity can be permanent after exposure to extreme intensity or can recover over time if the destructive power threshold was not reached. Although this constitutes an obstacle in the realization of commercial devices based on ChG glasses, proper calibration of the input power levels and protection of the samples from ambient light would allow for long term use of the components.

Figure 3.5 shows a cross section of a ChG chip. The silica substrate is covered by a 0.87 μm layer of highly nonlinear As₂S₃ material, etched by photolithography into a ridge waveguide of 0.38 μm height and 1 to 4 μm width. The active area of the chip is protected with a UV blocking polymer sheet. To avoid reflections due to index contrast at the facets of the chip,



FIGURE 3.5: Chalcogenide chip. Left: chip mounted on a coupling stage. Middle insert: microscope view of a lensed fibre coupled to a waveguide. Right: section of a chalcogenide ridge waveguide and mode profile. Mode profile is adapted figure from [67]

an anti-reflection coating is applied to the section of most samples. In our experiments, the chips had an approximate length of 6.5 cm.

A key property of the samples involved in high bandwidth experiments is their dispersion profile. The geometry of the waveguides is **dispersion engineered** to present a flat dispersion curve across the C- and L-band allowing for efficient phase-matched processes and walk-off free operation even with ultra wide bandwidth optical signals [68].



FIGURE 3.6: Dispersion profile of a dispersion engineered chalcogenide chip waveguide. Figure from [68]

Various features are implemented onto every chip and can be selected according to the purpose of the experiment by coupling to the respective inputs and outputs. The main variations available are straight waveguides of 1, 2, 3 and 4 μm width, as well as serpentines and "race track" designs for increased interaction length (up to 40 cm) within the active ChG area. Coupling of light into and out of the waveguide is done by aligning lensed fibres with its extremities. The single mode lensed fibres help focusing the beam into the area of the on-chip guided mode. The typical coupling loss at each facet is around 4.5 dB. Combined with a 3 dB propagation loss for straight waveguides, the overall insertion loss of the device sums up to about 12 dB. The total power that can be injected into the photonic chip is limited by the power handling of the samples. For our experiments with relatively high duty cycle, the average power was the limiting factor, with a maximum input at the lensed fibre of about 250 mW.

The combination of high insertion loss and limited input optical intensity were challenging for most of our experiments where the signals involved had high duty cycle. Due to the weakness of Kerr effects, the nonlinear mixing products received at the end facet of the waveguides had extremely low power levels and it was difficult to re-amplify them for detection with sufficient OSNR.

Figure 3.6 features a more detailed diagram of the facet of a 2 μm waveguide and emphasizes the importance of the polarization state of the light injected. The geometry of the waveguide is engineered such that the transverse magnetic (TM) mode has minimum dispersion in the 1450 nm - 1550 nm region of the spectrum. The transverse electric (TE) mode, however, exhibits higher dispersion level in this region and is not suitable for applications where low dispersion matters. Consequently, care must be taken to always align the polarization of the input field to the desired waveguide mode.

3.5.3 Other common nonlinear platforms

Many other nonlinear optical platforms exist and have different advantages and properties. We will not detail them here, given that we only used HNLF and chalcogenide waveguides in our experiments.

Among popular platforms, we can cite integrated semiconductor optical amplifiers (SOA) that present a fast saturation dynamics. The nonlinear effects typical of Kerr materials (SPM, XPM, FWM,...) can also occur inside a SOA. However, SOA have a limited response time that does not allow operation with signals over 40 Gbaud.

Periodically poled lithium-niobate (PPLN) waveguides are second-order platforms that permit ultra high bandwidth processing. Their operation at relatively low temperature can make them difficult to use in some situations.

Silicon chip waveguides are another promising type of devices. One of their major disadvantage, however, is the high level of TPA.

In the quest for compactness and integration, nonlinear optical platform evolved from bulky optical fibres and free space optical crystals to on-chip waveguides, reducing the dimensions by several orders of magnitude and allowing from multiple path to be written onto a single support. The most recent outcome of this trend is the development of photonic crystal (PhC) nonlinear waveguides increasing dramatically the strength of light-matter interaction and reducing the required propagation length by another factor 100. Precise design of the photonic crystal causes slowing down of the group velocity of light to increase the effective interaction with the material. It also allows for dispersion engineering of the waveguide. The effect of slowing down the propagation of the light inside the PhC is named *slow light*.

3.6 Nonlinear optics at work

In this chapter, we scratched the surface of the multiple effects that nonlinear optics offers. We selected a few of them for their interesting properties for signal processing purposes. In particular, SPM, FWM and XPM were detailed. These are the basis of all nonlinear operations introduced in the next chapters of this work.

We also described two of the existing platforms that exhibit Kerr nonlinearity: HNLF and integrated chalcogenide waveguides. They were intensively used in our experiments and their selection was a critical choice to achieve our results.

In the following chapters, we will detail schemes that exploit the background presented here in practical applications.

Ultra high baudrate signals and automatic compensation methods

The first part of our work is related to transmission of ultra high bandwidth single channel signals. These use extremely short pulses of light to transmit large amounts of data by time division multiplexing.

In the context of increasing demand for bandwidth in telecommunications, various strategies have been developed to keep up with this pace [42]. In section 2.1.4, we introduced the common methods adopted to encode extremely dense data onto optical carriers. Wavelength division multiplexing (WDM) [69], optical frequency division multiplexing (OFDM) [20] and advanced modulation formats (DQPSK, QAM, M-ASK) [70] are usually cited. Ultra-high symbol rate signals created by optical time division multiplexing (OTDM) of low bit-rate channels have long been envisaged as another promising approach but have more recently fallen out of favor for long distance communications.

Efforts towards increasing the baudrate were initially motivated by a historical observation: throughout the modern era, disruptive technologies allowing higher working frequency have been associated with dramatic reduction of cost per bit in telecommunications [71]. The frequency accessible with electronic components keeps increasing slowly, but at very high cost. In the last 15 years, optical techniques have smashed the records, enabling improvements by an order of magnitude. Although symbol rates as high as those presented in this work, of the order of the terabaud, might not be used commercially in the near future, it is likely that the technologies developed on the way will arouse interest for allying OTDM methods with other techniques. Serial transmission has advantages over parallel solutions in terms of routing and processing. Many nonlinear optical signal processing functions (for example 3R-regeneration [31, 72] and logic gates [73]) only apply to one binary channel at a time. Single channel signals can potentially reduce the cost of the system by processing of a higher bandwidth per device. Finally, the rate of operation of sequential instructions is ultimately limited by the baudrate of the data inputs.

Nevertheless, higher baudrate signals are more sensitive to channel distortions and handling signals of the order of the terabaud becomes extremely challenging. In this chapter, we first cover the techniques used to generate and monitor OTDM data. We then introduce novel approaches for stabilizing them and automatically compensating for various impairments fluctuating over time.

4.1 Generation and detection of OTDM data

Electronic data generators and electro-optical converters are typically limited to about 100 Gbaud. Higher baudrates are permitted by optical components, but no direct method exists for encoding such ultra dense optical channel from electronic data. However, higher baudrates are attainable by combining several low-bandwidth data streams into one by optical time division multiplexing. This domain was pioneered by M. Nakazawa in the 1990's with the groundbreaking demonstration of a 640 Gbaud link [26]. Symbol-rates above 1 Tbaud have been realized soon after [74] [75].

The total information is conserved by successively inserting a symbol from each source with a small temporal delay with respect to the previous one. To avoid overlap between neighbouring symbols, the time slots containing them must be extremely short. To combine N channels of bandwidth B into one at an aggregated bandwidth of N.B, the symbol period is $\frac{1}{N.B}$. As a consequence, the incoming signals must be encoded in Return-to-Zero (RZ) format on light pulses shorter than the symbol period.

The time-interleaving process is represented on figure 4.1. Each constituting channel j is delayed by an integer number of symbol periods $\frac{j-1}{N.B}$ before being added to the OTDM stream. Depending on the baudrate and optical source, the pulse width of the tributaries must be compressed in order to match the phenomenological criterion

$$T_{FWHM} < \frac{1}{2} \frac{1}{N.B} \tag{4.1}$$



FIGURE 4.1: Principle of time-interleaving of low baudrate channels. The N signals are first compressed and delayed respectively to each other, then combined into a single serial channel. For clarity, the timescale is not realistic

4.1.1 Interleaving stages

Experimentally, an OTDM implementation comprising N data generators is too costly and complex to run. Test beds usually employ a single low bit-rate data sequence encoded with a sufficiently long random pattern (PRBS 2^{31} -1). The original signal passes through successive interferometric stages that each double the baudrate. Inside each interferometer, the signal is split into two identical copies that are delayed relatively to each other and recombined. For N = 32, five successive stages are cascaded, with delays calibrated to obtain a final signal encoded with a 128 bits PRBS sequence.

Figure 4.2 describes the setup used in our OTDM experiments to generate signals at 640 Gb/s or 1.28 Tb/s. A pulse train at 40 GHz is produced by a modelocked optical clock synchronized with an electronic clock signal. The width of the optical pulses, 1.4 ps, is too broad to allow terabaud time multiplexing without overlap between the channels. Two successive compression stages using highly nonlinear fibres (HNLF) and band-pass filtering (BPF) [58] reduce the length of the pulses to about 275 fs (or 500 fs for the 640 Gb/s case), which are then encoded with a $2^{31} - 1$ pseudo random bit sequence (PRBS) by a Mach-Zehnder modulator (M-Z). The 40 Gb/s pulse sequence is then multiplexed to 640 Gb/s or 1.28 Tb/s with 4 or 5 interferometric multiplexing stages (Mux) interleaving 16 or 32 times the sequence from a single PRBS pattern generator.

The OTDM signal generation setup was originally designed by Prof. Jürgen Van Erps and Dr. Jochen Schröder [76]. Their initial configuration has been adapted with different bandpass filters and tuned in order to improve the pulse shape and adapt the centre wavelength of the signal.


FIGURE 4.2: Generation of a 640 Gb/s or 1.28 Tb/s PRBS signal. A modelocked fibre laser produces a pulse train which is then encoded with a Mach-Zehnder modulator. Two compression stages [58] ensure that the pulses are short enough to avoid overlapping after the time interleaving operation (16 or 32 times). The resulting signals are observed on an optical sampling scope trace (640 Gb/s) and with a second harmonic generation autocorrelator (1.28 Tb/s).

4.1.2 Bandwidth requirements

The fundamental relation of the time-bandwidth product originating from the properties of the Fourier transform dictates a link between the pulse width and the extent of the optical spectrum of the signals [77]. Pulses that present the smallest product $\Delta\nu$. $\Delta\tau$ allowed are called *transform limited* or *bandwidth limited*.

$$\Delta \nu \Delta \tau \geq 0.44$$
 (Gaussian pulse) (4.2)

$$\Delta \nu. \Delta \tau \ge 0.315 \quad (\text{Sech}^2 \text{ pulse}) \tag{4.3}$$

For ultra dense OTDM data, the channels are supported by a very wide spectrum. This imposes a strong limitation to the maximum baudrate of an OTDM transmission. Indeed most telecommunication systems rely on Erbium-doped fibre amplifiers (EDFA) that have a limited gain band. In our experiments, the 1.28 Tb/s signal occupies a big part of the C-band. Hence, all-optical signal processing operations involving effects such as FWM or XPM between multiple signals may require the use of an extended spectral band, for example both the C- and L-bands. This is the case for the OTDM demultiplexer and optical RF spectrum analyzer depicted in the following sections. Operation in the L-band makes the realization of the experiments more difficult, in particular due to the reduced performance of EDFAs in the L-band.

4.1.3 OTDM demultiplexing

Optoelectronic converters do not have the bandwidth required for direct detection of OTDM data. However, a single tributary channel can be extracted from the main stream and received at a baudrate accessible to photodetectors. Three main techniques have been demonstrated in the literature, using either nonlinear optical sampling gates [28], coherent detection with a pulsed local oscillator (linear optical sampling) [75] and time-lens time-to-frequency mapping [78]. In our case, the first method was used.



FIGURE 4.3: Demultiplexing of an OTDM signal by nonlinear sampling. Nonlinear mixing of the OTDM signal with a pulsed pump at lower repetition rate allows for extracting a single tributary channel. Left: time domain. Right: optical spectrum.

Nonlinear demultiplexing setup

Demultiplexing of high baudrate signals involves a second short pulse train (pump) at the repetition rate of the tributary to demultiplex. Passing the pump and OTDM data signals through a high speed photonic AND gate extracts one OTDM sub-channel. The wanted channel is selected by synchronization and adjustment of the relative delay between pump and signal. In this case, the AND gate operation is performed by degenerate FWM of the signal and pump at distinct wavelengths. This work exploits FWM in a highly nonlinear chip waveguide, but other schemes exist, for example based on a nonlinear optical loop mirror (NOLM).

The pulse train at the repetition rate of the tributaries is usually generated locally at the receiver, for example with a modelocked laser source. The synchronization of the pulse train with the OTDM data stream is a non trivial issue that involves clock recovery from the OTDM channel. Clock recovery at baudrates of 160 Gbaud and 640 Gbaud have been demonstrated using methods based on an electroabsorption modulator or nonlinear sampling with a phase locked feedback loop [79–81]. In the present work, back-to-back demultiplexing was performed without clock recovery, using the same clock source for both the data and pump modelocked lasers. This sufficed as a test bed for high bandwidth OTDM devices, given that no long distance transmission was attempted.

In our experiments, data and pump channels were centred at 1550 nm (Cband) and 1572 nm (L-band) respectively. The OTDM signal and the pump pulse train (FWHM=470 fs) were synchronized in time so as to coincide with the targeted tributary. The sampling operation took place in a dispersion engineered ($D = 29 \ ps \ nm^{-1} \ km^{-1}$ at 1550 nm) chalcogenide chip waveguide (see section 3.5.2) with sufficient bandwidth (about 50 nm) for propagating the OTDM signal and the pump with a dispersion induced walk-off negligible compared to the pulse width. The nonlinear parameter was $\gamma = 9.9 \times 10^3$ $W^{-1}km^{-1}$ and the attenuation inside the chip 60 dB/m over a total length of 0.06 m. Degenerate FWM generated an idler signal centred at 1605 nm thanks to the wide bandwidth [28] and anomalous dispersion characteristics of the chip waveguide allowing phase matching. For extreme rates of the order of 1 Thaud, the total bandwidth required exceeds 40 nm. Subsequent bandpass filtering and amplification of the idler channel produced the demultiplexed channel at 10 Gb/s. As detailed in figure 4.4, multiple successive bandpass filters were used to remove the unwanted pump, signal and ASE components from the optical field. Note that the final setup presented here is the result of a long evolution process outlined in appendix C.



FIGURE 4.4: Experimental setup for 640 Gbaud or 1.28 Tbaud demultiplexing down to 10 Gbaud tributary channels.

We modified the raw pump and signal channels using two FD-POP in order to improve their pulse shape and spectrum. Dispersion compensation of 2^{nd} , 3^{rd} and 4^{th} orders was applied, as well as spectral shaping in order to produce quasi transform limited gaussian pulses. This method also enabled flexible control over the centre wavelength and bandwidth of the signals. The effect of the reshaping can be seen for the data signal in figure 4.5.

Figures 4.6 and B.2 show the output spectrum of the photonic chip after nonlinear mixing occurred. The total power coupled to the input facet of the waveguide was limited to 23 dBm to avoid degradation or destruction of the sample. The coupling losses by lensed fibres $(2 \times 4.5 \ dB)$ and propagation loss along the selected TM mode of the waveguide (4.5 dB) summed up to an



FIGURE 4.5: Spectral reshaping of the OTDM signal by fitting a gaussian pattern onto the raw spectrum using a FD-POP.

aggregated insertion loss of about 13.5 dB. These losses, combined with the limited efficiency of the FWM process, produced an output idler signal with a power level around -45 dBm for a total pump and signal power of 23 dBm.



FIGURE 4.6: Error-free demultiplexing of a 640 Gbaud signal. Left: Optical spectrum at the output of the chip waveguide. Middle: autocorrelation traces of the pump (red) and signal (blue). Right: Eye diagram of the extracted idler.

The biggest issue to overcome was certainly the re-amplification of the idler signal from such ultra low power level. The low conversion efficiency of the Kerr process at the relatively low peak power imposed by the waveguide tolerance and high duty cycle of the signals produced an idler at a power close to the noise level of our most sensitive preamplifiers. Moreover, L-band amplification at 1605 nm is extremely challenging given the reduced efficiency of Erbium doped fibre amplifiers in this region of the spectrum. A low noise EDFA used just above its noise threshold was followed with a second booster amplifier and bandpass filters. This approach had been successfully demonstrated with a chalcogenide chip sample at a baudrate of 1.28 Tbaud in reference [28]. We attempted to reproduce these results with a view to use the system as a test bed for our automatic impairment compensator introduced later in this chapter. After numerous attempts with different chip samples, signal and pump generation schemes, we have not been able to achieve an error free bit-error rate at 1.28 Tbaud. The best error rate achieved to date was 5.2×10^{-5} at 1.28 Tbaud.

Repetition of this scheme with reduced bandwidth for both data signal and pump achieved error-free demultiplexing at 640 Gb/s. Bit-error rate was measured up to a value of 3.0×10^{-11} with data pulses of 510 fs of FWHM and pump pulses of 610 fs. With such broader pulses, the conversion efficiency remained similar to the one obtained in the 1.28 Tbaud configuration as the reduction in peak power was compensated by the lower repetition rate of the signal.

The most plausible explanation of the non reproducibility of the previous result at 1.28 Tbaud is the lower quality of data and pump signals, in particular in terms of timing jitter. For a 640 Gbaud signal, the timing jitter became negligible regarding the width of the broadened pulses and no longer impaired the error rate. These arguments are discussed in appendix B, based on jitter measurement of the various channels.

The OTDM demultiplexing setup was developed together with Dr. Evarist Palushani and Dr. Simon Lefrancois who both contributed significantly to the improvement of the experiment.

4.2 OTDM signal monitoring

The bandwidth of terabaud serial signals are far out of reach for any electronic device and cannot be measured with classical methods such as electrical sampling oscilloscopes or RF-spectrum analysis [82]. Moreover, high speed electronic devices are avoided as much as possible even in lower bandwidth systems due to their prohibitive cost. This raises issues regarding the monitoring of the quality of the signals in transmission systems.

Various methods have been proposed for monitoring ultra high baudrate signals. Ultra fast optical sampling oscilloscopes provide an averaged pattern of the signal in the time domain [83, 84], but present commercial devices hardly reach terahertz bandwidth except a few devices such as the Picosolve sampling scope. Optical second harmonic generation (SHG) autocorrelators have been used to access the pulse shape of OOK signals [85], but such devices are bulky and hardly integrable in real communication networks. Other methods based on an all-optical 2R regenerator [86] or a spectral broadening measurement [87] have been proposed.

Measurement of the RF-spectrum is a common technique for monitoring telecommunications signals. Electronic RF-spectrum analyzers are used for bandwidths typically up to 50 GHz [88]. However, these cannot be used for ultra-high baudrate signals that are far beyond the bandwidth of state-of-the-art electronics. In order to overcome this limitation, we made use of an all-optical equivalent of an RF-spectrum analyzer based on nonlinear optics [89] to infer the level of the impairments on the data stream. The method potentially allows for compact all-integrated measurement with a bandwidth of over 2.5 THz [67].

4.2.1 Principle of the optical RF-spectrum analyzer

We will start by defining the general notion of the RF-spectrum, then show how optical elements can be used to implement the concept.

Mathematically, the RF-spectrum, often called power spectrum, of a signal s(t) is defined as the Fourier transform of its modulus square $|s(t)|^2$.

$$U(f) = \mathcal{F}\left[|s(t)|^2\right] \tag{4.4}$$

It is important to notice that the RF-spectrum is different to the optical spectrum, which is the modulus squared of the Fourier transform of the field.



FIGURE 4.7: Principle of the optical RF-spectrum analyzer. A CW probe is copropagated with the signal under test in a Kerr medium and undergoes XPM. The sidebands appearing on the optical spectrum of the probe correspond to the RF-spectrum of the signal.

For a physical signal s(t), the definition corresponds to the Fourier transform of the intensity waveform. In the case of an optical field $s(t) = A(t)e^{-i\omega t}$, it is the frequency spectrum of the intensity envelope $I(t) = |A(t)|^2$ and does not reflect the fast carrier oscillation.

In optics, the modulus square of the Fourier transform of a field can be accessed by measuring its optical spectrum, for example by dispersing the beam on a grating device. To access the Fourier transform of the intensity waveform, we exploit the properties of cross-phase modulation effect in a Kerr nonlinear medium. Note that a two-photon absorption (TPA) detector can be used for a low-power autocorrelation of the signal, which is the Fourier transform of the RF-spectrum, but such measurement still requires scanning through a delay and does not provide a significant advantage over the method used in our experiments.

The Kerr effect in a medium of length L induces a nonlinear phase shift of the probe proportional to the intensity of the signal $\phi_{NL}(t) = k\gamma I(t)L$, where γ is the nonlinear coefficient and k is a constant. The CW probe field $E_p(t)$ experiences a change of its phase so that the modified field becomes

$$E_p(t) = E_p(t)exp(i\phi_{NL}(t)) \approx E_p(t)(1+i\phi_{NL}(t))$$
(4.5)

where a first order Taylor expansion approximation has been made with the assumption that the phase shift is much smaller than 1. The optical spectrum of the probe after cross-phase modulation is then proportional to the intensity spectrum of the signal under test (SUT). The principle is depicted on figure 4.7.

$$\mathcal{S}(f) = \left| \mathcal{F} \left[\tilde{E}_p(t) \right] \right|^2$$

= $\left| \mathcal{F} \left[E_p(t) \right] \right|^2 + k\gamma L \left| \mathcal{F} \left[I(t) \right] \right|^2$
 $\propto cste + \left| \mathcal{F} \left[I(t) \right] \right|^2$ (4.6)

A concrete implementation of the system is illustrated in figure 4.8. Copropagation of the SUT with a CW probe at another wavelength in a nonlinear medium induces spectral broadening on the probe due to XPM. The phase of the probe is modulated in proportion to the instantaneous intensity of the signal. Hence the optical spectrum of the probe reflects the Fourier transform of the square modulus of the optical field of the signal, which corresponds to its RF-spectrum. In these experiments, the nonlinear medium was a 6.5 cm long As_2S_3 chalcogenide waveguide which allows ultra high bandwidths up to over 2 THz [90], normally inaccessible with common nonlinear fibres. It also has the potential for an integrated signal monitoring solution [91].

The RF-spectrum of an optimized signal exhibits a strong DC component (the initial CW probe), a background power over the whole spectrum and tones at the carrier frequency and its multiples. This can be viewed on the measurement featured in figure 4.8. Distortion of the signal induces the appearance of tones at sub-multiple frequencies and a dramatic drop of the main tone power [92].



FIGURE 4.8: The RF-spectrum of the signal is measured with terahertz bandwidth by copropagating it with a cw probe at a different wavelength inside a highly nonlinear chalcogenide waveguide. XPM causes spectral broadening in the probe, reflecting the RF-spectrum of the signal. The tone power of the spectrum corresponding to the baudrate is extracted with a bandpass filter (BPF).

4.2.2 Application to impairment monitoring

As written previously, RF-spectrum analysis is a common technique for signal quality evaluation at bandwidths up to 100 GHz. Therefore the all-optical equivalent presented in the previous section can be used in a similar way for an OTDM channel. Techniques based on extracting characteristics of the RF spectrum have already been demonstrated for monitoring multiple impairments on high baud rate systems [90, 92]. In this work, the signal quality was monitored using the power of the fundamental tone in the RF spectrum (640 GHz or 1.28 THz according to the baudrate).

Figure 4.9 shows the RF-spectrum of the data signal measured with an optical spectrum analyzer (OSA). The various traces correspond to the SUT impaired with different levels of dispersion: optimized dispersion, and with 2^{nd} , 3^{rd} and 4^{th} order dispersion. The spectrum of the optimized signal has a maximum 1.28 THz tone power and no tones at submultiple frequencies. The addition of a dispersion offset for various orders all impair the RF spectrum by decreasing the fundamental tone power and generating side tones. The values of the dispersion orders β_m in figure 4.9 are given in the normalized units of dispersion length previously introduced in equation (3.44) for m = 2.

$$L_D^m = \frac{1}{\beta_m} \left(\frac{T_{FWHM}}{2\sqrt{\ln(2)}} \right)^m \quad \text{where } m \text{ is the order of dispersion.}$$
(4.7)

Interestingly, there is no need to measure the complete RF spectrum for this monitoring method. Simple narrow optical bandpass filtering with 30 GHz bandwidth centred on the main tone at the output of the RF-spectrum analyzer allows monitoring of the tone power on a slow photodetector.

We will show later that for other types of impairments, the same changes are observed: the power of the main tone reduces and new tones at lower



FIGURE 4.9: Evolution of the RF spectrum for deviations of the 2^{nd} (left), 3^{rd} (middle) and 4^{th} (right) orders of dispersion. For each series of spectra, the bottom graph corresponds to the optimized signal. The application of increasing dispersion reduces the power of the 1.28 THz tone (upper spectra). The dispersion orders β_m are given in the normalized units L_D^m (m = 2, 3, 4) (see equation 4.7).

frequencies are generated. This is caused by broadening of the signal pulses and overlap between subsequent pulses. Figure 4.10 shows the evolution of the tone power independently for the three first orders of dispersion, while all other parameters are fixed to their optimum values. It is interesting to note that there is no clear qualitative difference between the effects of the various orders of dispersion on the RF spectrum. The key innovation of our approach, explained in section 4.5, is that we do not have to distinguish between them.

4.3 Compensation of chromatic dispersion

As already mentioned, high bandwidth signals are increasingly susceptible to dispersion impairments. The ultra short pulses are strongly distorted by a non flat dispersion slope due to the extent of their spectrum. In addition to group velocity dispersion (GVD), higher-orders of dispersion significantly affects the transmission [57]. An accurate dispersion compensation solution is therefore crucial.

In present optical communications networks, the GVD of the fibre is balanced by spans of dispersion compensating fibre (DCF). At Tbaud rates, this method is no longer sufficient as higher orders of dispersion must be taken into account and even slight drifts of the dispersion coefficients (for example due to temperature fluctuations) can be detrimental. Also, the dispersion compensation must be extremely accurate as Tbaud pulses are broadened by a factor of 2 over a length of only 1 m of standard SMF. All previous experiments employed



FIGURE 4.10: Impairment on the signal quality (power of the 1.28 THz tone) as a function of three orders of dispersion. The dispersion orders β_m are given in the normalized units L_D^m (m = 2, 3, 4) (see equation 4.7).

elaborate compensation schemes using concatenated specialty fibres [93], phase modulators [94], tunable fibre Bragg gratings [95] [96] [97], MEMS [98], a spatial light modulator with a virtually imaged phase array [99] or parametric wavelength conversion [100] to compensate higher orders of dispersion. However, these compensation schemes were either static, narrow band, limited to GVD or did not include a dispersion measurement system. Therefore they were unable to compensate for temporal fluctuations due to stochastic phenomena like the temperature dependence [101] [102] of the dispersion coefficients, which can become significant in real telecom systems.



FIGURE 4.11: Example of phase pattern corresponding to combined values of β_2 , β_3 and β_4 (dashed black). The resulting phase is the sum of the quadratic (red), cubic (green) and quartic (bleue) orders of dispersion. Dispersion is emulated by implementing such spectral phase on a FD-POP.

In our case, the device used to control the dispersion state of the signal was a Finisar Waveshaper FD-POP [103]. The device can apply a custom spectral phase pattern to an optical field, based on a scheme similar to the one first demonstrated by Weiner et al. [104]. The light spectrum is spread spatially using a conventional grating onto a Liquid Crystal on Silicon (LCoS) array. The latter allows the phase of light reflected by every pixel to be independently controlled, so that a custom spectral phase can be applied. The modified spectrum is then recombined into the output fibre using the same grating. Further details on the working principle of a FD-POP can be found in Appendix A.

GVD corresponds to a quadratic spectral phase, 3^{rd} order dispersion to a cubic spectral phase, and so forth. Any combination of these linear effects can be obtained by summing the contributions of each order of dispersion as shown in figure 4.11. Given that the FD-POP is able to apply any phase pattern, it can create the spectral phase corresponding to the wanted amount of dispersion. The dispersion emulation is limited to about $\pm 60 \text{ps/nm}$ (for GVD) due to the delay dependent loss of the SPS [103]. However, this is more than enough to allow for residual dispersion compensation of terabaud signals in real world applications, typically drifts due to temperature changes which have been measured to be of the order of -0.003 ps nm⁻¹km⁻¹K⁻¹ for SMF in the worst case [102].

As dispersive effects are linear, the compensation device can be inserted anywhere between the transmitter and the receiver. In realistic reconfigurable optical networks, LCoS-based Wavelength Selective Switches (WSS) are already in use at network nodes and could in principle be used to achieve partial dispersion compensation [103]. Dispersion compensators based on digital signal processing can hardly be implemented in OTDM links, given that dispersion must be canceled before detection.

4.4 Compensation of polarization mode dispersion

The concept of PMD has been introduced in section 3.4.3 as the difference in group velocity between two orthogonal polarization states due to the birefringence of the transmission link. In general, this property varies with the wavelength. At a first order approximation, the wavelength dependence is neglected and PMD comes down to a differential group delay (DGD). Note that this assumption may not remain valid with the transmission of sub-picosecond pulses over long distances. DGD is defined as the delay between the two principal polarization components caused by the difference in travel time of an optical wave in the birefringent medium [105].

This effect can also be created by separating the principal polarization components, delaying one with respect to the other and recombining them [106]. The setup used as a DGD compensator in our experiments is shown in figure 4.12. The signal was split onto two orthogonal polarizations by a polarizing beam splitter (PBS). The polarization state of its input was set at 45° with respect to the principal axes of the PBS in order to match the polarization state arbitrarily set to 45° at the input of the DGD emulator inside our transmission link. This choice was artificial and would be subject to fluctuations in a real transmission application. Further development of our method could include addition of a computer driven polarization controller at the input of the DGD compensator, regulated as two additional degrees of freedom by the same optimization algorithm as used for driving our actual DGD compensator. Adding a polarization controller in the DGD compensator has been proved to improve the performance of the compensation scheme [107]. For higher-order PMD control, the wavelength dependence of the DGD should be included [105].



FIGURE 4.12: PBS split and recombine the two orthogonal polarization states. One arm is directed straight to the second PBS via a length of SMF and the length of the other is controlled with a programmable delay line. The degradation of the signal due to a DGD offset is recovered by applying an opposite DGD value using the compensator.

A relative delay was applied between both arms with a computer controlled programmable delay line (*General Photonics MDL-002*). This device permits a resolution of 1 fs, which was more than enough to accurately delay the 500 fs duration pulses. As we will see later, the 1 fs resolution was also required for fine stepping through the carrier cycle. A second PBS recombined the two polarizations to form the output beam. Polarization controllers (PC) were inserted into both arms in order to adjust the polarizations to match the transmitted states of the second PBS. For zero DGD, the delay on both arms must be identical.

In our setup, another kind of DGD generator, a commercial device (*General Photonics DynaDelay 40G*), was also employed to emulate the DGD fluctuations of the link. Its working principle, based on rotating birefringent crystals, differs from the compensator, but has a similar effect on the signal. However, the same device could not be used as compensator due to its limited resolution of 360 fs, which was too coarse compared with the pulse width of our

OTDM signals and would have made the compensation system inaccurate and unstable.

4.4.1 Adjustment of the interferometer

Initial setting of the programmable delay line was found by cancelling the spectral interference fringes at the output of the DGD compensator fed with polarized ASE noise, set with a polarization state at 45° with respect to the principal axes. A polarizer at 45° was placed at the output of the compensator to force the two output polarization states into the same plane to generate interference fringes, followed by an OSA. Fringes were observed on the spectrum with a period inversely proportional to the detuning of the delay between both arms as illustrated by the spectra on the right of figure 4.13. Zero relative delay value was found by setting the delay line so as to cancel the fringes.



FIGURE 4.13: Spectral interference patterns induced by relative delay of broadband signals. Left: illustration of two interfering beams originating from the arms of the X- and Y-polarizations. The delay prevents all the wavelengths to simultaneously coincide. Right: OSA measurement of interferences for various delays.

The diagram on the left of figure 4.13 pictures light waves recombining at the output of two arms of an interferometer with a relative delay (green and blue waveforms corresponding respectively to the arms for the X and Y polarizations in the case of the DGD compensator). When the signal has components at different wavelengths, introduction of a relative delay prohibits simultaneous constructive interference at every wavelength, creating spectral fringes.

4.4.2 Limitations of the DGD compensator

The polarization dependence of the RF-spectrum monitor used in parallel for our automatic compensation scheme introduced constraints on the DGD compensator. Indeed the continuous wave probe involved in the XPM process has to be polarized along the TM mode of the waveguide and requires the signal to have the same polarization state. The interferometric nature of the DGD compensator induces a polarization offset depending on the relative phase between the two arms, which would lead to a random polarization at the output of the device if no further care was taken. Sweeping through the phase to find constructive interference in the compensator (as shown in figure 4.14) ensures an optimum delay between both arms. Hence the polarization state at the input of the RF-spectrum analyzer is not affected by the action of the DGD compensator and enables for consistent measurement of the impairment.

The change of the DGD state was achieved in two stages. Firstly the delay was tuned to the rough value of the DGD wanted (with 0.1 ps resolution). Then the delay was stepped with 1 fs resolution over one period of the optical wave (about 5.2 fs for a signal at 1558 nm). During this second stage, the power of the 640 GHz tone of the RF-spectrum was recorded. The highest tone power detected determined the signal quality for that value of DGD. For this reason, each DGD optimization step during the experiment took about 10 seconds.

Although fine stepping through the delay greatly influenced the RF-spectrum measurement, the quality of the eye diagram of the signal did not change significantly as long as the device used to measure it was not polarization dependent. This was verified by monitoring the eye diagram with an optical sampling oscilloscope as one was stepping finely through the delay. The optimization process could have been dramatically sped up if a device similar to the one used as the DGD emulator, but with better resolution, could have been used as compensator since the *General Photonics DynaDelay 40G* is programmed to adjust the delay with consistent phase periodicity. Some additional noise appeared in our measurements due to the limited resolution of the delay line that could not reliably capture the maximum of the interference during the fine sweep. Note that this issue would not occur with a polarization insensitive RF-spectrum analyzer, for example built using buried waveguides.



FIGURE 4.14: Resolution limit of the programmable delay line. Time trace shows the RF tone power fluctuation resulting from sweeping the delay by 1 fs steps every 500 ms.

4.5 Automatic compensation of multiple impairments

In this work, we set our goal to control multiple sources of distortions and also compensate for them in real time. Environmental factors such as thermal changes, vibrations and changing stress on fibres may cause the parameters of the link to drift in time. Therefore, we aimed at combining signal monitoring with tunable compensators to maintain the transmission stable in spite of fluctuations.

Compensating the combined effects of multiple impairments requires controlling as many independent variables. Instead of measuring these separately, we base our approach on the single scalar information of the quality factor of the signal (the fundamental tone power measured from the RF-spectrum analyzer). For every parameter taken into account in the compensation, we first make sure that the quality indicator has a monotonically decreasing dependence around the optimum. This condition allows the existence of a global optimum for the system state, which can be found by running a multivariate scalar optimization algorithm on the N-D surface defined by the tone power (for N simultaneous parameters).

For instance, figure 4.10 suggests that the 2nd, 3rd and 4th order dispersion verify this monotonous dependence within a certain range around the optimum.

4.5.1 Automatic feedback optimization

For independent parameters verifying the condition of monotonicity, the map of the RF tone power presents a local maximum around the state of optimum signal. Within a given area in the space of the parameters $\xi = \{\xi_1, \xi_2, ...\}$, moving away from the optimum state in any direction leads to a drop in the RF tone. This area neighbouring the optimum is called the *domain of convergence* of the algorithm. We aim at finding the optimum combination of parameters $\xi^{(0)} = \{\xi_1^{(0)}, \xi_2^{(0)}, ...\}$ corresponding to a maximum RF tone power, also called the *target function* $f(\xi)$. In the cases considered in our experiments, the parameters optimized were GVD (β_2), two higher orders of dispersion (β_3 and β_4) and DGD (δ).

The use of a hill-climbing algorithm driven by the information of $f(\xi)$ allows to find $\xi^{(0)}$ on a trial/error basis.

Optimization algorithm

A decrease in the signal quality corresponds to a reduction in the tone power that is detected and activates the compensation procedure. The information provided by measurement of this single parameter of the RF spectrum is sufficient to control a blind optimization. Due to the fluctuations of the tone measured by the power meter, the algorithm must be robust to noise. Also, during the optimization process it is critical to keep the system operating within a narrow range around the optimum tone power in order to keep the data transmission up at all times. Those requirements prohibit the use of a classical conjugated gradient algorithm due to its instability in noisy configurations [108]. Indeed, this family of algorithms tends to create big swings in the variables optimized when the target function evaluation is inaccurate, which is undesirable for keeping the transmission working continuously. A fixed step steepest gradient method has been run successfully in simulations for simultaneous compensation of β_2 and β_3 . However, the method requires too many evaluations of the tone power for various dispersion states to be applicable when a third parameter is added. Here, we perform automatic compensation with a simple fixed step relaxation algorithm optimizing the tone power by trial and error with respect to the variables. Also called *fixed step Powell's* method [109], it was chosen for its simplicity and robustness to noise in the measurement. It was used with the parameters $(\beta_2, \beta_3, \beta_4 \text{ and } \delta)$ set as conjugate directions for the optimization.

The principle of the Powell's method is illustrated in figure 4.15. It first selects one variable as the active parameter, keeping the others fixed. Its value is changed by small steps in a random direction. If the tone power improves, a further change is made in the same direction otherwise the direction is inverted. After four recursive steps, the same process is applied to another variable. If the tone power has not fully recovered to the optimum value, the scheme is repeated until complete compensation of the impairments is achieved.

In the case of dispersion and DGD, we observed that the absolute optimum configuration could be approached only from within a window with a width of about one bit period on either side of the optimum, because the tone power changes monotonically as a function of the impairments approximately only within that window. This does not mean, however, that the largest errors that can be corrected are limited. Merely that tracking must start from within that window for the compensation to remain accurate. By running the monitor continuously, the compensated transmission link remains locked to the optimum in spite of any fluctuations of DGD and GVD on the link. Over time, the state of the parameters can be pushed very far from the initial configuration, as long as the instantaneous fluctuations are small enough to let the algorithm track the optimum. Once the algorithm is locked in, the drift of the parameters can cover the whole range allowed by the compensation devices.

Convergence

If the condition that the variables must evolve within the domain of convergence is not verified, the algorithm is likely to push the system away from the



FIGURE 4.15: Flow diagram of the optimization algorithm. Detection of a drop in the tone power activates the optimization process. The various parameters are successively considered in a loop modifying the active variable step by step so as to maximize the tone power.

optimum and reach another local maximum of tone power. Also, the measurement noise of the objective function can cause the parameters to hop in a wrong direction and eventually lead to a local maximum if the state is too close to the border of the domain of convergence.

Figure 4.16 compares three cases of the behaviour of a compensated transmission link affected by a sharp change in one parameter (DGD in this case). If the compensation feedback is inactive (left), the signal is permanently degraded by the change. When the compensation system is used within its working range (middle), fluctuations are cancelled and the signal recovers. If the compensator is active, but the fluctuations applied to the link are bigger than the permissible range (see for example figure 4.23, right graph: the tone power rises with a period equal to the period of the signal), the optimization process is likely to diverge to a local maximum leading to a permanent mismatch. Hence the signal cannot be recovered. Similar behaviour is observed for dispersion compensation.

The previous statement that the tone power reflects the quality of the signal must be considered carefully, in particular given the limitations discussed in section 4.4.2 about the DGD compensator. Indeed, the strong fluctuations appearing occasionally on the RF tone time trace (e.g. at 150 s in figure 4.25) are due to errors in the measurement and do not imply that the signal itself was impaired. These random errors do not affect significantly the stability of the link, as they influence the control of the compensators for only one step,



FIGURE 4.16: Time trace of the tone power monitoring a system subject to sharp DGD changes. (left) The automatic DGD compensation system is turned off, the degraded signal is not recovered. (middle) The automatic compensation system is turned on, the tone power quickly recovers its maximum value. (right) The compensation system is on, but the DGD change applied is too high to allow for stable compensation ; the system is pushed to a local maximum.

which is small enough to have negligible effect on the actual quality of the signal.

4.5.2 Simultaneous compensation of multiple orders of dispersion

A first set of parameters of great importance for telecommunication applications consists of GVD and higher orders of dispersion. We demonstrated automatic and simultaneous compensation of the second, third and fourth orders of dispersion (β_2 , β_3 and β_4) for a 1.28 Tb/s single channel OTDM signal. The signal was encoded using on-off keying (OOK) on a pulse train with 275 fs pulse width following the scheme presented in section 4.1. Feedback control from the RF-spectrum monitor to a FD-POP allowed the distortion level to be adjusted automatically. We evaluated the system on a test bed consisting of a transmission link emulator used to apply custom residual dispersion as well as temporal fluctuations. The automatic compensator kept the 1.28 Tbaud signal stable in spite of gradual or even abrupt changes in the link dispersion properties introduced by the emulator. This scheme is summarized in figure 4.17.

Our method extends a previous GVD only compensation scheme [76] by proving that various orders of dispersion can be optimized iteratively only by maximizing the tone power in the RF-spectrum as discussed in section 4.5. In this proof-of-concept experiment, we emulated GVD and higher-order dispersion fluctuations by applying phase filters corresponding to various combinations of 2^{nd} , 3^{rd} and 4^{th} order dispersion with a FD-POP. A second FD-POP controlled by the monitor was programmed to react in real time in order to



FIGURE 4.17: Automatic dispersion compensation scheme: the 1.28 Tbaud transmitter is followed by a link emulating both initial residual dispersion and dispersion fluctuations. The combination of a signal monitor with a dispersion compensator allows for automatic compensation of those fluctuations simultaneously for β_2 , β_3 and β_4 . The compensation scheme is based on maximizing the 1.28 THz tone power of the RF-spectrum of the signal (right insert).

act as the compensator.

Experiment

After passing through the dispersion emulation/compensation system, the signal was amplified and combined with a CW probe before propagating through a As_2S_3 dispersion-engineered waveguide [91]. Upon propagation through this highly nonlinear ($\gamma = 10^4 \text{ W}^{-1} \text{ km}^{-1}$) waveguide the cw-probe was spectrally broadened by cross-phase modulation (XPM), resulting in an optical spectrum proportional to the RF-spectrum of the input signal. We isolated the tone at 1.28 THz with a fibre grating bandpass filter followed with an ultra narrow square filter with 0.17 nm bandwidth in order to reject completely the signal and the CW component of the probe and measure the tone power with high contrast using a slow photodiode (PD). The power measured after filtering was typically -55 dBm for an optimized signal. The extremely low level of the tone, due to the weakness of the Kerr effect, did not constitute a big issue in this experiment, given that low bandwidth (kHz) photodetectors with sensitivity of the order of -80 dBm are common. Figure 4.18 depicts the setup used for automatic dispersion compensation, including the RF-spectrum analyzer.

As multiple orders of dispersion were compensated simultaneously, the cross effects between the various orders had to be taken into account. The actual β_2/β_3 tone power map presented an elliptical peak whose principal axes made an angle with respect to the dispersion coordinates. This made the global optimum harder to find for the optimization algorithm. We cancelled this effect by applying a linear coordinates transformation consisting of rotating and



FIGURE 4.18: The dispersion emulation system consists of a FD-POP and a short fibre link. It is followed by the compensation system controlled by a chip-based all-optical RF-spectrum analyzer.

stretching the reference plane in order to obtain a quasi symmetric peak as shown in figure 4.19. This improved significantly the speed and accuracy of the optimization algorithm. The transformation led to the definition of the coordinates $\{\xi_2, \xi_3\}$ in which the algorithm evolved, replacing $\{\beta_2, \beta_3\}$ (the effect is weak for β_4 , hence did not require such transformation). The coordinates transformation did not vary with environmental fluctuations and changes in the transmission link, leading to the conclusion that the effect is due to the internal configuration of the FD-POP. Figure 4.19 shows a color-map of the tone power as a function of ξ_2 and ξ_3 in the units of dispersion lengths (L_{D_2}, L_{D_3}) . The plot clearly reveals a point of maximum tone power corresponding to the optimized signal (note that the offset with respect to (0,0) is due to the fact that the dispersion compensation of the transmission link was not perfectly adjusted initially).



FIGURE 4.19: The colormap shows the tone power as a function of the dispersion state in the modified dispersion coordinates $\{\xi_2, \xi_3\}$ (see text). The three contour plots highlight the optimum zone of the plot at -1.5 dB, -3 dB and -4.5 dB.

The fixed step relaxation algorithm optimized the tone power with respect

to 2^{nd} , 3^{rd} and 4^{th} order dispersion. The signal was monitored continuously. Once the tone power dropped below a threshold value, the algorithm adjusted the dispersion on the second FD-POP to recover the optimum tone power. Figure 4.20 demonstrates the ability of our scheme to accurately track and compensate 2^{nd} , 3^{rd} and 4^{th} order dispersion simultaneously. We emulated a drift of the dispersion by discrete steps via phase profiles applied to the first FD-POP. It can be seen that our compensation algorithm recovered from the initial drop of the tone power when the dispersion was abruptly changed. After a short time, the tone power returned to its maximum value.



FIGURE 4.20: Automatic dispersion compensation time-line, depicting the temporal evolution of the 1.28 THz tone power of a signal impaired with sharp random dispersion changes. The tone power is quickly recovered thanks to the optimization-algorithm.

Although it would have been interesting to compare how the changes in tone power translate in terms of BER, this measurement could not be realized due to the difficulty of obtaining a BER through demultiplexing, as explained in section 4.1.3. Moreover, simultaneous measurement of BER and RF-spectrum would have required using two photonic chips in parallel, adding one level of complexity to the experiment.

4.5.3 Simultaneous compensation of GVD and DGD

PMD fluctuation is another important source of time-varying distortions in high baudrate links [102, 110, 111]. It usually goes along with dispersion impairments so that both effects have to be compensated at the same time. Automatic PMD compensation has been demonstrated at rate up to 160 Gb/s based on optical sampling monitoring technique and DGD compensators using a polarization controller and a birefringent fibre [112, 113]. Recently, the use of Nyquist pulses has been shown to reduce PMD sensitivity of OTDM transmission [114].



FIGURE 4.21: The transmitter is followed by a link emulating both the initial bias and the temporal drifts in the GVD and DGD. The automatic compensation system is inserted before the receiver to cancel those impairments using the signal received from the alloptical RF-spectrum analyzer (the signal monitor). The figure illustrates how the link distorts the data pulses and this is detected by the signal monitor which activates the DGD and GVD compensators. GVD causes pulse broadening and DGD pulse splitting, and these both impair the RF-spectrum by decreasing the 640 GHz tone (blue trace).

In the context of coherent transmission formats, another common dispersion and differential group delay (DGD) compensation technique uses postprocessing of the received signal in the electrical domain by digital signal processing (DSP) [115, 116]. This method has gained substantial interest recently; however, the statement that DSP can mitigate DGD and chromatic-dispersion impairments is not straightforward in the context of OTDM signals. These signals must be demultiplexed to a lower bit rate before they can be converted into an electrical signal. Demultiplexing using sampling techniques [26, 75] would require extracting all of the OTDM channels to allow for DSP compensation of the signal. This would remove the ability to extract only one channel at a time out of the OTDM signal. Inserting an optical distortion compensator before the demultiplexer allows for clean detection of one OTDM channel with optimum pulse width without having to demultiplex all the channels in parallel.

In this section we demonstrate the first real-time impairment cancellation system for GVD and DGD for a 640 Gb/s single channel signal. Simultaneous compensation of the two independent parameters is done by feedback control of separate GVD and DGD compensators based on the same monitor used for dispersion compensation in section 4.5.2. In this experiment, the accessible bandwidth was limited to 640 Gb/s due to the too coarse resolution (360 fs) of the DGD emulator simulating the fluctuations of the link properties. Its use on a 1.28 Tb/s channel would have caused DGD jumps at the limit of the range of stable operation of the monitor.

Figure 4.21 summarizes the strategy used for automatic compensation of DGD and GVD. The signal quality was monitored with an all-optical RF-spectrum analyzer with terahertz bandwidth based on a chalcogenide waveguide chip [67].

Experiment

Figure 4.22 shows the entire compensation system. It comprises a transmitter generating the 640 Gb/s OTDM signal encoded with a Pseudo-Random Bit Sequence (PRBS) data. The optical signal was transmitted through a link made of a FD-POP dispersion emulator, an interferometric DGD emulator, a short fibre section (50 m SMF and 10 m DCF) and an EDFA. The automatic compensation system was inserted after the link in order to recover the signal before the receiver.



FIGURE 4.22: The main building blocks of the compensation experiment are successively: a 640Gb/s OTDM transmitter, a link emulating DGD and GVD fluctuations with a DGD emulator and a FD-POP, a DGD compensator, a GVD compensator (FD-POP) and a RF-spectrum analyzer used as broadband signal monitor. Feedback from the monitor to the compensators enables automatic compensation.

We already showed in the previous section that the convergence criterion was satisfied by GVD. Similar analysis was carried for DGD by mapping the evolution of the RF tone power as a function of the level of DGD applied. Figure 4.23 summarizes this information. The graphs on the left of the figure contain the RF spectra for increasing values of delay. On the right hand side, the isolated power of the 640 GHz tone is plotted against the delay. These are connected to the eye diagrams of the 640 Gb/s data measured with an optical sampling scope in order to visualize the effect of the impairment.



FIGURE 4.23: Evolution of the tone power as a function of DGD (left and right graphs) validates the choice of that parameter as signal quality indicator. Indeed an increase of the DGD impairment (as visualized on the series of eye diagrams) causes a drop of the tone power within a range of more than 1 ps around the optimum.

Unlike chromatic dispersion monitoring, the tone power measured does not evolve monotonically with the level of DGD applied as seen on the right plot of the figure. The value is maximum at zero impairment and decreases to a minimum for a DGD of half a bit length. It then increases again up to a local maximum for a DGD value of one bit length. The tone power monitored oscillates with a period of one bit length as a function of DGD. This effect translates the fact that the waveform is the most deteriorated as pulses on both polarizations do not coincide. At the opposite, when the delayed pulses are aligned in time, the resulting waveform is similar to a pulse sequence, but with a lower peak power on average over the data pattern. The monitor then exhibits a local maximum of tone power, although lower than the absolute maximum of an optimized signal.

Results

Experimental results of automatic compensation are given for DGD-only and simultaneous DGD and GVD fluctuations. The first case demonstrates the stability of the method over a large span of DGD exceeding the duty cycle of the signal, while the second one demonstrates that the concept can be extended to the conjugate effect of DGD and GVD, by controlling two separate compensation devices.

In figures 4.24 and 4.25, the bottom graphs show time traces of the 640 GHz tone power measured after each step. The upper graphs record the evolution of the parameters controlled. In the case of figure 4.24, only DGD fluctuates,



FIGURE 4.24: (middle) Time trace of the DGD emulated by the link (dashed) and the DGD applied by the compensator in the opposite direction (solid). (bottom) Evolution of the tone power. The vertical dashed lines show where the DGD changes have been applied. (top) A series of eye diagrams confirms the compensation of the DGD impairment on the signal.

similarly to the experiment reported in [76] for GVD. The opposite of the offset introduced by the emulator is shown by the dotted trace. The state of the compensator is plotted in blue. After an initial settling time of about 50 s, the DGD states of the compensator successfully tracked that of the emulator. The eye diagrams shown on top of the figure demonstrate that the compensation system was working properly with open "eyes" being recovered after the changes of the emulator setting. The reaction of the system to a degradation of the signal is relatively rapid and the DGD applied by the compensator immediately moves in the correct direction. The compensation trace appears as a piecewise stair function due to the fact that the optimization algorithm proceeds by trial and errors with a fixed step size.

An increase in the number of degrees of freedom to be successively optimized by the hill-climbing algorithm results in a slower overall response of the system. In the case of multiple parameter optimization, shown in figure 4.25 for DGD and GVD, the compensators take longer to stabilize to the correct values due to the increased complexity of the optimization process. The states of the DGD and GVD compensators do not immediately follow the emulators and sometimes oscillate before converging. In order to push the system to its limits, random state changes were applied between the times 250 s and 600 s, without waiting for the complete stabilization of the system after each change.



FIGURE 4.25: (top) Time trace of the DGD emulated by the link (dashed) and the DGD applied by the compensator in the opposite direction (solid). (middle) Evolution of the GVD emulated by the link (dashed) and the GVD applied by the compensator in the opposite direction (solid). (bottom) Evolution of the tone power. The vertical dashed lines show where the DGD and GVD changes have been applied.

As a consequence, the tone power did not have the time to recover its optimum value and the compensators did not match the emulators during this transient period. However, when letting the system recover after the time 550 s, the compensation device successfully brought the tone power back to its highest value and the states of the compensators cancelled the emulators.

4.6 Discussion

Recovery time: Currently, the recovery time when a specific value of dispersion or DGD is applied by the link emulator is relatively long (typically of the order of several tens of seconds for large perturbations). This is caused by the use of a relatively slow optimization algorithm, combined with the time required to update the phase pattern to the FD-POP (≈ 1 s), the adjustment time of the DGD compensator (≈ 10 s) and the fetch time from the powermeter (≈ 0.2 s). In our simple algorithm, the parameters are optimized sequentially, i.e. the algorithm will first try to optimize second-order dispersion for 4 steps

and then switch to try to optimize third-order dispersion, and so on. A more efficient algorithm could however reduce this time. With improved compensators and better feedback strategy, recovery times below one second should be feasible. In any case, the main origin of dispersion and PMD drifts, the temperature dependence of the properties of the link, is an inherently slow effect (of the order of hours or tens of minutes). Therefore, in a deployed system the changes would be less abrupt, making it much easier for the algorithm to track. Thus our experiments present a worst-case scenario.

Measurement noise: Reducing the noise induced in the RF-spectrum monitor by the DGD compensator would allow a reduction in the step size set in the optimization algorithm, smoothing the "stairs" in the evolution of the compensator states and improving the stability of the compensated signal. As discussed in section 4.4, the DGD compensator is a major source of noise due to inaccurate phase control leading to polarization fluctuations that affect the measurement of the RF-spectrum. Replacing the home made DGD compensator by a device similar to the emulator, but with better resolution, would improve the results a lot by avoiding the issue of phase sensitivity of the DGD value applied.

Divergence and re-locking: In case of large abrupt change in the link parameters, the regulation system can lose the optimum working point and diverge towards local maxima. In such situation, the transmission is interrupted and re-locking is required. The recovery process can be complex in the case of automatic control of multiple parameters and require measuring the tone power map over a wide range of all of the parameters. For DGD only compensation, the locked-in regime corresponds to a clear maximum of the measurement, so that a simple scan over the values of DGD allows finding back the DGD offset required from the compensator.

Accessible parameters: In the future, automatic higher-order PMD compensation could be implemented based on the same technique, by considering each PMD order as additional degrees of freedom in the optimization process. The DGD compensator should then be updated to allow compensation for higher-order PMD. Addition of a computer controlled polarization controller at the input of the existing DGD compensator would already allow for first- and second-order PMD control [106, 107]. PMD can also potentially be compensated using a dual polarization FD-POP.

4.7 Conclusion

This chapter has shown how large amounts of data can be transmitted with ultra dense time-multiplexed signals. Short optical pulses were used to interleave multiple data channels without overlap between the symbols up to 1.28 Tbaud. At the receiver, a demultiplexing operation must be done to extract the data of a single tributary.

Through the practical issues encountered, we have experienced how challenging this technology is. Data encoded onto sub-picosecond optical pulses are extremely sensitive to impairments such as dispersion, differential group delay, jitter and timing drifts. Fluctuations of the parameters of a link are unavoidable in real world implementations, in particular due to temperature drifts, vibrations or stress. Hence propagation of dense OTDM signals over long distances is ambitious.

This work attempted to address some of the difficulties through impairment compensation driven by closed-loop feedback action from a quality monitor. An all-optical RF-spectrum analyzer provided the information on the quality of the high bandwidth signal and drove a multivariate optimization algorithm to maximize it by controlling impairment compensators. The scheme has been demonstrated for simultaneous compensation of second, third and fourth order dispersion, as well as DGD. The solutions developed might be enough to make transmission of serial data practical over short- to mid-haul distances.

Without limiting our approach to the extreme baudrates demonstrated here, the automatic compensation methods remain valid at rather moderate rates such as 80 Gbaud or 160 Gbaud and may improve the reliability of the transmissions for a minimal added complexity. Combined with a WDM approach, lower symbol rate OTDM may be a good intermediate for applications requiring the advantages of a serial channel while keeping a reasonable sensitivity to impairments.

However, in the current state of the technology, we are reserved about whether the benefits of using terabaud OTDM instead of existing and wellestablished techniques are worth the substantial effort. In particular, the demultiplexing operation remains an important issue. 78

5 Nonlinear optical processing in communication

Transmission of information requires signal processing at multiple stages. Operations on the data itself, such as compression, integrity checking and encryption are usually implemented in the transmitter and receiver at the ends of the link. At a lower level, encoding [117] and multiplexing are also crucial functions. For mid and long distance communication, amplification and regeneration of the signals can be required to compensate for the attenuation and distortion of the logical levels incurred upon propagation. Routing of data to various addressees is another key function in networks.

Since the era of electrical transmissions, these operations have been performed by electronic components. The signals being already in the electrical domain, these techniques did not raise particular issues. The introduction of optical fibre communication technology made a breakthrough in terms of range and capacity of network segments. In present optical telecommunication networks, information processing is done by digital signal processing (DSP) logic after optoelectronic conversion. This provides extremely versatile and powerful solutions to most computational needs, but comes at the cost of increased latency, power consumption and limited bandwidth since analog signals are converted to the electrical domain and processed by algorithms at the bandwidth of digital circuits [118]. Moreover, the cost of such devices increases significantly with the baudrate. Multiplexed transmission such as WDM and OTDM also requires demultiplexing and separate processing of every channel by as many electronic devices. Integration of the multiplexer, demultiplexer, opto-electronic conversion and processing logic onto single chips [119, 120] promises to facilitate the handling of multiplexed channels and reduce the costs of the parallelization of the operations. However, latency, energy consumption and heat dissipation remain challenging even in these hybrid solutions [121].

All-optical signal processing is a technology that could remove the need for double optoelectronic conversion and allow for processing directly at the bandwidth of optical components. However, many of these techniques rely on nonlinear optics and remain challenging. Over the years, the photonics research community has developed a range of methods applicable to telecommunication purposes. In this chapter, we bring our contribution to this effort by introducing a novel all-optical processing scheme compatible with advanced modulation formats.

5.1 You can buy bandwidth, not latency

In most common cases, the performance of a telecommunication system is expressed in terms of bandwidth. The total capacity, although limited for a single channel, can in principle be increased as much as needed by adding parallel channels. For example by laying multiple optical fibres to multiply the bandwidth.

As seen in section 2.4, the travel time taken by information is however, mostly incompressible. Nevertheless delay due to the processing of data is a non-negligible part of the overall latency of current optical networks and could be reduced. Some optical schemes have the potential of operating functions with extremely low latency and could replace current electronic devices. Here we introduce novel low-latency all-optical systems that could substitute to DSP in some applications.

5.2 Energy concerns

In recent years, the energy cost of processing and transmitting information has become a non negligible share of the world's energy usage [122–124]. Estimates of the total power consumed vary according to the studies, but most of them report numbers of a few percent of the global electricity production [125].

Although optical technologies are often cited as the future of computing, this can only happen if they are as energy efficient as existing systems. Given the already high power requirements of IT networks, the world cannot afford an increase in the energy cost per bit of information processing and transport. The question of the potential energy efficiency of optical logic is a highly controversial debate, yet some fundamental limits have been emphasized in recent studies [16, 125].

In-depth comparison of state of the art DSP and all-optical processing technologies has come to the conclusion that optics is not well suited for mass computing and large scale circuits with numerous cascaded operations [125]. However, one can easily envision all-optical solutions for performing elementary operations that can be carried out using simple schemes [126].

The schemes introduced in this chapter are analog methods that perform operations without digitizing the signals. This approach is advantageous for power consumption as it only requires one or few processing nodes to execute the wanted function. This comes at the cost of very specific architectures that can only be used for the application that they were designed for. In contrast, common usual DSP based systems provide more functionality and flexibility of use, but require a large number of transistors that would not be possible on an optical device. Although comparison of an experimental setup with mature commercial circuits is difficult, we will show that our analog photonic approach should at least not perform worse than electronic DSP regarding power consumption.

5.3 Existing all-optical operations

Numerous all-optical analog and digital operations have been demonstrated. Other technologies are still at an experimental state, but promise great improvements in the future of optical communications [127, 128]. Nonlinear optical processing functions mostly rely on semiconductor optical amplifiers (SOA) [129] and nonlinear wave mixing in HNLF or integrated nonlinear waveguides [130]. In parallel, advances in digital signal processing have made electronic solutions more and more competitive. All-optical techniques will find a place in the commercial landscape only if they demonstrate an advantage over existing DSP devices.

In this section, we do a short and non exhaustive review of some of the optical signal processing functions developed for telecommunications. The purpose of this discussion is to put our work into context.

5.3.1 Regeneration

The degradation of the Optical Signal-to-Noise Ratio (OSNR) and nonlinear distortions are the major limitations for the reach and bandwidth of optical fibre links. These impairments build up along the path and cannot be reversed with simple linear methods. This sets a maximum transmission distance for a given bit rate.

OSNR reduction is mainly caused by amplified spontaneous emission (ASE) from Erbium doped fibre amplifiers (EDFA). Amplification is required to counterbalance fibre attenuation, with EDFA modules typically spaced by about 100 km. Cascaded amplification causes the total amount of noise to grow and thus a reduction in OSNR, which induces an increase in the bit-error rate

(BER). Nonlinear distortions modify the shape of the eye diagram of the data and cause crosstalk with other channels. These alterations can be complex functions of all signals transmitted in parallel through a single fibre.

Regeneration methods are classified by their three levels of action. 1-R regeneration is reamplification of the signal without modification of its shape. 2-R regeneration reamplifies and reshapes the signal to compensate for degradation of the logical levels. 3-R regeneration reamplifies, reshapes and retimes the optical pulses to remove timing jitter.

One method employed for restoring degraded data consists of applying a decision threshold to the power levels of the signal. Thresholding restores the discrete states of the data and removes noise and irregularities. Timing can also be corrected by matching the pulses to a clock. Electronic thresholders convert the optical signal into the electrical domain, apply a decision on the value of each symbol, then re-encode the regenerated data onto a new optical carrier [131].

Amplitude regeneration

Numerous all-optical schemes have been demonstrated to avoid opto-electronic conversion. One well-known 2-R system, introduced by P. Mamyshev, consists of using SPM in a Kerr medium to cause spectral broadening of the signal [132]. Bandpass filtering of the resulting spectrum with an offset with respect to the initial carrier frequency ensures that small input powers are blocked, which restores the "0" logical level. For increasing input power, the intensity filtered rises quickly once spectral broadening has reached the frequency offset, then plateaus. This enables some restoration of the "1" logical level.

Phase regeneration

With the development of coherent formats encoding information onto the phase of light, the presence of phase noise has become a new challenge. Regeneration of the phase for coherent formats has been demonstrated by phase sensitive amplification in parametric amplifiers [133]. Different schemes can be used, for example parametric amplification by FWM in a Kerr medium [134]. One possible architecture uses two pump signals to transfer energy to amplify a third wave. The gain transfer function of the amplifier presents a maximum when pumps and signal are in phase, which amplifies the signal components whose phases have not been perturbed and attenuate the others.

5.3.2 Nonlinearity compensation

Section 2.3.4 reviewed nonlinear limits to optical fibre capacity and underlined the importance of detrimental nonlinear effects in mid and long haul networks. However, it is possible to undo nonlinear impairments, at least partially. The approach differs from a 3-R regenerator by the fact that the signal is not regenerated by thresholding of a distorted signal, but by applying the inverse nonlinear transformation that impaired it.

One recent method proposed for "beating the nonlinear Shannon limit" relies on optical phase conjugation (OPC). The scheme involves conversion of the optical field into its complex conjugate at the middle of the fibre link to cancel the distortions induced by both halves [32]. Such operation can be made by FWM process of the data signals with a pump at another wavelength. Another scheme uses twin conjugated waves propagated in parallel [47]. This last scheme claims beating the maximum capacity imposed by the nonlinear Shannon limit.

5.3.3 Routing

Routing of information consists of recognizing the addressee of a data packet and directing the corresponding payload to the correct destination. Static routing of data flow is presently operated in flexible networks by wavelength selective switches (WSS) able to direct different wavelength channels to multiple output fibres. However, the slow reconfiguration of these devices does not allow for real time switching between channels.

Basic schemes for all-optical header detection and routing have been proposed to all-optically detect a header pattern identifying the address of destination of a data packet. SOA Mach-Zehnder interferometers are a popular platform for such systems [135, 136].

5.3.4 Logic operations

In the quest for an optical computer able to perform high level functions in the manner of programmable digital processors, researchers have been aiming to create building blocks for logical operations. In spite of the difficulty to harness optical nonlinearities, a multitude of elementary circuits have been demonstrated. Starting from basic Boolean logic gates such as AND, OR, NOT or XOR [127], the degree of complexity of the operations available in the "photonic toolbox" has increased with the realization of full adders and slightly more advanced functions [33, 137].

This list of functions, although far from being exhaustive, gives an idea of the effort made by the photonics research community to develop new schemes for all-optical processing. One interesting conclusion is the fact that all the operations demonstrated are relatively basic in terms of logical complexity. This chapter introduces new methods to complete the "photonic toolbox". We will now introduce two key building blocks involved in our experiments: a multipath interferometric circuit and an optical dot-product.

5.4 Multipath interferometric circuit

An important function used in our experiments consists of splitting, delaying and recombining optical signals. Such multipath interferometric circuit function (MIC) is known in signal processing as delay tap filter. A diagram of a MIC involving three copies of an input signal is shown in figure 5.1.



FIGURE 5.1: Multipath interferometric circuit. Left: power and phase transfer functions. Blue, green and red traces correspond to different MIC configurations. Right: diagram of the equivalent MIC.

The effect of a linear MIC is fully represented by its spectral phase and amplitude transfer function, which can be realized in different ways. One method consists in reproducing the circuit in figure 5.1 (right) using fibre or free space optics components or photonic integrated circuits. Another, more flexible solution, is to configure a Fourier domain programmable optical processor (FD-POP) to apply the corresponding spectral transfer functions. In the experiments reported below, we chose the latter technique to allow convenient switching between different structures.

In the following, we consider a MIC where the incoming field $E_S(t) = A_S(t)e^{i\phi_S(t)}$ is equally split into N parallel arms, each delayed by an integer number of symbol lengths $\Delta L_j = (j-1)\Delta L_{\text{symbol}}$ (j indexes the arms). Before recombination of the arms, each path is attenuated by β_j and offset in phase by ϕ_j according to the desired MIC configuration. The field after propagation attenuation delay phase offset

in arm j is $E_j(t) = \frac{1}{\sqrt{N}} A_S(t) e^{i\phi_S(t)} \sqrt{\beta_j} e^{ik\Delta L_j} e^{i\phi_j}$. At the output of the device the delayed complex fields are passively combined resulting in a

total field:

$$E_{\rm tot}(t) = \sum_{j=1}^{N} E_j(t) = \frac{A_S(t)e^{i\phi_S(t)}}{\sqrt{N}} \sum_{j=1}^{N} \sqrt{\beta_j} e^{i(k\Delta L_j + \phi_j)}$$
(5.1)

with $0 \leq \beta_j \leq 1$. Let us define $\nu_j = \frac{c}{\Delta L_j}$ being the frequency offset associated to the delay j such that $k\Delta L_j = 2\pi \frac{\nu}{\nu_j}$. The amplitude transfer function is written

$$\mathcal{H} = \frac{|E_{\text{tot}}(t)|^2}{|E_S(t)|^2} = \frac{1}{N} \left| \sum_{j=1}^N \sqrt{\beta_j} e^{i\left(2\pi \cdot \frac{\nu}{\nu_j} + \phi_j\right)} \right|^2$$
$$= \frac{1}{N} \left(\left(\sum_{j=1}^N \sqrt{\beta_j} \cos\left(2\pi \cdot \frac{\nu}{\nu_j} + \phi_j\right) \right)^2 + \left(\sum_{j=1}^N \sqrt{\beta_j} \sin\left(2\pi \cdot \frac{\nu}{\nu_j} + \phi_j\right) \right)^2 \right)$$
(5.2)

and the phase transfer function is

$$\phi = angle(E_{tot}(t)) = atan\left(\frac{Im(E_{tot}(t))}{Re(E_{tot}(t))}\right)$$
$$= atan\left(\frac{\sum_{j=1}^{N}\sqrt{\beta_j} \sin\left(2\pi . \frac{\nu}{\nu_j} + \phi_j\right)}{\sum_{j=1}^{N}\sqrt{\beta_j} \cos\left(2\pi . \frac{\nu}{\nu_j} + \phi_j\right)}\right)$$
(5.3)

Current FD-POP technology allows for custom control of the spectral phase and amplitude with a resolution of 10 GHz and can implement the transfer functions 5.2 and 5.3 for a MIC involving delays up to about 50 ps. At 40 Gbaud, the method can create MIC filters involving up to 3 successive symbols. For an OTDM signal at 640 Gbaud, the system allows to handle packets of up to 33 bits.

5.5 Optical dot-product

An original contribution of our work to the expansion of the set of all-optical functions is the first realization of a vector operation between high bandwidth WDM optical fields. In our approach, we leverage the parallelism available to optics by using WDM signals to operate multiple elementary arithmetic operations at the same time. We demonstrate the versatility of the technique in sections 5.6 and 5.7 by implementing a linear crosstalk compensator for two data channels at 40 Gb/s and a format transparent hash-key error detection system.

Advanced modulation formats producing multilevel phase-and-amplitude encoded signals are now widely used in optical communication systems [43].
Presently, very few mathematical operations can be carried out all-optically on such signals. Their processing relies on coherent detection and DSP. For example, the fast Fourier transform (FFT), turbo codes and pattern recognition are essential functions in most optical information networks for encoding, error correction, impairments mitigation and routing [138, 139].

Here we introduce a novel all-optical scheme for calculating the dot-product between optical fields based on a combination of linear and nonlinear optics by Fourier-domain optical processing and four-wave mixing (FWM). All-optical and electro-optic matrix operations have been demonstrated in the past using free space linear optics, spatial light modulators [30, 140] and nonlinear media [141]. The method demonstrated here is applicable to any encoding format and potentially enables operation at ultra high baudrate thanks to the fast response of the Kerr nonlinearity. We believe that such high speed optical arithmetic operation has potential applications in the field of all-optical signal processing as it corresponds to the calculation of a complex linear form in linear algebra. In particular, summation of frequency components with harmonic coefficients can be the base of an optical FFT. In a broader context, an alloptical dot-product implementation is a valuable addition to ongoing efforts for developing an optical signal processing toolbox [127, 128]. Such a toolbox is desirable in particular to enable ultra-high bandwidth processing systems with low latency [130, 142].

We demonstrate an all-optical field dot-product at 40 Gbaud with up to three intensity modulated signals. The error-free operation of the scheme is demonstrated for a particular case with two signals. Moreover, the accuracy of the dot-product operation is quantified in section 5.6 by configuring it to remove a signal component introduced intentionally and measuring the biterror-rate improvement. Following the comment at the end of the last section, the current FD-POP implementation at 40 Gbaud is limited to dot-products between 3 signals. However, other types of MIC or the use of higher baud rates can overcome this restriction.

5.5.1 Principle

The aim of this work is to calculate a dot-product between a time dependent vector of values $\mathbf{E}(t)$ and a static vector of complex coefficients **a**. Mathematically, this operation is a linear form in the complex space resulting in a scalar $E_{I_{tot}}(t) = \sum_j a_j E_j(t)$ where $E_j(t)$ and a_j are the components of the two vectors. In optical terms, this would correspond to a weighted sum between orthogonal light fields $E_j(t)$, for example signals encoded on separate WDM channels.

The method involves a combination of linear and nonlinear optical processing. A FD-POP first applies attenuation and phase values corresponding to the modulus and argument of the coefficients a_j to the fields. Mixing of the modified fields with pump fields in a Kerr medium frequency-converts the WDM channels to a unique idler frequency where they add or subtract coherently. We show that this scheme operates a dot-product operation of the analog values contained in up to three complex optical fields of phasors E_{S1} , E_{S2} and E_{S3} with a vector of complex coefficients (a_1, a_2, a_3) normalized so that $|a_j| \leq 1 \forall j = 1, 2, 3$. The following equations are written for the case of three signals, but can be extended to any number of channels. With the coefficients expressed with their modules $|a_j|$ and arguments ϕ_j , this corresponds to the phasor in equation (5.5).

$$E_{I_{tot}} = a_1 E_{S1} + a_2 E_{S2} + a_3 E_{S3} \tag{5.4}$$

$$= |a_1| e^{i\phi_1} E_{S1} + |a_2| e^{i\phi_2} E_{S2} + |a_3| e^{i\phi_3} E_{S3}$$
(5.5)

The working principle of the device is shown in figure 5.2. The WDM signals to be processed are encoded onto a flat frequency comb with a frequency separation $\Delta \nu$. The dot-product device sets them with custom attenuation levels and relative phases before combining them with pump waves that have half their spectral separation $\frac{\Delta \nu}{2}$. Subsequent nonlinear mixing inside a Kerr medium generates frequency converted idlers via degenerate FWM. Due to the choice of the spectral separation of the pumps and signals, the idlers fall on the same frequency and add or subtract coherently according to their phase relations.



FIGURE 5.2: Principle of the all-optical dot-product. Relative phase and magnitude coefficients are applied to three WDM signals before propagating through a Kerr nonlinear medium for generation of idlers at a same frequency. Signals with opposite phases (represented with opposite arrows), lead to idlers cancelling each other.

The incoming WDM signals $\mathbf{E}_{S1}(t)$, $\mathbf{E}_{S2}(t)$ and $\mathbf{E}_{S3}(t)$ are phase locked due to the fact that they originate from the same frequency comb source. The pump signals $\mathbf{E}_{P1}(t)$, $\mathbf{E}_{P2}(t)$ and $\mathbf{E}_{P3}(t)$ are also phase locked, however they do not need to be locked to the signals. Their electric field is expressed in equations (5.6) to (5.11).

$$\mathbf{E}_{S1}(t) = E_{S1} e^{i2\pi\nu_{S1}t} \tag{5.6}$$

$$\mathbf{E}_{S2}(t) = E_{S2} e^{i2\pi(\nu_{S1} + \Delta\nu)t}$$
(5.7)

$$\mathbf{E}_{S3}(t) = E_{S3} e^{i2\pi(\nu_{S1} + 2\Delta\nu)t}$$
(5.8)

$$\mathbf{E}_{P1}(t) = E_{P1} e^{i2\pi\nu_{P1}t} \tag{5.9}$$

$$\mathbf{E}_{P2}(t) = E_{P2} e^{i2\pi(\nu_{P1} + \frac{\Delta\nu}{2})t}$$
(5.10)

$$\mathbf{E}_{P3}(t) = E_{P3}e^{i2\pi(\nu_{P1}+\Delta\nu)t}$$
(5.11)

In these expressions, E_{S1} , E_{S2} , E_{S3} , E_{P1} , E_{P2} and E_{P3} are the phasors of the fields. We consider that their amplitudes are A_{S1} , A_{S2} , A_{S3} , $A_{P1} = A_{P2} = A_{P3} = A_P$ respectively and their wave vectors k_{S1} , k_{S2} , k_{S3} , k_{P1} , k_{P2} and k_{P3} . The first order FWM idlers generated from these waves have phasors $E_{I1} = A_{I1}e^{i\phi_{I1}}$, $E_{I2} = A_{I2}e^{i\phi_{I2}}$ and $E_{I3} = A_{I3}e^{i\phi_{I3}}$. The wave vectors of the generated idlers are determined by the phase matching condition of the FWM processes [57]:

$$k_{I1} = -k_{S1} + 2k_{P1} \tag{5.12}$$

$$k_{I2} = -k_{S2} + 2k_{P2} \tag{5.13}$$

$$k_{I3} = -k_{S3} + 2k_{P3} \tag{5.14}$$

From these relations and considering that signals are phase locked and pumps are phase locked, the explicit phase relations of the idlers can be written [143]:

$$\phi_{I1} = -\phi_{S1} + 2\phi_{P1} = -\phi_{S1} + 2\phi_P \tag{5.15}$$

$$\phi_{I2} = -\phi_{S2} + 2\phi_{P2} = -\phi_{S2} + 2\phi_P \tag{5.16}$$

$$\phi_{I2} = -\phi_{S3} + 2\phi_{P3} = -\phi_{S3} + 2\phi_P \tag{5.17}$$

where the phases of the signals are ϕ_{S1} , ϕ_{S2} and ϕ_{S3} and the pump phase is set so that $\phi_{P1} = \phi_{P2} = \phi_{P3} = \phi_P$. We consider the phase of the incoming phase-locked signals as a reference and shift their relative phases respectively by $\phi_{Sj} \rightarrow \phi_{Sj} + \phi_j \quad \forall j = 1, 2, 3$ in order to implement the argument of the coefficients in the dot-product. In the particular case where (a_1, a_2, a_3) are real, these phases are 0 (addition) or π (subtraction). The coherent idlers add or subtract accordingly due to interference. In figure 5.2, we consider real coefficients (a_1, a_2, a_3) . Signals with opposite phases are symbolized by arrows pointing in opposite directions. Note that the principle remains valid when the vectors **a** and **E**(t) are complex valued.

In the undepleted pump approximation of degenerate FWM, the amplitude of the idler is proportional to the amplitude of the signal. As a consequence, the relative amplitudes of the idler waves can be controlled by power attenuation coefficients $(|a_1|^2, |a_2|^2, |a_3|^2)$, which are applied to the signal waves. Expressed in terms of field amplitude, this results in equations (5.18) to (5.20).

$$A_{I1} \propto |a_1| A_{S1} A_{P1}^2 \tag{5.18}$$

$$A_{I2} \propto |a_2| A_{S2} A_{P2}^2 \tag{5.19}$$

$$A_{I3} \propto |a_3| A_{S3} A_{P3}^2 \tag{5.20}$$

The relations above allow writing the phasor of the total generated idler $E_{I_{tot}}$ as a function of the characteristics of the incoming fields and the applied attenuation and phase shifts.

$$E_{I_{tot}} = A_{I1}e^{i\phi_{I1}} + A_{I2}e^{i\phi_{I2}} + A_{I3}e^{i\phi_{I3}}$$

$$\propto \left(|a_1| A_{S1}e^{-i(\phi_{S1}+\phi_1)} + |a_2| A_{S2}e^{-i(\phi_{S2}+\phi_2)} + |a_3| A_{S3}e^{-i(\phi_{S3}+\phi_3)} \right) \underbrace{e_{i,2\phi_P}^{i,2\phi_P}}_{\text{absolute}}$$

$$= a_1 E_{S1} + a_2 E_{S2} + a_3 E_{S3}$$

The latter expression matches the result expected in equation (5.4). Note that ϕ_P results in an absolute phase offset of the idler, which is not relevant to communication systems using coherent modulation formats. Hence, no phase relation is required between signals and pumps. Although other first-order cross products are generated from the pumps and signals, these appear at other frequencies than the idler of interest and can be removed by bandpass filtering. The method can be generalized to a dot-product between more than three channels by using more wavelength channels and pump waves.

In practice, the attenuation $|a_j|^2$ and phase shift $|\phi_j|$ are applied via attenuation and phase control using a FD-POP device schematically represented in the inset of figure 5.2.

5.5.2 Experimental setup with 3 signals

Figure 5.3 describes the experimental implementation of the scheme. A pulse train from a 40 GHz modelocked (ML) fibre laser centred at 1550 nm with 2.2 ps pulse width was spectrally broadened by SPM in two successive HNLF broadening stages. The resulting flat frequency comb was split into a pump pulse train and a data signal modulated, for example with a 64 or $2^{31} - 1$ (depending on the experimental feasibility) pseudo random bit sequence (PRBS) via a Mach-Zehnder modulator (M-Z).

Three independent data channels were derived from this signal by spectral slicing and generation of a one bit-delay between each of them with a Finisar Waveshaper FD-POP [144]. The device allows for delays up to 60 ps, with



FIGURE 5.3: Experimental diagram. A comb source is split into a pump pulse train and a data signal encoded using a Mach-Zehnder modulator. Spectral carving and selective delaying of the data signal creates three distinct bit streams. After following different path, pumps and signals are recombined and controlled in phase and intensity by a Fourier domain programmable optical processor (FD-POP) and generate an idler in a highly nonlinear fibre.

increasing insertion loss for long delays. The two channels were separated with the shortest possible delay (1 bit) in order to reduce OSNR impairments.

The dot-product device was based on a second FD-POP controlling the phase and magnitude of the signals involved in the operation, an optical amplifier and a 30 m section of HNLF. The square root of the attenuation coefficients to the channels corresponded to the absolute value of the coefficients in equation 5.21. The relative phase between the signals controlled the argument of the coefficients, or their sign in the case of real values. The idler wavelength was then extracted using a narrow (70 GHz) bandpass filter and measured via a sampling oscilloscope with 65 GHz bandwidth. Note that only the intensity of the output signal was observed, given that an intensity detector was used.

The optical spectrum after the HNLF is shown in figure 5.4. The dashed and dotted traces correspond to configurations with only S_1 and S_2 or S_2 and S_3 respectively that were used for calibration of the relative phases. When the pair of signals had the same bit pattern and were set with opposite phases, a dip appeared in the spectrum of the idler as a signature of the destructive interference. The power contrast due to interference between identical signals in the cases of addition and subtraction exceeded 20 dB. The solid trace features the dot-product operating with all three signals with a given set of phase and amplitude coefficients.



FIGURE 5.4: Optical spectrum after the HNLF. (dashed trace) Signal 1 and Signal 2 are equal, $\phi_{S1} = \phi_{S2} + \pi$; (dotted trace) Signal 2 and Signal 3 are equal, $\phi_{S2} = \phi_{S3} + \pi$; (solid trace) dot-product with three signals involved. (Right) Spectrum of the corresponding idlers superimposed and filtering setup.

5.5.3 Qualitative results of dot-product between 3 signals

Figure 5.5 presents the time trace of the idler intensity measured with the sampling scope. The initial data stream S was a 64 bit on-off keying (OOK) pattern. S_1 , S_2 and S_3 were obtained by delaying S by -1, 0 and 1 bit-delay (25 ps) respectively. The results are given for coefficients $(a_1, a_2, a_3) = (1, -0.4, 0)$, (0.3, -1, 0.7), (0.3, -1, 0.3) and (0.5, -1, -0.5) from left to right and top to bottom on the figure.

The theoretically expected envelope (solid blue trace) of the output signal is overlayed with the data measured. The agreement of the traces gives a good confirmation that the system is working as a dot-product of the signals.

5.6 Linear distortion compensation

In order to provide a quantitative evaluation of the dot-product function, we designed an experimental configuration where the expected output sequence



FIGURE 5.5: Sampling scope traces of dot-product output signal. The solid blue traces are simulations of the expected intensity levels.

was a known binary pattern, to allow for direct bit-error rate (BER) measurement. Two initial 40 Gb/s OOK data streams were voluntarily distorted according to a linear transformation represented by a matrix **X**. Basic algebra shows that this distortion can be undone by multiplication with the corresponding inverse matrix \mathbf{X}^{-1} . The transformation was reverted for one channel via the all-optical dot-product device. After cancellation of the impairment, the resulting recovered channel was analyzed through BER and eye diagram measurement.

The applied linear distortion can be compared with a simplified version of heterodyne linear crosstalk which is a recurrent issue in Wavelength Division Multiplexed (WDM) optical networks [145, 146]. In real systems, this effect is mostly due to optical add-drop multiplexer (OADM) and demultiplexer imperfections [147], where a data channel leaks into an adjacent channel due to imperfect filtering or spectral overlap in dense WDM networks [148]. However, at least in its present form, our system is not able to compensate for realistic crosstalk with non trivial spectral distribution and involving signals that are not phase locked. We take the analogy with heterodyne crosstalk to validate the system quantitatively with a concrete case, without claiming for a practical linear crosstalk compensation solution.

5.6.1 Principle

The principle of the experiment (figure 5.6) consists of creating of a crosstalk impairment between two data channels, followed by the dot-product configured to recover one data channel. Given that the crosstalk is linear, it can be reverted by a linear operation of the impaired channels. For each separate channel, this operation consists in a dot-product of the impaired signals with a vector of static coefficients.



FIGURE 5.6: Two WDM signals with controlled amount of distortion are generated. The dot-product device is configured to revert the distortion so as to restore the initial binary signal.

Distortion generation: The linear transformation of the two signals by multiplication with the matrix \mathbf{X} was applied inside the first FD-POP. The operation can be represented mathematically by a matrix product $\mathbf{D} = \mathbf{XS}$ where \mathbf{S} is an array of the signals before modification (for example optical fields), \mathbf{D} represents the distorted signals, and \mathbf{X} is the transformation matrix. Distortion can therefore be cancelled by applying the inverse transfer matrix \mathbf{X}^{-1} to the degraded signals $\mathbf{S} = \mathbf{X}^{-1}\mathbf{D}$, supposing that the adjacent channels are known. Here the matrices are normalized to verify conservation of energy and expressed with their coefficients. The expression of \mathbf{X}^{-1} is given for the particular case of 2-by-2 matrices with the coefficient $\alpha = \frac{1}{ad-bc}$.

$$\mathbf{X} = \left(\begin{array}{cc} a & b \\ c & d \end{array}\right) \tag{5.22}$$

$$\mathbf{X}^{-1} = \alpha \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$
(5.23)

The matrix \mathbf{X} is applied by an interferometric technique inside the FD-POP. The method emulates the effect without applying an actual power transfer between the channels thanks to the manner used to generate the incoming signals by delaying a single initial bit-pattern. Both channels are identical replicas delayed by one bit-length ΔL . As a consequence, the data of one channel can be obtained from the other one by delaying it by one bit-length in the opposite direction. A linear combination of the data of channel 1 and channel 2 can then be created from a single channel by adding a delayed fraction of this channel to a non-delayed one. The extra-diagonal coefficient b of the matrix \mathbf{X} corresponds to a leak from channel 2 into channel 1 and can be synthesized by adding a fraction b of each symbol of channel 1 into the next one in the manner of a MIC illustrated in figure 5.7 (right).



FIGURE 5.7: Attenuation and phase profiles implemented inside the FD-POP to distort channel 1 (left) and equivalent delay line interferometer (right).

Power and phase transfer functions of such MIC can be derived as detailed in section 5.4. These are plotted on the left of figure 5.7 for various values of b. The FD-POP is configured to apply these spectral transfer functions to the two channels, which is equivalent to a distortion by the matrix **X**.

Distortion compensation: The dot-product device is used on the distorted signal to undo the linear impairment. The attenuation level of the channels translate the absolute value of the matrix coefficients of \mathbf{X}^{-1} , while their relative phase (either 0 or π) set the signs. The FD-POP is configured so that both channels are π out of phase and the channel to be subtracted is attenuated by b^2 . The idler wavelength is then filtered using a bandpass filter and measured using a 40 Gb/s bit-error rate tester and a sampling oscilloscope.

5.6.2 Results

The optical spectrum after the HNLF is shown in figure 5.8. The blue and black traces correspond to signal 1 and signal 2 being identical and undistorted, with equal and opposite phases respectively. When the phases were opposite, a dip appeared on the blue trace at the idler wavelength with 20 dB of contrast, which was the signature of the destructive interference between the idlers. The green trace features both signals affected by a mutual crosstalk of 30 % and the compensator set to cancel this distortion. Signal 2 was attenuated by a factor $b^2 = 0.185 = -7.3dB$ and flipped by a π phase shift. Note that the spectral shape of the signals changed slightly when the distortion is applied, due to the FD-POP spectral shaping operation.



FIGURE 5.8: Optical spectrum after the HNLF. (black) Signal 1 and Signal 2 are equal, $\phi_{S1} = \phi_{S2}$; (blue) Signal 1 and Signal 2 are equal, $\phi_{S1} = \phi_{S2} + \pi$; (green) Signal 1 and Signal 2 are different and affected by a power transfer of 30%, $\phi_{S1} = \phi_{S2} + \pi$.

The method was assessed for coefficients of 0.2 and 0.3 with symmetric matrices

$$\mathbf{X_{0.2}} = \begin{pmatrix} 0.8 & 0.2\\ 0.2 & 0.8 \end{pmatrix}$$
(5.24)

$$\mathbf{X_{0.3}} = \left(\begin{array}{cc} 0.7 & 0.3\\ 0.3 & 0.7 \end{array}\right) \tag{5.25}$$

Figure 5.9 presents BER measurements for the back-to-back case (compensator bypassed) and with the compensator active. Curves are shown in the case of non distorted signals (blue), crosstalk with 20 percent leak (green) and 30 percent leak (red). A significant improvement of the power penalty was observed in both cases of crosstalk coefficients of 0.3 and 0.2. The wavelength conversion operation on an undistorted signal was causing a 1 dB power penalty compared to the back-to-back operation (B2B clean) at error free level.

5.6.3 Discussion

Our measurements confirm the ability of the system to operate a dot-product operation. Comparison of the BER curves for the clean converted idler and the



FIGURE 5.9: Bit-error rate testing of the compensator. Back-toback (B2B) and compensated curves are measured respectively at the wavelength of the signal and of the converted idler.

back-to-back cases shows that the operation introduces a minimum increase of the power penalty of about 1 dB at error free level. This added penalty is due to introduction of noise in the reamplification of the idler product. Moreover, the significant decrease in BER due to the action of the dot-product on distorted signals shows that the scheme reliably carries out the operation. The drop in signal quality of the recovered channel, compared to a clean channel, is mostly due to the optical signal-to-noise ratio (OSNR) degradation caused by the coherent subtraction. Indeed, subtraction of the idler signals leads to a total idler of lower power than the sum of their powers.

This OSNR decrease limits the range of coefficients practically accessible by our method. Indeed, choices of coefficients leading to subtraction of signals of similar magnitudes cause coherent cancellation of their respective idlers and the total idler power level is significantly reduced.

The coherent dot-product operation requires the incoming signals to be phase-locked. This condition is automatically satisfied provided that the signals originate from the same transmitter, such as a comb source encoded with monolithically integrated modulators. Pump channels must also be phaselocked, however no phase relation is required between signals and pumps. In our experiment pumps and signals both originated from the same source, although this is not required for the technique to function properly. Real world implementation would rely on a distinct pump source driven by a clock recovery module.

In the next section, we are using the dot-product to implement a signal comparator for detection of errors in a data transmission link.

5.7 All-optical error detection by hash code comparison

Detection and correction of mistaken data is a critical function in information systems. Therefore, numerous error management techniques have been developed, including forward error correction (FEC), checksums, cyclic redundancy checks and more generally hash functions [149]. The underlying principle of any hash method is the computation of a number as a signature of a data packet mapping the original data to a value in a smaller mathematical space [150]. Identical blocks of data yield the same hash code, while different ones should have different signatures, heralding an error. In a transmission system, failure can be detected with high probability by comparing the hash before and after propagation. With a higher degree of complexity, FEC algorithms are able to repair corrupted data without re-transmission, and are in use in most current telecommunication networks [151] allowing coding gains of over 10 dB [152– 154]. These integrity-check methods imply the transmission of some redundant information (i.e. the hash code) and usually result in increased latency and power consumption due to the required processing.

In the context of high bandwidth optical communications, these algorithms are part of the DSP at the transmitter and the receiver [138]. The processing operations introduce constraints in terms of bandwidth, energy efficiency and latency [155, 156]. An implementation of a hash function directly in the optical domain could significantly reduce these burdens. A recent report established the feasibility of a semiconductor optical amplifier (SOA) based hash encoder and decoder. However, the proposed scheme only works for on-off keying (OOK) signals of limited bandwidth [157, 158] and cannot be generalised to advanced modulation formats which are now used extensively in communication systems [70].

Here we introduce and demonstrate an all-optical, format transparent, hash code verified transmission link comprising of a low latency hash code generator and comparator. In this scheme, the hash code is generated using a FD-POP on the data signal, while the hash code comparator is implemented via coherent optical subtraction based on multiple FWM processes in a nonlinear medium. Data validation consists of comparing hash signals calculated before and after transmission. In contrast to most electronic methods, which incorporate the hash codes into the bitstream or packets, we transmit the codes on a separate wavelength channel thus avoiding the need for sophisticated bit-rate adjustment, check symbol insertion and segmentation schemes. Experiments are reported for a 40 Gb/s single channel binary phase shift keying (BPSK) signal and a 80 Gb/s single channel quadrature phase shift keying (QPSK) signal and the approach can be easily scaled to higher symbol rates and modulation formats without modification.

5.7.1 Principle

Hash based error detection codes require transmission of redundant information along with the data. In most electronic systems it is concatenated to the payload of each data packet. In our method however, the original data channel remains unchanged while the hash code is transmitted on a parallel wavelength channel as depicted on figure 5.10. At the receiver side, integrity of the data is checked by comparing the hash code from before to the one after transmission via all-optical signal subtraction.



FIGURE 5.10: Principle of the all-optical hash code verified link. Hash codes are generated inside the transmitter and communicated through the network along with the data in an adjacent channel. At the receiver side, hash codes are recalculated from the data channel and compared to the transmitted hash. On the diagram, red components indicate optical devices and green the electrical domain.

The functional diagram in figure 5.11 illustrates the successive transformations of the optical channels. The same data is encoded onto two phase locked WDM channels. One channel is kept as it is, while the other is used as a base for computing the hash signal (hash 1) in a first hash generator device. Both channels are then propagated through the link. At the receiver, the comparison is performed by calculating the hash (hash 2) from the received data channel using the same hash function and combining it with Hash 1 and two phase-locked pump waves. The four waves, occupying different wavelengths, are sent through a nonlinear medium to carry out the hash comparison. Upon nonlinear propagation, the two hash channels are translated to the same wavelength channel (idler) by degenerate FWM with the two pump waves. If the relative phase between the hash channels is set to π , the wavelength conversion will result in a coherent subtraction, i.e. if the hash codes are the same, then destructive interference cancels the optical power at the translated wavelength. A transmission error causing the hash codes to differ would unbalance the interference term and trigger a spike at the idler wavelength, producing an error signal which can be extracted by bandpass filtering.



FIGURE 5.11: Functional diagram. Arrows aligned with a same horizontal level represent signals at a same wavelength.

The hash code of an N-symbol packet is calculated by multiplying N successive symbols with custom weight and sign coefficients and coherently adding them. Such an operation can be realized using a multipath interferometric circuit (MIC) with tunable phase offset and attenuation, represented on the right part of figure 5.12. The sign of the coefficients is controlled by the relative phase of the different interferometric arms, while the weight can be set via attenuation. We use a FD-POP configured as a custom MIC as explained in section 5.4. Unlike previous methods [137, 157], this hash generation scheme is compatible with advanced modulation formats and allows for easily reconfiguring packet length and the hash function without any physical change to the setup. Wavelength selective delay, phase control and attenuation are implemented in the FD-POP [144, 159], as detailed in section 5.7.4. The resulting hash signal takes the form of a multilevel phase- and amplitude-coded channel at the same rate as the data. Note that, because the hash code is generated from N data symbols, its required bandwidth can be reduced to 1/N of this value by sampling, as further explained in section 5.7.6.

The comparator placed at the end of the transmission exploits our optical dot-product scheme to perform coherent subtraction of the hash codes. A dot-product with two channels can be configured as a complex subtracter by implementing a static vector with opposite coefficients $\mathbf{a} = (1 - 1)$.

$$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = H_1 - H_2$$
(5.26)

Given that the general case handles phase and amplitude and is format transparent, the subtracter is also. Hence, in contrast to previous systems that only worked for binary intensity coded signals, our hash comparator enables comparison of multi-level phase and amplitude encoding.

5.7.2 Implementation of the hash generator

The hash is calculated by coherent addition of N_S successive symbols with custom phase and amplitude offsets. This is performed in a MIC with calibrated phase shifts and attenuations on each arm illustrated on the right of figure 5.12. We implement the MIC with a FD-POP, as in section 5.4.



FIGURE 5.12: Hash code generation by Fourier domain optical processing. Left: power and phase transfer functions programmed in the FD-POP. The spectral profiles applied to the data signal reproduce the characteristics of a multipath interferometric circuit (MIC) coherently adding successive bits. Blue, green and red traces correspond to different MIC configurations. The other channel is left untouched through a bandpass filter transfer function. Right: equivalent MIC implemented in the FD-POP.

The pass channel of the wavelength selective MIC is transmitted through a flat top bandpass filter with constant phase. The hash channel results from application of a MIC function with each arm delayed by an integer number of symbol lengths, attenuated by β_j and offset in phase by ϕ_j according to the desired hash function equation.

Figure 5.12 shows examples of spectral transfer functions for various MIC configurations. The red trace corresponds to a 2-symbol packet with the hash definition $H_N = B_N + B_{N+1}e^{i\pi}$, where H_N and B_N are the optical field amplitudes of the hash and data at symbol N. The green and blue traces correspond to 3-symbol packets with hash definitions respectively $H_N = B_N + B_{N+1}e^{i\pi} + \frac{1}{3}B_{N+2}e^{i\pi}$ and $H_N = B_N + B_{N+1} + \frac{2}{3}B_{N+2}e^{i\pi}$. Other hash definitions can be obtained by choosing linear combinations of the B_N with other coefficients and phases.

The hash codes for 40 Gb/s BPSK and 80 Gb/s QPSK signals of 64 symbols are shown on figure 5.13. Changing the hash parameters leads to different hash sequences for the same initial signal. In this experiment, hash code and data signal are transmitted through a short fibre link of 50 m of SMF and 20 m of dispersion compensating fibre (DCF). Note that the time traces and eye diagrams represent only the intensity of the hash channel. Information is also



FIGURE 5.13: Experimental intensity plots of the hash codes for BPSK (first and second columns) and QPSK (third and fourth columns) signals. Information encoded in the phase of the hash signals is not represented. Time traces (top row) and eye diagrams (bottom) of the 64-bit patterns are measured with a sampling oscilloscope. Blues traces show simulation results for the corresponding bit patterns.

contained in its phase, which is not shown in the traces but is taken into account by the coherent signal comparator. Noise is mostly introduced in the signal generation by the phase modulator delivering the $\frac{\pi}{2}$ phase shift.

From the principle of hash used to check data integrity, different hash codes imply that the packet contains an error, although identical ones mean that the packet is likely to be correct, without proving it with certainty. The degree of reliability of packet validation depends on the hash algorithm and the number of bits involved in the hash code. Our proof-of-principle experimental setup performs hash calculation over up to 3 symbols (6 bits in the QPSK case), limited by the delaying capacity of the FD-POP. Real communication systems, however, usually perform hash operation on much larger packets in order to reduce the probability of an error remaining undetected, allow for using FEC mechanisms and reduce the number of verification operations [160]. Present telecommunication systems use packet lengths of 16 bits (IPv4 header checksum) [161] to 1992 bits for a Reed-Solomon (255,249) FEC code [162], or even longer packets. Increasing the number of bits considered in the hash could be done either by calculating the hash with an integrated MIC with larger delays or using advanced modulation formats such as QAM coding a higher number of bits per symbol.

5.7.3 Implementation of the hash comparator

The dot-product scheme introduced in section 5.5 can be configured as a comparator as depicted in figure 5.14. The two phase-locked hash channels to be compared are co-propagated with two pumps separated by half their spectral separation $\frac{\Delta\nu}{2}$. The pumps are set with equal phases $\phi_{P1} = \phi_{P2} = \phi_P$ while the hash signals are set with equal magnitude and opposite phases $\phi_{H2} = \phi_{H1} + \pi$ for coherent subtraction. Given that if the two initial copies of the signal come from the same broadband source, H_1 and H_2 are automatically phase locked. Following the same reasoning as previously, the idlers sum up as

$$E_{I_{tot}} = E_{I1} + E_{I2} = A_{I1}e^{i\phi_{I1}} + A_{I2}e^{i\phi_{I2}} \propto \left(A_{H1}e^{-i\phi_{H1}} + A_{H2}e^{-i\phi_{H2}}\right) \underbrace{e^{i2\phi_P}}_{\text{absolute}}$$

= $(A_{H1} - A_{H2})e^{-i\phi_{H1}}$ (5.27)



FIGURE 5.14: Principle of the all optical coherent signal comparator based on a dot-product scheme. The total idler product cancels when both hash are equal. Mismatch between hash codes causes an idler spike.

The subtraction in complex space implies that the total intensity at the idler frequency cancels only if $A_{H1} = A_{H2}$. Any mismatch in phase or amplitude between the hash signals translates into an idler spike signalling an error. Hence the extracted idler defines the error signal since it is zero as long as the hash codes are equal and is nonzero value when they differ.

5.7.4 Experiment and results

The experimental setup is shown in figure 5.15. The signal was generated from a mode-locked fibre laser that delivered a 40 GHz pulse train (2 ps FWHM) at 1550 nm. The spectrum was broadened inside a HNLF and filtered to obtain a quasi-flat comb spectrum over a bandwidth of 6 nm [58]. The pulse train was then encoded using two successive phase modulators driven by two independent pseudo-random bit sequences (PRBS) of either 64 bits (figures 5.13 and 5.17) or $2^{31} - 1$ bits (figure 5.19) and biased to deliver a π and $\frac{\pi}{2}$ phase shift in order to generate an 80 Gb/s QPSK signal [163]. The 40 Gb/s BPSK signal was generated by bypassing the second modulator.

The hash code H_1 was initially calculated at the transmitter with a first FD-POP. After a short link made of 50 m of SMF and 20 m of DCF, the



FIGURE 5.15: Experimental setup. A wideband flat frequency comb was QPSK-encoded with two successive phase modulators (PM). Hash codes were calculated before and after transmission inside a FD-POP. Combination with a pump pulse train and wavelength-selective phase control by the second FD-POP enabled coherent hash comparison inside a 30 m section of highly nonlinear fibre (HNLF). PM: phase modulator ; BPF: bandpass filter.

hash H_2 was re-calculated from the transmitted signal with a second FD-POP using the same phase and amplitude transfer function as described for the transmitter. The transmitted hash channel H_1 was left untouched. The relative phase between H_1 and H_2 was adjusted to π by the second FD-POP to result in coherent subtraction.



FIGURE 5.16: Experimental optical spectrum at the HNLF output. Insert: the total idler cancels out for identical hash and not if they differ. Solid: both hash signals equal; short dahses: one error every 512 symbols; long dashes: hash functions different for both signals.

The pump was coupled to a second input port of the FD-POP and shaped to obtain a dual pump configuration with half the spectral separation of the hash signals. The hash and pump were combined inside the FD-POP to form the input of the comparator. After amplification to a total optical power of 300 mW and out-of-band noise filtering, nonlinear mixing occurred in a 30 m span of HNLF. The signals both had powers 10 dB lower than the pumps.

Initial adjustment of the relative phases between signals was done by minimising the power of the idler with both hash channels equal. A dip was observed in the spectrum of the idler. A simple method to check that the hash comparator was working properly consisted in artificially causing a mismatch between the hash by generating it using different functions before and after the link. Different channels should result in a nonzero error signal for most symbols.

Figure 5.16 shows the output spectrum measured after the HNLF for three configurations: solid green: the two hash functions were identical, resulting in minimum idler output level due to coherent subtraction (H1 - H2); dashed black: the hash functions were programmed differently at the receiver and the emitter, resulting in unbalanced coherent subtraction and therefore output pulses indicating errors at every symbol; dotted blue: the hash functions were equal, but one bit every 512 bits presented an error that triggered an idler spike at the comparator output. The latter result with localized error was obtained with a variation of the setup designed to inject a single error in the signal. This modified experiment is explained in more details in the following subsection "Detection of a single error" and its corresponding figure 5.18.

Error detection was performed in the time domain, by narrow band (70 GHz) spectral filtering at the frequency of the first order FWM products (detailed in the inset) to extract the idler. After amplification, the hash verification signal was viewed on a sampling scope with 65 GHz bandwidth.

The amplitude of the error signal translates to the degree of mismatch between the hash codes and the degree of packet reliability within the limit of small errors. Indeed a small phase and amplitude offset applied to one symbol causes a proportional change in the hash function. Note however, that stronger perturbations or changes to multiple symbols have very small, but nonzero, probability of cancelling out in the hash calculation.

Figure 5.17 relates the comparator output to the spectrum of the extracted idler for an 80 Gb/s QPSK data. When both hash functions were generated using the same parameters before and after transmission, the mismatch signal after the comparator was constant zero, reporting no error (bottom left). If the hash signals were generated with different hash functions on both generators, the output was nonzero heralding transmission errors (bottom right) at every symbol.

Detection of a single error

Figure 5.18 shows a variation of the setup introduced in figure 5.15 in order to create localized errors in the data signal in the case of BPSK encoding with $2^{31} - 1$ PRBS data. The two replicate channels encoded with a first



FIGURE 5.17: Time traces of the error signal after bandpass filtering of the idler. Top: Idler channel filtered out. Bottom left: both hash signals equal. Bottom right: hash signals different. Hash code mismatch is reflected by a nonzero error signal after optical subtraction.

phase modulator (PM) were split into two arms. One passed through a second phase modulator driven by a separate pattern generator operating a π phase shift to one bit every 512 symbols. The other arm was delayed and adjusted in phase with a phase shifter (PS) to exactly match the path lengths of the interferometer. Both paths were recombined inside the first FD-POP operating the hash calculation by assigning the two wavelength channels to different ports of the device. This structure kept the two data channels identical, *except* for introducing a one symbol error every 512 symbols. The relative phase between the two paths was set manually via the phase shifter and phase locking was maintained by placing the interferometric structure inside an enclosure to isolate it from vibrations and thermal drift.

A single bit error yields an idler mismatch in the comparator for the symbols that have different hash codes. For a hash function calculated for data packets of N_S symbols, the error signal exhibits N_S spikes, due to the fact that all of the N_S hash codes involving the mistaken bit are altered. Figure 5.19 shows the comparator output for a single bit error when the data packets contained 3 bits ($N_S=3$). Similarly, a single bit error during transmission in a real system would change the hash code of the data channel, leading to a spike at the comparator output.

A phase change of π is a large perturbation of the signal and a real world transmission would already exhibit errors with alterations smaller than these.



FIGURE 5.18: Variation of the signal generator to create localized errors in a BPSK signal. The two replicate channels were split onto two arms, one being affected by a π phase shift of one bit period every 512 symbols. Both paths were recombined into the first FD-POP acting as a wavelength selective switch. PM: phase modulator ; PS: phase shifter.

As a consequence, smaller symbol changes should be detectable. In the previous results presented for unequal hash functions (see figure 5.17), various magnitudes of symbol mismatch can occur according to the data sequence and hash functions applied to it. In this case, small phase and amplitude changes are created and are still detected (although with a smaller magnitude of the error signal), as seen on the right time trace on figure 5.17. This shows that the hash comparator is also able to monitor small symbol alterations.

Tolerance to long distance transmission

In order to verify the validity of the technique for propagation through a longer path, we implemented the system on a 40 Gb/s OOK link made of 20.72 km of SMF and a matching spool of DCF (340 ps/nm) as presented on the left diagram of figure 5.20. Residual dispersion was further compensated at the end of the link using the FD-POP and applying an additional quadratic spectral phase in the same manner as described in section 4.3. The hash code channel was configured for 3-bit packets according to the equation $H_N = B_N - B_{N+1} + \frac{1}{3}B_{N+2}$.

The only significant issue related to the addition of the link was a slow scale timing drift between the transmitted signal and the pump (that follows another path) due to ambient temperature fluctuations. The change in path length caused by the environmental temperature variations desynchronized the pump and signals significantly within a period of about 10 s. This issue, however, would not show up in a real system where the pump pulse train would be generated based on a clock recovery module from the received signal. In such scheme, the pump is synchronized with the signal received at the end of the link. Importantly, the coherent comparator was not affected by the phase drift between hash and pump channels as mentioned in the principle section.

Figure 5.20 features the optical spectrum of the idlers and the corresponding eye diagrams for two cases. The solid green trace shows the destructive interference between idlers when transmitted and received hashes were equal

107



FIGURE 5.19: Time traces of the error signal after bandpass filtering of the idler in the case of a single error. Top: Idler channel filtered out. Bottom left: both hash signals equal (no error). Bottom right: a single error causes a spike in the error signal.

and no error occurred. The dashed black trace marks the increase in total idler power when the comparator was unbalanced by configuring both hashes with different settings, which was equivalent to errors occurring at every bit. The corresponding time trace of the error signal, measured with a sampling oscilloscope, still comprised symbols apparently without error (error signal at zero), which is due to the fact that the data packets were formed of only three bits. Hence there is still a high probability of having equal hash codes for different signals.

5.7.5 Integrated hash comparator with ultra low latency

The latency introduced by our system is primarily determined by the 150 ns propagation time through the 30 m of HNLF. This latency, although very low, is still not negligible. It can be almost entirely removed by unwinding the HNLF so that it is part of the transmission link; however this is impractical for real world systems due to the space required. In this section, we show that the hash comparison scheme remains valid when the portion of HNLF is replaced by a short highly nonlinear waveguide. The integration of the function on a chip is also a valuable improvement for reducing the footprint of the device. Note that the pigtail fibre length inside the FD-POP used in the setup is of a few meters that can be removed in a real world system.

The experimental setup, presented in figure 5.21, was similar to the one detailed previously. Instead of the 30 m of HNLF, nonlinear mixing of the two hash and two pump channels occurred in a 6.5 cm highly nonlinear ($\gamma =$



FIGURE 5.20: Hash code verified link of 20.72 km. Left: composition of the link. Right: Idler channel filtered out in case of both equal (green) and different (black) hash function definitions. The corresponding waveforms show the error signal in the time domain.



FIGURE 5.21: Experimental setup of the integrated hash comparison system. The shaded area represents the chip-based signal comparator.

 $10^4 W^{-1} km^{-1}$) dispersion engineered chalcogenide chip waveguide and generated idlers that were extracted by a 70 GHz bandpass filter.

Similarly to other experiments using the photonic chip, the major challenge of this result was the extremely low idler power due to the weakness of the Kerr effect. This translated into noise on the error signal after reamplification.

Figure 5.22 shows the output spectrum after the photonic chip. As done previously, two configurations are featured: the two hash signals were identical, resulting in coherent cancellation of the idlers (dashed trace); the two hash signals were different (generated with different hash functions before and after transmission), resulting in unbalanced coherent subtraction (solid trace).

Time traces of the error signal in the two cases are illustrated at the right of figure 5.22. When both hash codes were equal, the comparator output was constant zero, reporting no error. If the hash signals were different, the output was nonzero.



FIGURE 5.22: Output spectrum and extracted error signal from the hash comparator. Dashed: the two hash signals were identical. Solid: Hash generated with different functions, errors appeared at every bits.

5.7.6 Discussion

Extension of the method to higher symbol rates

The potential for high baudrate operation enabled by the ultra fast reaction time of the Kerr effect exploited in the comparator offers a solution for error management at symbol rates inaccessible to electronics.

Our implementation of the hash generator is particularly well-suited for ultra high baudrate OTDM signals. The delay of the FD-POP is limited to about 50 ps [103], which at 40 Gbaud only allows hash calculation for packets of up to 3 symbols (note that this still corresponds to a bit-length of 6 in the case of QPSK). However, increasing the symbol rate to 640 Gbaud, for example, would enable processing of packets of up to 32 symbols. Also hash code based verification of an OTDM channel only requires one instance of the hash generation and comparison infrastructure, while WDM systems would need separate nonlinear hash comparators for all channels.

Latency

The latency introduced by the operation is governed by the propagation time through the devices added to the optical link. The time-of-flight in the 6.5 cm chip waveguide is approximately of 0.325 ns. The impact of the hash code generation can thus be approximated by the travel time through the FD-POP instrument, coming to about 1.5 ns delay per device. Hence the total added delay is of about 3.3 ns. Note that the latency of the hash generator could be significantly reduced by replacing the FD-POP by integrated multipath delay line interferometers.

This represents roughly two orders of magnitude of improvement compared to state of the art DSP-FEC processing. In this comparison, the latency of DSP is evaluated by the number of CPU cycles required. Optimistic data from state-of-the-art commercial DSP-FEC technology report latencies of 15.5 ns for encoding operation and 96.2 ns for decoding, summing up to 111.7 ns [164]. Finally note that the bandwidth efficiency of the technique can be improved by a factor N_S by sampling and filtering the hash channel in order to keep only one hash value per data packet made of N_S symbols.

We acknowledge the fact that DSP based error correction mechanisms provide much more sophisticated solutions. These introduce significant latency that is not straightforward to quantify, given the variety of different methods. Our scheme is to be considered as a compromise between DSP-FEC operation introducing high latency and a link without error detection.

Energy consumption

Evaluation of the energy per bit requirement of a deployed system based on our technique is not easy. Here we provide an approximate energy assessment. Given that the hash generation schemes can be implemented on passive circuits such as LCoS based FD-POP or integrated multipath delay line interferometers, we consider that they do not contribute to the energy consumption of the system. As a consequence, the power requirements are dominated by the signal comparator.

In principle, the power level needed for the data signals involved in degenerate FWM is relatively low and roughly comparable to the power required for optoelectronic detection by classical systems. Hence, we do not consider it in the energy assessment. The energy required to produce the pump used in our scheme (of the order of 20 dBm) needs to be taken into account. It could for example be delivered by a semiconductor modelocked laser amplified by a low power EDFA to reach an average power of 20 dBm. Low power semiconductor lasers do not need more than 5 W of DC power supply. Commercial low power EDFA exist with power consumption under 3 W (for instance Alnair EFA-200C). Reamplification of the idler signal after filtering is made with a second EDFA.

The total estimated power requirement of our method would be on the order of 11 W, with no significant dependence on the baudrate or modulation format. Commercial DSP-FEC technology offers solutions requiring 2.5 W per channel at 10 Gb/s, not including opto-electronic conversion (for instance Vitesse VSC8494 Quad Channel Universal 10G). Hence the total power requirement for a data rate similar to our 40 Gb/s BPSK demonstrator would be 10 W or 20 W for 80 Gb/s QPSK. At these bit rates, our technique has a power requirement comparable to state-of-the-art electronics.

Hash channel sampling for improved bandwidth efficiency

In the experiment described in this paper, the hash channel has the same bandwidth as the data channel because a hash code is generated at every symbol period. However, it can be downsampled to $1/N_S$ of the original baudrate to keep only one hash code per data packet made of N_S symbols, which is enough for error detection. This could be done for example, by sampling the hash with an electro-absorption modulator [165] followed by a narrow-band filter or, for ultra high bandwidth OTDM signals, by nonlinear optical sampling [74, 166]. The resulting lower bandwidth hash channel would be less subject to distortions and have better spectral efficiency.

Effect of dispersion

Residual chromatic dispersion can cause timing misalignment of the symbols in the data and hash channels transmitted at different wavelengths. To counter this, the FD-POP provides a flexible way to compensate for group velocity dispersion as well as higher orders of dispersion, at the same time as they apply the hash phase masks [103]. This is the method used in our 20.72 km link transmission to remove the residual delay due to imprecise dispersion compensation by the DCF section.

Also, in chapter 4 we have demonstrated solutions for automatic compensation of fluctuations of multiple orders of dispersion using a FD-POP as tunable compensator [167]. In presence of strongly varying dispersion, the FD-POP devices already used for the hash function generation can be feedback-controlled to cancel dispersion in real time.

Phase relations

The comparator requires the received data and hash signals to be phase locked. This condition is automatically verified provided that the hash is generated from the signal itself and that both channels propagate along the same path of fibres and amplifiers. Although the pumps involved in the comparator must also be phase locked, no phase condition is required between signals and pumps.

However, accidental desynchronization of the two channels or introduction of phase noise in long amplified links may degrade the action of the comparator. A perturbation $\Delta \phi$ of the relative phase between hash and data channels would introduce a maximum power inaccuracy in the error signal of the order of $\frac{I_{\text{ADD}}}{2} (1 - \cos(\Delta \phi))$, where I_{ADD} is the maximum total idler power when the phase relation is set in the adder configuration ($\phi_{H1} = \phi_{H2}$). A dependence of the total idler power of this type has effectively been measured when sweeping the relative phase over a 2π range (figure 5.23). In the case of our demonstration of a 20.72 km link reported in paragraph 5.7.4, no effect of phase noise could be observed and the extinction ratio of the coherent subtracter was similar to the one measured for the back-to-back transmission.



FIGURE 5.23: Theoretical and experimental inaccuracy in the hash subtraction as functions of the phase perturbation between the two hash signals. The measurement was realized by feeding two equal signals in the coherent comparator and sweeping over their relative phase.

The effect of a longer link could not be studied with the present experimental setup due to the slow time drift discussed in section 5.7.4. The temperature fluctuations, mainly, caused the signal and pump channels to walk off over a time scale of about 10 s, making longer links even more difficult to realize. However, a real system would include a clock recovery device driving the pump to remove this issue.

Because pumps and signals must verify a precise relation of their respective frequencies, it could be thought that they would require phase locking, given that frequency is a derivative of phase. However, phase locking is only required between signals and between pumps separately. The set of signals and the set of pumps can be independent.

In other terms, locking of the relative frequencies (frequency separation between the signals and between the pumps) is needed and not the absolute frequencies. Frequencies can vary, as long as the relation between frequency separations is maintained. In the experiment, the repetition rate is used as a reference for the exact frequency separation. Signals and pumps have same repetition rate, which translates into spectral lines at the exact same interval for pumps and signals. These are used to quantify the frequency separation (e.g.: signals separated by 8×40 GHz \rightarrow pumps separated by 4×40 GHz). This method can be seen as a certain form of frequency locking (through the repetition rates), but is not a direct phase locking of the carriers themselves.

5.8 Conclusion

In this chapter, we have introduced some new functions expanding the "photonic toolbox" for telecommunication applications. The core active element of our new systems is an ultrafast vector operation between WDM channels combining nonlinear optics with linear interference effects.

We started by the implementation of flexible complex MIC using a FD-POP. This recent technique simplified significantly the testing of the new devices and was an integral part of the error detection system.

The novel concept of all-optical WDM dot-product has then been introduced and demonstrated as a building block for more advanced functions. The method operates a sum, in the complex field, of coherent optical signals weighted by controllable complex coefficients. The idea behind the technique is the coherent addition of the signals by wavelength shifting to the exact same frequency. The process is based on simultaneous multiple degenerate FWM effects between the signals and a set of pump waves having precisely half their spectral separation. A demonstration of dot-product used as a linear distortion compensator gave a quantitative evaluation of the function.

A complete all-optical error detection system with ultra low latency was then realized, combining two MIC for generation of redundant hash signals and a dot-product for comparison of the optical hash. The scheme is format transparent and can be used with high baudrate signals. Its photonic chip implementation offers latency performances about two orders of magnitude better than state-of-the-art DSP.

To conclude the chapter we compare these optical solutions to cutting-edge DSP technology. This constitutes a big debate inside the optics community, with diverging points of view amongst experts in the field. The usual arguments towards all-optical processing are power efficiency and low latency. However, the claim concerning power has been contested in the case of complex systems involving multiple parallel or cascaded operations. The high peak power required by nonlinear optical interactions is indeed a major flaw and makes the practical advantages of all-optical techniques questionable.

Such statements are mostly based on the principle that optical processing should mimic the architectures of classical digital methods, which typically require a large number of binary logic gates to implement a single operation. All-optical devices, however, are more likely to perform well in the context of analog processing. In the analog approach, the function is not broken down into elementary binary operations of a digitized version of the signals. Instead, it is carried out directly by a well-suited physical effect. Table 5.1 presents existing DSP and optical solutions to some key signal processing applications.

The potency of DSP resides in their flexibility of use and ability to reproduce a wide range of functions by programming their behaviour. Binary architectures also enable numerous cascading of operations without degradation of the signals. They benefit from over 60 years of evolution of transistorbased electronics and a very mature fabrication industry. Nonlinear optical technologies are, quite the opposite, still subject to many developments at the fundamental level. Presently, nonlinear optical processing has a notable advantage only in a few niche applications. However, the fast progress in hardware and subsystems promises more competitive optical solutions in the future. The evolution of integrated optics starts enabling compact structures with a large number of elements and even allows for hybrid chip devices comprising both optics and electronics on the same substrate [168, 169].

Function	Optical processing	DSP
Amplitude	Nonlinear thresholding	O/E conv. + electrical
regeneration		thresholding
Phase	PSA	Coherent detection +
regeneration		digital phase recovery
Nonlinearity	OPC	ADC + digital nonlinearity
compensation		compensation
Dispersion	DCF, grating, FD-POP	ADC + digital dispersion
compensation	(linear optics)	compensation
Format	All-optical format	Coherent detection +
conversion	conversion	re-encoding
Packet	Few solutions	Digital header processing
routing		
Binary logic	Nonlinear logic gates	Electronic logic
Multilevel	Few solutions	ADC + digital processing
logic		

TABLE 5.1: Electronic and all-optical solutions for some key processing functions.

The major strength of nonlinear optics is bandwidth. The ultrashort timescale of second- and third-order effects, combined with the frequency and polarization multiplexing abilities, largely outperform the capacity of electronics. Thanks to this, light is the candidate of choice for data transport. Processing of the signals without a double conversion to the electric domain has an obvious advantage in terms of reducing the complexity of the networks. Other qualities such as robustness and quasi inexistent interactions by radiation with the environment can also benefit some specialized sectors (defence, space and industrial environments with high radiation levels,...)

We consider that the current trend tends towards unification of DSP and optical processing into half-breed systems where the advantages of both technologies can be leveraged [170–172]. Rather than a confrontation of optics versus electronics, it appears reasonable to think that these can be best allies.

6 Optical neural networks

6.1 Neural network processors

The ranked #3 biggest supercomputer in the World, *Blue Gene*, computes over 17,000 peta-operations per second and draws 7.9 MW of power. A human brain consumes 25 W of power, performs real time pattern recognition, precisely controls 640 muscles and is capable of a level of abstraction that *Blue Gene* is a complete stranger to¹.

6.1.1 History and intuitive vision of neural networks

The concept of neural networks arose with a phenomenological study of the architecture of biological information processors: brains (see figure 6.1). Despite decades of in-depth analysis of cerebral tissues, scientists could not explain the complex operations of the human mind from the organization of neurons in terms of an algorithm. The working principle of a single neural cell, a neuron, is well understood [174]. In a simplified way, the head of the neuron receives signals from other neurons and fires an output towards other neurons if the total incoming excitation exceeds a threshold. However, no underlying "code" governing the connections could be inferred, so that the brain appears as a complex organization empirically built without a plan.

Basic mathematical models of neural interactions came to the conclusion that it is possible to process information without a step-by-step algorithm.

¹Blue Gene has been used by IBM as a simulation platform to run biological brain models. It has been able to "simulate" an entire cat's brain and 4,5% of a human brain [173].



FIGURE 6.1: Neural networks are adapted to a task by optimization, similar to brains that are the result of a long evolution.

Instead, a generic system consisting of a large number of interconnected elements with similar properties (cf. the biological neurons) was optimized recursively [175].

Biological neurons are specified cells that receive electrochemical signals from other neurons through connections called axons. In a simplistic vision of their mechanism, every excitation from another neuron increases or reduces the "excitation level" of the target neuron. Once the sum of the excitations reaches a threshold, the neuron fires a signal through its own axon towards another neuron, as illustrated by a filling bucket in figure 6.2. In artificial neural networks (ANN), this process is modeled by a node, or "artificial neuron", receiving signals $x_i(t)$ from its inputs (labeled by the index *i*) and producing an output y(t) which is a nonlinear function of the inputs. An ANN is formed by a network of interconnected nodes.



FIGURE 6.2: Left: representation of a biological neuron. Right: model of artificial neuron inspired from biology.

The inability to find a general optimization method to adapt these models to more complex tasks left the field at the level of scientific curiosity until the discovery of backpropagation algorithms [176]. Their introduction enabled finding an optimum network configuration with a high convergence probability thanks to backward propagation of the output error through the system. The details of these algorithms is beyond the scope of our research. In the rest of the chapter, we will suppose that solutions of the optimization problems can be found.



FIGURE 6.3: Analogy with a network of masses connected with nonlinear springs.

An ANN can be seen as a "big" and complex arrangement of linear and nonlinear interconnections. Stimulation of the network by an incoming signal excites some of its elements that transform and transmit the input to other elements. A mechanical analogy can be made with a set of masses connected with nonlinear springs of various properties. One can imagine that for a particular choice of springs, the movement of one mass could be a useful filtered function of an incoming excitation on the network. For example, it could attenuate some oscillation frequencies and amplify others. This case is hypothetical and only used to form a mental image of the working principle of an ANN.

The optimization, or "learning", consists of starting from a generic nonlinear system with many degrees of freedom and finding a set of parameters for which it solves the wanted task. One common method, called "supervised learning", applies a known input signal to the system and stores its output. An error value is calculated from comparison between the output and the expected answer. An algorithm iteratively adapts the parameters with a view to minimize the error.

It is interesting to note that a Darwinian scientist believing in evolution will see the brain as a result of a long optimization process rather than the product of a methodical design by an Architect. Consequently, it should not be surprising that the brain does not have a clear organization following a stepby-step algorithm. Similarly, ANN are built through an optimization process and the exact function of their components is usually unclear. Due to their complexity, it is impossible to predict their behaviour analytically. It is also not possible to determine where a given piece of information is located, introducing the notion of distributed computing and distributed memory. Only the state of the entire system contains the information required to produce the wanted function of the input signal fed into the system. It is usually impossible to break it down into subfunctions.

The introduction of neural network systems can seem disturbing for physicists, given that in general no logical reasoning can be made to explain the way the system processes the information. Although, considering the immense richness of general nonlinear phenomena embraced in section 3.2.2, one should not be surprised to observe a wide variety of new complex phenomena. The ensemble dynamics of a large number of nonlinearities simultaneously operating at different working points is a key element in the operation of an ANN.

6.1.2 Terminology and classes of neural networks

Any nonlinear system is, in theory, capable of information processing. The organization of biological brains is an example of a nonlinear system based on a relatively simple concept (a network of building blocks having similar properties) and able to solve complex tasks. The field of ANN focuses on certain types of architectures or "topologies" based on a few main criteria:

- simplicity: the system must be simple enough to be simulated or implemented on a dedicated hardware;
- usefulness: the system must be complex enough to be able to carry out useful operations;
- training: the optimization process must be as easy as possible;
- compatible with the tasks: every problem requires specific properties from the system.

The training is usually the most delicate part in the use of ANN. The choice of the system is made so as to be compatible with known optimization algorithms. A very simple, yet powerful type of ANN is the feed-forward neural network (FFNN). It comprises a number of layers, each made of an ensemble of neurons. The output of each of the neurons of one layer is directed to every neuron of the next layer only. Each neuron works as a sum of its inputs, then transformed by a nonlinear function. It is called "feed-forward" because the information cannot circulate backward to a previous layer. Training algorithms usually have a good probability of convergence. The model is well suited for tasks such as function approximation, classification problems or compression.

A more general type of structure, called "recurrent neural network", allows cyclic patterns where the output of a neuron can flow back to its input, eventually after passing through other intermediate neurons. The domain of application of such systems is much wider, including, for example, time series prediction and chaotic behaviours. However, the training process is much more complex and usually becomes inextricable for large systems. More recently, a new type of recurrent networks has been introduced independently by Herbert Jaeger [177] and W. Maass [178] respectively under the names of "echo state network" and "liquid state machine", presently better know as "reservoir computer". The major innovation consists in defining randomly the topology of the nonlinear network and training only a linear output layer by regression, which greatly simplifies the training.

Many other structures of ANN have been used, combining smaller networks of different types or with properties specific to a certain task. In this work, we exclusively focus on the case of FFNN, because of their simplicity.

6.1.3 Mathematical model of an artificial neural network

The most common models of neural networks use nonlinear systems made of a large number of identical nonlinear nodes. These are interconnected in the form of a network where the input of one node is a linear combination of the output of other nodes [179]. In the case of a FFNN, the network is organized in cascaded layers.



FIGURE 6.4: Comparison of a neural network architecture with a generic nonlinear system.

Figure 6.4 compares a general nonlinear system to the typical form of a simplistic FFNN. In chapter 3, we saw that the general case can be described by a Volterra series and potentially leads to extremely complex effects. In order to avoid using this heavy mathematical description, neural network systems are often defined as a discretized case of a nonlinear system with strong saturating nonlinearities represented by a simple explicit equation. The response of a neuron, schematized on the right of figure 6.2, is typically of the form (6.1), where y_j is the output of neuron j, \mathcal{F} is a scalar transfer function, a_{ij} a matrix defining the relative weights of each of the M input signals x_i in the excitation

of neuron j and b_j is a constant bias value.

$$y_j = \mathcal{F}\left(\sum_{i=1}^M a_{ij}x_i + b_j\right) \tag{6.1}$$

The canonical form of a three-layer FFNN with one output y and two inputs x_1, x_2 is defined by a state equation of the form (6.2). The nodes receiving directly the input signals form the "input layer" (with as many nodes as there are inputs). They send the signals to a second layer of neurons called "hidden layer". The end node combining the outputs of the hidden nodes forms the "output layer".

$$y = \sum_{j=1}^{N} w_j \mathcal{F}\left(\sum_{i=1}^{2} a_{ij} x_i + b_j\right) + b_{\text{out}}$$
(6.2)

Seen as a flow diagram, these operations consist of sending the inputs to all of the nodes with different coefficients a_{ij} plus a constant bias level b_j . The total excitation at each node is then processed by the nonlinear transfer function \mathcal{F} and a weighted sum with coefficients w_j of their contributions gives the output of the system. Finally, an output bias b_{out} adjusts the absolute level of y. Various types of transfer functions \mathcal{F} can be used, most of the time saturating nonlinear functions. A very common choice is the hyperbolic tangent defined by

$$\mathcal{F} = tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{6.3}$$

6.1.4 Optimization of neural networks

A large number of degrees of freedom are left undefined in equation 6.2. The definition of the connection matrix a_{ij} , input bias b_j , output weights w_j and output bias b_{out} allows for adapting the architecture to particular applications. No general rule exists to find these deterministically, but multivariate optimization algorithms have been developed to efficiently solve this as a blind error minimization problem.

The adaptation operation, commonly called *training*, is similar to the human learning process: a known sequence of inputs is presented to the ANN, for example a series of handwritten characters. For this particular set of inputs, the expected answer (the values of the characters) is known. The algorithm records the output of the system and compares it to the expected answer. In case of errors, it attempts to tune the parameters of the network in order to maximize the quality of the answer.

Compared to a human child learning the difference between good and bad food, the input sequence corresponds, for example, to a set of aliments available for eating. The output of the neural network is the decision of eating them or not and the error signal could be either a correction by an adult or a good or bad taste to confirm or infirm the choices. Throughout learning, the child changes his criteria of selection of proper food as a function of the feedback received.

This arduous nonlinear optimization has been declared inextricable for a long time and left neural networks as a mathematical curiosity until the recognition of a general method in the mid-1980's [180, 181], although it had been first discovered in 1969. Backpropagation algorithms are now a well known and stable method for training ANNs of type (6.2) [180, 182]. We will not detail further the working principle of these algorithms and will consider that the optimum of our simple ANN model can be found. In other words, we suppose that, given an ANN structure and input and output training sequences, it is possible to find the optimum parameters $\{w_j, a_{ij}, b_j, b_{out}\}$.

Direct training of hardware neural network

In the few experimental architectures realized with a view to demonstrate our new principle of frequency multiplexed neurons, we have adopted either a deductive approach (for the optoelectronic system) or a computer simulation (all-optical network) to determine the parameters of the neural network. For our basic logic gates (section 6.4.2), we were able to guess the coefficients thanks to the simplicity of the problem. In the case of our all-optical FFNN, the adequation of the experimental response with the simple theoretical model allowed to predetermine the parameters by running simulations.



FIGURE 6.5: Iterative optimization of a hardware neural network. A computer model determines an approximate set of weights. Processing of given learning samples through the system gives an output signal that is compared with the expected answer. Measurement of the error provides feedback to the optimization algorithm for the next iteration.

In a more general system involving a larger number of neurons, the modelling approach is unlikely to perform as good due to divergence from the ideal behaviour caused by crosstalk, gain saturation, etc. However, this would not obligatorily impair the performance of the system. A direct optimization of
the parameters based on the measured output of the real system could in principle adapt the weights so as to make the most of it as it is [183, 184]. The automated training process would measure the response of the system to a sample input, compare it to the wanted output and use the error calculated to drive an optimization algorithm, as illustrated by figure 6.5. The code would then compute the next set of experimental parameters with a view to improve the output signal. Iteration of this procedure would likely converge to an optimum in the same fashion as usual backpropagation codes do for digital neural networks. To improve the convergence of the algorithm, an initial guess of the parameters could be found by a simulation approach similar to the one used in our experiment.

6.1.5 A simple case: a 2-bit ADC

In order to help the reader in understanding the design process of a neural network, we will go through the successive steps of building an analog-to-digital converter (ADC) using a FFNN architecture. The choice of this example is motivated by the fact that we report demonstration of all-optical ADC operation later in this chapter.

ADCs are widely used for interfacing digital electronic circuits with analog environments. They consist in converting a continuous variable x(t), for example a voltage in the range $[0 \ V, 1 \ V]$, into a digital representation with a precision of D digits. In particular, an analog-to-binary ADC takes one analog signal as input and returns D binary signals representing a quantized version of the input. Here, we use a Gray code to encode the values, which is easier to compute by a neural network.

Gray code encoding

Gray code, or reflected binary code, is a binary representation of a number with the particularity that only one bit changes at a time when continuously increasing the number represented. This concept is well known in applications where undecided transient states must be avoided, such as linear encoders. The principle of Gray code digitizing of a continuous variable is shown in figure 6.6 in the case of encoding three bits.

Neural network structure

We aim at building a 2-bit analog-to-Gray ADC and choose to try a three layer FFNN structure with one input (the analog signal) and a hidden layer comprising N hyperbolic tangent neurons, described by equation (6.4). This choice is quite arbitrary, although motivated by the fact that it is one of the simplest non trivial ANN architecture and is consequently a good starting



FIGURE 6.6: Construction of binary code and Gray code for the three first bits. Gray code is a unique representation of a value with only one bit change at a time. MSB - most significant bit ; LSB - least significant bit.

point when approaching a new problem.

$$y = \sum_{j=1}^{N} w_j \underbrace{tanh\left(a_{1j}x_1 + b_j\right)}_{=Neuron_j} + b_{out}$$

$$(6.4)$$

No rule exists to predict the minimum size N_{\min} of a neural network for a given task. The topology is designed by trial and error, although common sense and know-how can help finding an initial guess. Using an excessively large number of neurons, with the hope of being sure to have enough, is usually a bad strategy. Large systems tend to be more sensitive to noise and the learning process is more likely to diverge. The generalization capability of the ANN is also reduced, in a similar way as fitting a set of points with a polynom of very high degree is likely to suffer from overfitting.

Training

We trained the parameters a_{1j} , w_j , b_j and b_{out} by running a Levenberg-Marquardt backpropagation algorithm for different values of N [185]. The sample sequence used in the training was generated by a random smooth analog input signal and the corresponding digitized Gray code values at every time step. We found out that the smallest network able to synthesize the 2-bit Gray code ADC function with negligible error is made of two hidden neurons $(N_{\min} = 2)$. Note that for a 2-bit analog-to-binary ADC, $N_{\min} = 3$, which justifies the choice of Gray code for a neural network of minimum complexity.

 $w_2 = 1$

After convergence for N = 2, the algorithm provided the following coefficients.

$$a_{11} = -339.92 \qquad \qquad a_{12} = 253.69$$

 $w_1 = 1$

$$b_1 = 294.27$$
 $b_2 = -98.37$

 $b_{out} = 0$



FIGURE 6.7: Evolution of the $tanh(G \times s)$ transfer function as a function of the gain G. For the limit $G \to \infty$, the function tends to the hard limit.

The high magnitude of the input weights a_{1j} and bias weights b_j comes from the fact that the algorithm tends to drive the nonlinearities in their highly saturated regime to mimic the shape of a step transfer function (or "hard limit") and obtain sharp transitions between the states. The argument of the hyperbolic tangent can be substituted by normalized coefficients $a_{1j} \rightarrow \tilde{a}_{1j} = \frac{a_{1j}}{G}$ and $b_j \rightarrow \tilde{b}_j = \frac{b_j}{G}$ multiplied by a gain G. The gain is chosen as $G = max_{i,j} |a_{ij}|$, where $max_{i,j}$ is the maximum value over the parameters iand j, so as to normalize the coefficients to 1. This leads to the new expression of the neuron values with the parameters written below.

Neuron_j = tanh
$$\left(G\left(\tilde{a}_{1j}x_1 + \tilde{b}_j\right)\right)$$
 (6.5)

$$\tilde{a}_{11} = -1$$
 $\tilde{a}_{12} = 0.746$
 $\tilde{b}_1 = 0.866$ $\tilde{b}_2 = -0.289$

6.1.6 Applications and potential of neural networks

Nowadays ANNs are used in a wide range of applications related to sorting, classification, modeling, decision making [186], signal analysis and pattern recognition [187], speech recognition [188, 189], hybrid car energy management [190], face detection [191] or stock market prediction [192]. Their particularity is the ability to perform machine learning without writing an explicit code of operation. Unlike classical algorithms, neural systems are optimized from example situations and the exact meaning of every internal parameter is usually unclear.

At a lower level of sophistication, neural networks are also able to operate simple logic functions with few resources. For example, two regular artificial neurons (made of a single nonlinear optical element with the scheme introduced later in this chapter) are enough to perform an XOR gate, while a CMOS technology implementation would require 4 transistors. Unlike most electronic computers that handle binary information, ANNs are in general analog computing devices. Thanks to redundancy and distributed information architecture, they are usually robust to noise and have generalization abilities that are a characteristics of "intelligent" systems able to infer from examples.

One can wonder why it is interesting to implement an ANN in optics although we are barely able to build a couple of standard logic gates [17]. Optical nonlinearities have complex dynamics and are extremely difficult to use in classical schemes that tend to mimic CMOS systems with optics. This is where neural networks can be of great interest, because they do not critically depend on the exact nature of the nonlinearity [183]. Also the complex interactions inside nonlinear optical materials (crosstalk, phase distortion, cross FWM products, Raman effect, etc.) are automatically compensated by the optimization process and can even be seen as an advantage to make the dynamics of the whole network even richer.

Neural networks are also well suited to process multilevel signals in logic operations. Although basic attempts have been made to handle multilevel logic with optical circuits [193], they remain very basic. Neural networks could provide a model to operate complex analog functions with such signals.

6.1.7 Hardware neural networks

Before considering the implementation of ANN with optical components, numerous architectures have been investigated for ANN processing. The most popular platform is certainly digital electronics. Software simulations, running on general purpose CPU, are the most widespread form of commercial ANN systems [194]. Specific digital devices are also a topic of active research, for example using field-programmable gate arrays (FPGA) [195, 196]. Analog and pseudo-analog electronic circuits have also been proposed and built [197–199].

More exotic neural network hardwares have been demonstrated. Among them, we can cite the measurement of the wave patterns formed in a bucket of water for speech recognition [200] or recording of cat brains activity with electrodes [201]. The diversity of the possible implementations supports the proposition that dynamical systems of very different natures have the capacity to process information.

6.1.8 Existing optical neural network schemes

On top of the systems cited in the last section, optics has also been investigated as a candidate for a neural network processor. A wide range of methods have been invented to translate neural processing into optical components. However, as stated last year by D. Woods and T. Naughton in Nature Physics [202], "*The* most successful optical computing paradigms will be those that use phenomena that are unique to optics." By these, they refer mainly to the multiplexing ability (wavelength, polarization) and the high speed reactions possible with light, already mentioned in our introduction chapter. Until now, however, the experiments have hardly taken advantage of them.

The first obvious idea that may come up for designing an optical neural network (ONN) is to reproduce the wiring pattern of the functional diagram on figure 6.8 (left) with optical components (right) [17, 203]. One can easily figure out a network of power splitters with calibrated splitting ratios directing the input signals to N amplifiers and nonlinear optical waveguides following the same pattern as the theoretical graph. The formal diagram on the right of the figure, which is only given as a visual example, does not specify the exact nature of the nonlinearities neither considers the eventual interferences that could occur at the nodes.



FIGURE 6.8: Basic spatially multiplexed implementation of an ANN. The architecture of the theoretical ANN (left) is directly translated into an interconnection of separate optical elements to form an ONN (right).

Given that the neurons are defined by physical paths separated spatially, we call this type of architecture "space multiplexed neural networks". For reasons detailed in section 6.2.2, this configuration is not optimal.

Most ONN schemes that have been proposed and demonstrated in the past were free space optics and based on the concept of spatial multiplexing. Optoelectronic systems combining arrays of optical sources and detectors, spatial light modulators and electronic nonlinearities arouse a lot of interest at the end of the 1980's [203]. However, the critical nonlinear component was still implemented in the electrical domain. Photorefractive crystals have also been envisaged to create a reconfigurable 3D holographic interconnection matrix between neuron channels [204]. Saturable amplifiers have been used as an all-optical nonlinear transfer function at low bandwidth [205].

More recently, a time-multiplexed architecture of a recurrent neural network of type reservoir computer has been introduced, where a single nonlinear element (all-optical [206] or opto-electronic [18]) processes all the neurons sequentially. This model permits implementation of large networks with a number of neurons up to N = 100 using a very simple experimental setup. However, the bandwidth of the system must be N times higher than the bandwidth of the input signal, which imposes a strong limitation on the processing speed.

In the next section, we introduce an alternative method for representing the ANN structure with optical components using the spectral parallelism of optics to reduce the complexity of the system.

6.2 Frequency multiplexed neural network

Neurons can be seen as separate information channels. Each channel receives multiple data inputs, adds them up and transforms the sum according to a nonlinear transfer function. Instead of separating the channels by traveling through different physical paths, we isolate the N neurons of the network by encoding them onto N distinct carrier frequencies copropagating along the same waveguide. The use of WDM induces constraints on the active components carrying out the operations on the optical signals, i.e. bias addition, nonlinear transfer and output summation.

For the saturating transfer function, we exploit nonlinear processes that can occur simultaneously at different frequencies without interfering excessively. The application of wavelength selective attenuation and phase shift is performed by FD-POP devices that also act as flexible filters for shaping the channels.

6.2.1 General principle

Figure 6.9 shows a simplified block diagram of a frequency multiplexed optical neural network. Its main building blocks are a configurable signal combiner (a FD-POP), an optical nonlinear transfer function and an optical dot-product operation. The FD-POP applies static levels of attenuation for different frequencies, defining the input weights. In this case of the figure, the nonlinearity comprises one amplifier, one nonlinear medium and a filtering device to exploit SPM or FWM based effects. Many other options could be considered for the nonlinear transfer function. Finally, the weighted addition of the N output signals is carried out by the all-optical dot-product introduced in chapter 5.



FIGURE 6.9: Building blocks of a frequency multiplexed ONN. Right insert: frequency spectrum of the input of the nonlinear transfer function showing the logical organization of the ONN.

The figure illustrates the organization of the neural network structure by detailing the various components of the optical spectrum after the FD-POP. The N neurons are encoded onto N separate frequency channels. Each neuron j receives power contributions from all input signals in proportions dictated by the input weight coefficients a_{ij} and input bias b_j . The addition of the input signals at a single frequency is represented by stacked arrows.

The input and bias signals must all extend over the whole bandwidth occupied by the N neurons. In our experimental setup, this is achieved by encoding the signals onto a wide flat frequency comb. Also, the signals and bias must be phase locked before merging inside the FD-POP to avoid random interference instabilities.

6.2.2 Advantages of frequency multiplexing

Logical complexity

Frequency multiplexing exploits the parallelism offered by WDM optics to reduce the hardware complexity. It theoretically enables implementation of neural network architectures of any size on a fixed number of physical components. The argument that the amount of nonlinear elements does not scale with the complexity of the logical function is critical in the context of nonlinear optics that struggles with integration and cascading.

On the opposite, the optical realization of a neural network by space multiplexing is prohibitive due to the multiplication of nonlinear waveguides and amplifiers required as the number of neurons increases. In that obvious case, the complexity of the hardware scales linearly with the complexity of the logical ANN architecture.

Reconfigurability

The frequency parallel technique using a FD-POP described later allows for reconfigurability of the network without any physical change in the setup. The topology of the ONN is defined by filtering patterns, attenuation and relative phases applied by the FD-POP device. Switching to a different neural network function or architecture can be made by simply reconfiguring it.

In contrast, a space multiplexed implementation with hardwired connections between nonlinear elements would result in a static configuration. A change in the number of neurons, topology or task solved would require non trivial rewiring.

Reduced multipath interferences

In the frequency multiplexed architecture, all the signals travel through the same path in a single waveguide, eliminating all potential interference issues inside the processor.

In the vision of the simple ONN sketched on the right of figure 6.8, every node combining signals from multiple optical waveguides (at the input of the nonlinearities and at the main output) is subject to create random interference patterns if the system is not perfectly stabilized.

6.3 Implementation

6.3.1 Signal generation

The input of the optical nonlinearity combines all the inputs in different proportions for each neuron channel. To allow every neuron j to receive $\sum_{i=1}^{M} a_{ij}x_i + b_j$, every one of the initial input signals and bias must be present at every frequency channel. As a consequence, generating the data at one single wavelength, as it would commonly be done for data transmission, is not enough. Each of the input signals and bias must be encoded on N redundant channels. Here, we used equally spaced channels for simplicity and better spectral efficiency by reducing the guard bands between neurons.

The method used in our experiments was similar to the one described in chapter 5. It employs a flat frequency comb source generated with a 40 GHz modelocked fibre laser centred at 1550 nm and spectrally broadened. The wide optical spectrum resulting from the two HNLF stages is reduced to a bandwidth of 6 nm in its flat top region to avoid formation of ultrashort pulses with high peak power that could damage the modulator. The pulse train is encoded using a Mach-Zehnder modulator and a FD-POP is used for spectral slicing of the broadband signal into N channels.

The technique permits flexible definition of the centre frequency and bandwidth of each channel. Also, the comb lines constituting the spectrum are naturally phase locked and are regularly spaced

6.3.2 Nonlinear transfer function

The richness of nonlinear optics offers multiple ways to nonlinearly transform information. For the purpose of an ONN, we seek a saturating transfer function. It must also be simple to implement and compatible with the frequency multiplexed architecture. Here we compare three suitable effects based on SPM, FWM and two-photon absorption (TPA). Our choices were inspired by the use of similar setups for 2R-regeneration that also requires a saturating nonlinear transfer function [207].

SPM based nonlinearity

In a Kerr medium, SPM causes spectral broadening proportionally to the peak power of the optical pulses. Bandpass filtering with a frequency offset with respect to the carrier of the incoming signal selects a part of the spectrum that grows progressively when the input power rises and reaches a ceiling after a certain threshold. This effect has already been described in section 5.3.1 as a thresholder for binary signal regeneration.

Although the intensity transfer function of this scheme presents a saturation, its shape is not ideal. SPM also induces phase distortions that are unwanted for phase sensitive combination of the channels. The process also requires a wide bandwidth due to the guard bands needed to avoid spectral overlap of the broadened channels. For these reasons, we did not explore SPM any further.

FWM based nonlinearity

Generation of an idler product via degenerate FWM with a strong pump and weak signal is known to present a linear growth with the intensity of the signal and quadratic with the pump. Hence FWM could seem not well suited for implementing a saturating nonlinearity.

$$A_{\rm idler} \propto A_{\rm signal} \times A_{\rm pump}^2$$
 (6.6)

However, relation (6.6) is only valid under the assumption of an undepleted pump. FWM process with signal intensity comparable to the pump leaves this regime and exhibits saturation. This phenomenon can be understood as an inversion of the roles of signal and pump in the FWM mechanism, once the signal reaches a higher intensity than the pump. The initial generated idler then becomes a second-order idler with significantly lower efficiency. The transition from one regime to the other leads to a saturation of the idler. It is important to note that, thanks to the phase matching condition inherent to FWM, the phase relation between signals, pumps and idlers are maintained independently of the power, unlike SPM.

Later in this chapter, we will show measurements of the FWM intensity transfer for various configurations in good agreement with the ideal tanh(x) function (see figure 6.17). This was one of the main arguments for using FWM in our final experiment.

TPA based nonlinearity

TPA is a nonlinear process causing an intensity dependent loss through the medium. It was a potential candidate for creating a saturating transfer function. However, we will see later that FWM allows a significant simplification of the setup by combining two steps into one. As a consequence, we did not investigate TPA any further.

6.3.3 Bias summation

The general equation (6.2) of a neural network involves addition of a bias signal at the input of every neuron with weight b_j . This component allows for changing the working point of the nonlinearities and is essential for most useful applications.

An optical equivalent of adding a constant bias value can be made by coherently combining the input signals with a continuous wave source or a pulse train in the case of pulsed signals. This operation was performed inside a FD-POP where the data signal and bias paths were directed to two separate ports of the device.

Both bias and data signals originated from the same comb source and followed different paths. The slight variations of path length due to ambient fluctuations caused random changes in the relative phase between them that were compensated using a piezo-electric phase shifter. An enclosure isolated the interferometric setup from the environment for improved stability.

6.3.4 Dot-product

The last operation carried out by a neural network for generating the final output signal is the weighted sum of the outputs of all the neurons. For a frequency parallel ONN, it comes down to summing the N frequency channels at the output of the nonlinear transfer function.

The next section presents a simple scheme where the dot-product is operated with an optoelectronic component. However, the simplicity comes at the cost of a very restricted system due to the incapacity to handle negative values. Finally, section 6.4 presents an all-optical architecture implementing the general case of a dual layer feed-forward neural network.

6.4 Optoelectronic reconfigurable logic gates

The all-optical dot-product operation of chapter 5 is a delicate part of the system. We aimed at performing a preliminary test of frequency parallel processing without having to use this complex nonlinear mixing scheme.

Although the optoelectronic dot-product has very few practical applications and is limited by the bandwidth of the detector used, it was a useful milestone to validate the principle of frequency multiplexing of neurons. It led to the realization of XOR and AND logic gates between WDM channels without optical nonlinearity.

6.4.1 The easy way: power adder

An immediate method for calculating a weighted sum of WDM neurons is to selectively attenuate each channel according to the output weights w_j and detect the signal with a broadband photodetector. The detector outputs an electrical signal proportional to the total optical power in all the channels, which operates a sum. Wavelength selective attenuation can be performed with a FD-POP inserted before the detector. This partial solution is only valid for addition of positive signals weighed by positive coefficients. The intensity detector looses information on the phase, that contains the sign of the signals. The bias b_{out} can be added as a supplementary optical channel of constant power.

Mathematically, the power adder translates into an absolute value of the neurons outputs and weights resulting in a neural network equation of type (6.7), where y, x_i , b_j and b_{out} correspond to optical powers.

$$y = \sum_{j=1}^{N} \left| w_j \mathcal{F}\left(\sum_{i=1}^{M} a_{ij} x_i + b_j\right) \right| + b_{out}$$
(6.7)

In practice, such restricted system gives poor results and very few applications can be done with it. Simulation of the system for many simple, yet



FIGURE 6.10: Using an intensity detector as a channel adder. Right graph: configuration of the FD-POP to apply the weight coefficients w_j .

useful, tasks and attempt to run a backpropagation algorithm did not converge to a valid set of parameters. This can be understood as the fact that neurons cannot have an inhibitive effect (negative values) in the sum.

To improve the simple scheme, a balanced photodetector can be used to enable addition and subtraction of signals. On top of applying the attenuation coefficients, the FD-POP is also configured to route the positive and negative contributions to two different output ports connected to the respective inputs of the balanced photodetector. As a consequence, a neuron can have either an inhibitory or additive role in the dot-product.



FIGURE 6.11: Using a balanced photodetector as a channel adder and subtracter. Right graph: configuration of the FD-POP to apply the weight coefficients w_j and direct the channels to the ports of corresponding signs.

However, this naive approach still does not provide a general implementation of the dot-product $\sum_{j} (w_j \text{ neuron}_j)$. Although negative weights are now allowed, the power adder returns the absolute value of the neurons output. It is also equivalent to replacing the transfer function by $|\mathcal{F}|$ as in equation (6.8).

$$y = \sum_{j=1}^{N} w_j \left| \mathcal{F}\left(\sum_{i=1}^{2} a_{ij} x_i + b_j\right) \right| + b_{out}$$
(6.8)

In spite of the improvement, the functionality of the neural network remains strongly limited and it does not work for most real world tasks. A simulation of the system and backpropagation optimization for many tasks generally showed that no satisfactory set of weights could be found, even for large networks of over 10 neurons. Two simple particular cases where the method could be used are reported in section 6.4.2.

Note that one could think of using coherent detection to reconstruct both phase and amplitude of the optical fields, hence preserving the sign of the values encoded in the phase. However, coherent detection could not be performed if the total bandwidth Ω of all the neurons was higher than the bandwidth of the detector (about 42 GHz). Such an approach would loose some of its interest, given that coherent detection requires a DSP postprocessing operation that removes part of the advantages of all-optical processing.

6.4.2 Logic gates

A first simple demonstration of information processing using a frequency multiplexed ONN architecture was made without optical nonlinearity and with a power adder dot-product module, as introduced in the previous section. Understanding of this section is not required for the rest of the work, but can help in building a better mental picture of how a very basic neural network functions.

We designed the structure of this simplistic network without resorting to a backpropagation optimization algorithm. Instead, we inferred the parameters by analysis. The setup described in figure 6.12 was configured with the FD-POP to create one or two optical neuron channels. We will now show that such circuit can operate AND and XOR logic operations between phase-locked signals.



FIGURE 6.12: Experimental diagram of an optoelectronic proofof-principle WDM neural network.

Given that the power adder dot-product scheme does not require any particular phase relation nor frequency spacing between the channels, they could in principle be generated using separate continuous wave laser sources. However, this approach raises the issue of glitches which is further developed in appendix D. To overcome the problem, we employed the frequency comb source introduced previously.

The initial pulse train was encoded with data and passed through a FD-POP to carve two WDM channels. It also applied a multipath delay line interferometer transfer function in order to generate relative delays and pseudo mixing between the signals in a similar fashion as introduced in section 5.6.1. The coefficients of the distortion matrix were determined by the design of the network structure explained in the following flow diagrams.

XOR gate configuration

Figure 6.13 gives a flow diagram representation of an XOR logic gate between two signals. A and B are the two input signals (for example originating from different paths) that are transmitted to the two wavelength channels C and D with fixed weights. At point E, they are converted into a positive electronic signal by a photodetector that provides the output signal. In this case, channel C is not needed and has a constant value 0, which makes the XOR operation a trivial case with a single neuron.



FIGURE 6.13: Flow chart of an XOR gate by subtraction of two signals and absolute value.

Table 6.1 describes the state of the various nodes of the flow chart for every state of input signals. The values at the output node E correspond to an XOR operation of the inputs. Configuration of the FD-POP with coefficients corresponding to the input weights and output weights of the flow chart realized the experimental XOR function.

A	0	1	0	1
В	0	0	1	1
C=0	0	0	0	0
D = B - A	0	-1+0	0+1	-1+1
E = D	0	1	1	0

TABLE 6.1: Logical states of the different nodes of the network configured as an XOR gate.

Measurement of the output of the system was made for a user-defined 64 bits bit-pattern. The right part of figure 6.14 shows time traces of the input signals (top) and gate output (bottom). It can be observed that the output signal is indeed an XOR operation of the inputs. The dotted lines are used as reference to show corresponding bits.



FIGURE 6.14: Measurements of XOR operation. Left: bit-error rate curves and eye diagram (inset). Right: time traces of the input signals and XOR output. B2B: back-to-back

The left graph of the figure features bit-error rate curves of the XOR output signal. We limited the measurement to a BER of 10^{-7} due to the slower acquisition rate of the BER tester for custom bit-patterns. The modelocked fibre laser used as source was impaired by random mode hoping causing the absolute position of the frequency comb to change abruptly with an average period of a few minutes. Given that the spectral pattern applied by the FD-POP needs to be precisely aligned with the spectrum of the signal, BER measurement without modification of the experimental parameters could not be achieved with a longer integration time.

AND gate configuration

Conception of an AND gate with the same setup is also possible by reconfiguring the FD-POP. Figure 6.15 shows the flow diagram for an AND function between two signals. Again, A and B are the two input signals that are added with given weights to the two wavelength channels C and D. The subtraction required at point E is made possible by assigning the channels to different ports of the balanced photodetector. Here, both neuron channels are needed.



FIGURE 6.15: Flow chart of an AND gate based on two WDM neurons.

In a similar fashion, table 6.2 describes the states at the various nodes

A	0	1	0	1
В	0	0	1	1
C = A + B	0	1+0	0+1	1 + 1
D = A - B	0	1 - 0	0 - 1	1 - 1
E = C - D	0	0	0	1

for every combination of input signals. The values at the output node E correspond to an AND operation.

TABLE 6.2: Logical states of the different nodes of the network configured as an AND gate.

The BER assessment of the system is reported below. It is interesting to note that the error rate was lower after the gate than for the back-to-back detection of the input at same received power. This is due to a pattern effect, in particular the fact that the average number of pulses ("1" level) present on the output signal after the gate was smaller than the 50% ratio of the two input signals due to AND operation. For two random and independent input signals with equiprobability of "1" and "0" states, the resulting AND product would have a fraction of 25% of bits at the "1" level. As a consequence, the average power detected was lower for a comparable peak power of the pulses. The use of both arms of the balanced photodetector also partially cancelled the signal noise.



FIGURE 6.16: Measurements of AND operation. Left: bit-error rate curves and eye diagram (inset). Right: time traces of the input signals and AND output.

In this architecture, no optical nonlinear element was used. However, the signals were processed by the nonlinearity consisting of the absolute value operation by the intensity detectors. Also note that in figures 6.14 and 6.16,

we chosed to fit the BER measurements with curved lines (dashed) inspired by the alignment of the data points instead of straight lines.

6.5 All-optical frequency parallel neural network

We now combine the previous partial attempt of an optoelectronic neural network with the optical dot-product introduced in chapter 5 to create an alloptical and general feed-forward neural network architecture. The most elegant part of this work is the fact that the nonlinear transfer function and the dot-product are integrated into a single physical device.

6.5.1 Combining nonlinearity and dot-product

An all-optical operation corresponding to the mathematical expression of a dot-product has been introduced and used in chapter 5. It can be used for recombining of the signals from the hidden-layer of the neural network, in the manner of the sum over the index j in equation 6.2. However, it requires cascading the dot-product operation with another nonlinear optical component applying the saturating transfer function to each of the neurons. Such solution is not trivial and would increase the complexity of the setup.

In section 6.3.2, we concluded that a FWM based nonlinear transfer function was an ideal choice to reproduce a classical hyperbolic tangent. On the other hand, our all-optical dot-product also involves multiple FWM processes. This common point suggests combining both operations into one single step. However, it is not a straightforward mode of operation and it first needs to be verified that the dot-product scheme still performs well in the saturation regime with significantly higher signal levels and that the shape of the nonlinear transfer function remains adequate when combined with the interferometric idler adder.

The linear behaviour of degenerate FWM is only valid in the approximation of the undepleted pump. For signal intensities comparable to the pump, the growth of the idler power presents a saturation, as seen previously. Here we chose to study the dot-product scheme with strong signals. Experimental measurement of the transfer characteristics of the dot-product in the saturating regime exhibited a surprisingly good match with the wanted hyperbolic tangent behaviour despite the presence of multiple signals.

Transfer function for one channel

The nonlinearity was first studied regarding one input signal. Figure 6.17 shows the superposition of an experimentally measured transfer function with



FIGURE 6.17: Comparison of the experimental transfer function with a tanh(x) function. Negative input values (cf. signal shifted by a π phase shift) has been inferred from the data in the first quadrant of the axes.

a classical tanh(x) fit. The measurement was made while another channel had nonzero constant power. The corresponding constant background power at the output was subtracted from the measurement to focus on the power dependence in the channel of interest only.

The experimental data points corresponding to negative values of the input signal were inferred from the data measured for positive values. The sign of the signal is encoded by application of a π phase shift. Given that the absolute phase of the input does not influence the power of a single idler created by degenerate FWM, the power transfer function is the same for positive and negative inputs. Thanks to the phase matching condition of FWM, a phase shift of the signal is reflected by an equivalent phase shift on the idler. Hence, the sign of the signal is conserved through the nonlinearity, while the amplitude is affected by a saturation. As a consequence, the study of the transfer function for positive signals can be extrapolated to negative values by a symmetry regarding the origin.

Transfer function for two channels

The study of the response of the system with respect to a single input signal is not enough to characterize the ONN architecture. To take into account the eventual mutual influence of multiple signals, we measured the idler power as a function of two incoming signal channels.

The measurement setup used to obtain these two-dimensional power maps comprised a flat comb spectrum source described previously. A FD-POP sliced the spectrum of the signal and pump channels, applied variable attenuation levels and controlled their relative phase. An optical amplifier boosted the total power up to 300 mW before passing through 30 m of HNLF. A tunable



FIGURE 6.18: Experimental setup for nonlinear transfer measurement of the dot-product. Inset shows the output spectrum of the HNLF. Solid: signals add coherently ; dashed: subtraction of equal signals.

narrow-band 70 GHz filter extracted the idler channel to be detected by a power-meter. A computer program controlled the FD-POP and the power-meter to automate the data acquisition for various combinations of relative powers of the signals and pumps.

We observed that the scheme conserves the independence of multiple signal channels. This can be an important issue in the case of frequency multiplexed neurons, given that their concurrent processing in a single nonlinear medium is likely to cause cross-effects by complex nonlinear interactions. Figures 6.19 and E.1 compare the transfer function measured experimentally from the system with the expected transfer of a classical neural network comprising two neurons with tanh(x) characteristics.

Figure 6.19 (top) shows the measured output power of the idler after a 30 m section of HNLF used in a dot-product scheme. The idler was extracted by sharp narrow band filtering (70 GHz) using a FD-POP. The remarkable symmetry of the graph regarding the bisector of the axes shows a quasi perfect independence of the two neurons (labeled *Signal 1* and *Signal 2* on the plots).

The bottom plots present a simulated version of the transfer function calculated for a system with two tanh(x) neurons, corresponding to the canonical FFNN in equation (6.2). The equation of the 2D map is

$$y = |tanh(S_1) + tanh(S_2)| \tag{6.9}$$

where y is the output of the dot-product and S_1 and S_2 correspond to Signal 1 and Signal 2. By the exact symmetry of the expression regarding S_1 and



FIGURE 6.19: Transfer function with two neurons measured experimentally (top) and theoretical simulation(bottom). The color maps in the middle represent the output power as a function of the intensity of the two neurons. One-dimensional transfer profiles sliced according to the axes are featured in the left and right series of graphs.

 S_2 , slices parallel to the two axes are identical. Hence only one series of unidimensional plots is featured. Comparison of the theoretical and measured figures reveals the strong similarity of behaviour of the saturating dot-product with the canonical transfer function of a classical neural network.

A similar study was carried out in the case of subtraction between the neurons. In this situation, the dot-product was configured to operate coherent subtraction of the two idlers. These results are detailed in appendix E.

Note that in the experimental setup both signals and pumps were amplified

through the same EDFA. The total power delivered by the amplifier was constantly 300 mW and the power given to the pump channel was reduced for high signal powers: $P_{pump} = P_{tot} - P_{signal1} - P_{signal2}$. Hence the pump power was comprised between 300 mW (no signal) and 175 mW (both signal at 62.3 mW), creating a distorsion of the data measured, in particular when both signals had maximum power. As a consequence, the static response showed on the figures is affected by a slight artificial roll-off of the idler power measured for high signal power.

6.5.2 Experimental demonstration of the ONN

The pulse train from the frequency comb spectrum source was split onto two arms of a polarization maintaining 50/50 coupler. One path was modulated by a Mach-Zehnder modulator (MZM) to form the data channel. The other one was delayed so as to exactly match the length of the data path and controlled in phase by a piezo-electric phase shifter (PS). In the case of the results presented in the next section, the data signal was an intensity ramp generated by driving the MZM with a sine wave at 1 GHz. However, this signal would be in principle any phase-and-amplitude modulated form of encoding.



FIGURE 6.20: Experimental diagram for transfer characterization of the neural network. A first FD-POP coherently combines bias and data signal into the neuron channels and carves the pump channel out of the bias input. The dot-product in saturation regime combines the neurons. PS - phase shifter

The two paths, corresponding respectively to the modulated data and constant bias/pump signals, were combined together inside a first FD-POP. The device was programmed so as to compose the channels of the two neurons (N1 and N2 on figure 6.20) from fractions of the data and bias signals according to the input weights a_{ij} and bias weights b_j . To apply custom phase and attenuation levels to both input ports and combine them coherently into a single output port, we took advantage of the ability of the FD-POP to address a single channel to multiple ports. The non modulated input was also carved spectrally to create the pump channel.

The resulting optical field was then amplified using a low noise EDFA and directed to the saturating dot-product device. Although the phase and amplitude of the channels involved in the dot-product could have been controlled by the first FD-POP, we kept this operation separate by using a second device for ease of programming. The rest of the experimental setup was similar to the one of the dot-product detailed in chapter 5.

It is important to note that the coherent addition of signals inside the first FD-POP and inside the dot-product sum the fields and not the intensities. This point must be taken into account for the configuration of the various coefficients.

6.5.3 Results for a 2-bit analog-to-digital converter

Feed-forward neural networks are capable of numerous applications ranging from function fitting to automated decision, pattern recognition, classification and data compression. The parameters of the general purpose ONN architecture developed in the previous section must be tuned to accomplish a target function. From the wide range of possible applications, we selected one that could reasonably be implemented on our system. The easiest non trivial demonstration of a neural network function would ideally be done with the smallest possible number of neurons, with a minimum of two. We made the choice of a Gray code ADC as a compromise between ease of implementation and practical interest of the function [208].

As concluded in section 6.1.5, analog to Gray code conversion for the second bit can be operated with only two neurons. Interestingly, it also requires only one input data signal (the analog signal to digitize) and a bias, which removes the need for generating multiple phase locked data signals. Gray code naturally avoids the issue of undetermined transient states due to an imperfect hard limit transfer function. We also selected the ADC as a demonstration of principle for its intuitive understanding by a public not familiar with neural networks and their more abstract applications. It is a well known function and a challenging topic of active research in photonics [209, 210].

Network training

The optimization of the parameters of the system was obtained by offline simulation of a model of neural network, followed by manual tuning of the various experimental parameters. Common algorithms involve iterative optimization by backpropagation of the output mismatch through the system. It was not feasible to implement this process directly on our experimental system without a phase stabilization of the interferometer by feedback and it would have been a complex process unless automated.

Instead we optimized the parameters based on a simple model of the system described by the canonical equation of a neural network with a hyperbolic tangent as nonlinearity and only one input, as in equation (6.10).

$$y = \sum_{j=1}^{2} w_j \underbrace{tanh(a_{1j}x_1 + b_j)}_{=Neuron_j} + b_{out}$$
(6.10)

This optimization has already been done in section 6.1.5 where the parameters were expressed as a function of a gain G corresponding to the following equation of the neurons.

$$Neuron_j = tanh\left(G\left(\tilde{a}_{1j}x_1 + \tilde{b}_j\right)\right)$$
(6.11)

Note that in our optical system, the gain G corresponds to the level of the signals relative to the pumps in the saturating dot-product scheme. Indeed, saturation occurs when the power of the signal becomes comparable to the pumps or higher. We used this set of normalized coefficients to configure the input layer of the experimental neural network (defined by the first FD-POP). The gain G was controlled by adjusting the relative power of pumps and signals. G could not take values as high as in the simulation, although we could reach a sufficient level of saturation to demonstrate non trivial neural network operation.

In the results reported in the demonstration of principle in the next section, the coefficients were implemented with one notable modification. Instead of using $w_1 = w_2 = 1$ given by the backpropagation algorithm, the output weights were changed to have opposite signs ($w_1 = 1$ and $w_2 = -1$). This produced the opposite of the second bit of Gray code, as shown later in figure 6.23.

Several reasons justify this change:

- with both neuron outputs subtracting from each other, the network is more symmetric and imperfections partially cancel;
- this allows us to push the nonlinearity to further gain values while the deviations from ideal behaviour roughly cancel on both channels;
- the visual effect on the response of the system is more pronounced: the expected output is maximum at zero input, which demonstrates the proper action of the bias signals.

$$w_1 = 1$$
 $w_2 = -1$

Finally, manual tuning of the parameters and instruments was done with a view to improve the results, account for mismatch between the simple analytical model and the experimental setup and compensate for environmental fluctuations.

Measurement of neural network output

We tested the ADC ability to digitize an analog signal, which required generation of a multilevel data. For simplicity, we modulated a pulse train with a sine wave using a Mach-Zehnder modulator in order to obtain a slowly increasing intensity ramp shown on figure 6.21. This method allows for clear identification of the intensity transfer function of the neural network and has been used in reference [210] to characterize a 2-bit ADC.



FIGURE 6.21: Input signal. A pulsed power ramp is created by carving a 40 GHz pulse train with a Mach-Zehnder modulator driven by a 1 GHz sine wave. The inset details the pulsed structure.

The optical spectrum at the output of the HNLF is plotted in figure 6.22 (c). The signal is configured to have an average power 5 dB higher than the pump. This regime corresponds to the optimum working point of the experimental setup (motivated later in figure 6.23). The idler extracted with a 70 GHz bandpass filter is shown in the time domain in 6.22 (b). The top diagram corresponds to the input signal after frequency conversion inside the HNLF. (b) shows the response of the neural network in the second bit of inverted Gray code ADC configuration and compares the experimental data to the expected ideal ADC output.

According to the expected transfer function of the second bit of the inversed ADC (represented on figure 6.22 (b) by a blue trace), the output should be '1' for an input power comprised in [0, 0.25] and [0.75, 1] and '0' for the rest. The pattern measured (yellow) corresponds to the wanted answer, with smoothed variations instead of steep transitions due to the non ideal nonlinearity. This smooth transition was also observed in reference [210].

Figure 6.23 (a) shows a simulation of the influence of G on a system with neurons having an ideal tanh(x) transfer function. The green dashed trace represents the intensity ramp given by the Mach-Zehnder modulator. The solid blue traces are the transfer functions of the neural network with various gain levels. At the limit of high gain, the response presents sharp transitions.



FIGURE 6.22: Experimental neural network output. (a) Time trace of the input signal. (b) Output in ADC configuration (yellow) and expected output (blue lines). (c) Optical spectrum after the HNLF in ADC configuration.



FIGURE 6.23: Evolution of the neural network response with the input gain of the nonlinear transfer function. (a) Simulation. Dashed: Input signal envelope ; solid: Output envelope for various input gains G=[4, 8, 16, 80]. (b) Measured input signal after wavelength conversion. (c-f) Measured outputs for various values of pump attenuation.

The subfigures (c) to (f) show the evolution of the output measured for various signal-pump relative intensities. Higher signal power corresponds to a higher value of G. Although the difference between the regimes is not as clear as in the simulation, the data show a similar trend. At low gain (G = 4 in the simulation), the output exhibits two peaks along the input ramp, one being lower than the other, which also appears on subfigure (c). When the gain increases, the relative heights of the two peaks tend to reach the same level, which is also the case in subfigures (d) and (e). For very high gain (G = 80), the theoretical response tends to a rectangular function, although the effect rolls off in the experiment as the system saturates and the waveform no longer evolves: (f) does not tend to a rectangular response.

Performance and limitations

Although at its embryonic stage of development, our ONN was able to perform a 2-bit Gray code ADC function at 40 Gb/s, which is the state-of-the-art performance obtained from SPM based all-optical ADC [209, 210]. The maximum possible bandwidth of this type of ADC is discussed in section 6.6.1. A more thorough investigation of the ADC should measure the effective number of bits (ENOB) of the device. Other all-optical and optoelectronic ADC have been invented, partly driven by the need for high speed waveform analysis and satellite communication technologies [211]. A method based on microring resonators has been proposed [212]. Soliton self frequency shifting allowed experimental demonstration of 4-bit quantization at 3.3 GS/s, with a potential for higher bandwidths [213]. An optoelectronic ADC with optical time-stretch preprocessing has achieved a sampling rate of 480 GS/s with an ENOB of 5.17 [214].

An ideal ADC would present sharp transitions between the states of each bit. For operation in this regime, step transfer functions, also called hard limiters, are required although such effect is difficult to obtain with an optical system. In particular, our saturating dot-product provides a smooth transition between stable states as shown on figure 6.17. This issue, related to the level of saturation of the nonlinearity, has been discussed before and is due to the limited magnitude of the gain G applicable on the real system. It is also present in the other schemes of all-optical ADC.

For this last reason, the Gray code encoder was not a perfect choice for demonstrating our ONN with a real task. Other functions can potentially be solved by neural networks without requiring hard limiters.

6.6 Practical functions

6.6.1 Complexity limits

The frequency parallel scheme offers numerous advantages in terms of hardware complexity versus logical complexity. However, the size of the network that can be practically implemented on such platform is finite. Various reasons forbid the realization of a neural network of arbitrary dimension.

Bandwidth \times Neurons product

Frequency parallel ANNs encode the multiple neurons on distinct frequency channels. Each channel has a given bandwidth dictated by its baudrate. Given that the transmitting bandwidth of optical components is limited, the total aggregated bandwidth of all the channels must fit within this range.

For a network of N neurons, each having a bandwidth B and implemented on a system of minimum bandwidth Ω , the Bandwidth \times Neurons product condition is expressed:

$$\overline{N \times B < \Omega} \tag{6.12}$$

On top of this, one must note that an additional guard band is usually required to avoid interference between channels.

One of the most limiting factors of Ω is the bandwidth of the dot-product platform. In our experiment, we used a 30 m section of HNLF that provided a 3 dB bandwidth of about 10 nm. This range was sufficient for our needs, however, nonlinear devices with wider available bandwidth such as on-chip dispersion engineered waveguides may be desirable for bigger ANNs.

Parasitic effects

The steps involving nonlinear optics, i.e. the dot-product and nonlinear transfer function, are also the origin of unwanted effects such as XPM, SPM and higher-order FWM products. These introduce parasite phase distortion and power transfers that can become detrimental when the number of cross combinations increase. For the small system studied, we have not detected significant influence of these phenomena.

In the broader context of blind optimization of the global system, however, these side-effects might become beneficial. By "enriching" the dynamics of the network, they could potentially improve the computational capabilities of the ONN. Note that the equation governing the ANN would then not anymore match the canonical definition given in section 6.1.3.

Adjustment complexity

A larger number of neurons implies initial adjustment of the relative phases and amplitudes in the dot-product and bias adder modules following the procedures explained previously. These operations may become fastidious for large systems and require long term stability of the laser sources and components.

6.6.2 Training

As said previously, the configuration of a neural network is usually optimized starting from a random set of weights until it converges towards a minimum error. Here we summarize the implications of an ONN on the training process.

The optical implementation sets physical limits to the range of admissible parameters due to the power handling of the components, OSNR degradation for small inputs and divergence from ideal behaviour for extreme values. On top of this, direct optimization of the system would require automating the measurement of the error based on the experimental output for each guess of weights. This approach is discussed in section 6.1.4.

Convergence of the optimization process can be dramatically sped up by judicious guessing of the initial set of weight coefficients [215]. A good approximation of the optimum of the real system can be deduced from simulation of a simplified model of the ONN. For a given application, we first simulated various ANN architectures in order to determine the minimum requirements of the network in terms of number of neurons, bias and symmetries to accomplish the task. For the selected architecture, we then extracted the initial guess of coefficients from the backpropagation solution.

Learning algorithms running on computer models tend to push the magnitude of the parameters to extreme high and low values to match the target output as close as possible. These ranges of values suppose ideal transfer functions that are not physically realizable. In order to be able to use the common neural network learning methods, we approached our optical nonlinearities with classical functions extending over the whole range $[-\infty; +\infty]$. A first rough approximation is given by the functions tanh(x). We used a Levenberg-Marquardt backpropagation algorithm to train the networks based on initial noisy sample sequences.

The calculated solutions typically feature excessively high coefficients that can be reduced, at the cost of a slightly increased mismatch with the target output. In the case of our small systems, a simple consideration on the theoretical weights usually suffices to infer a realistic configuration. Some examples of informed tuning of the parameters are provided in the following sections. Note that systematic methods exist to restrict the magnitude of the weights inside the optimization algorithm, such as the penalization method.

6.6.3 Possible applications of a simple ONN

To conclude on the experimental demonstration of frequency multiplexed ONN, we summarize three possible extensions of our work to solve practical functions.

Nonlinear distortion compensation

Neural nertwork processing could be used to undo unwanted nonlinear distortions occuring in long haul transmission. Beating the nonlinear Shannon limit is a present focus of research in photonics. A neural network scheme implemented on digital signal processing has for example been proposed to compensate for Raman crosstalk in WDM networks [216].

MAX gate

The development of multilevel and coherent modulation formats has added another challenge for optical signal processing. Few all-optical schemes exist for such signal and processing relies exclusively on DSP. Optical neural networks are inherently analog and can perform multilevel and complex-valued operations. Possible applications would be multilevel logic gates such as MAX or MIN [217, 218] or more advanced functions.

Analog-to-digital conversion

A more advanced version of our 2-bit Gray ADC could give access to more digits. It would require networks made of a larger number of neurons, i.e. more frequency channels. This has been shown possible by simulations proving that backpropagation optimization of more neurons could perform quantization of more bits.

Sharper transitions between the states are also desirable. This could be achieved by either driving the nonlinearities harder, employing nonlinearities with steeper transfer function or possibly using networks with more neurons than the minimum required for a given task.

6.7 Future of neural network in optics

The recurrent argument that optics cannot compete with digital signal processing is, as stated earlier, based on the principle that it should mimic the functioning of electronic circuits. However, this observation might mean that photonic hardware is just not well suited to classical digital methods and that a paradigm shift is needed in the way optics is used for computation. It is in this sense that we investigated neural networks as a new way to exploit optical effects in advanced signal processing applications. Many schemes have been proposed in the past to implement neural networks with optics [17, 18, 204, 219]. Most of them were either optoelectronic, free-space imaging or low-bandwidth solutions. Also, most of them did not properly exploit the advantages of optics for multiplexing [202]. To our knowledge, the scheme that we proposed here is simpler and better suited for useful signal processing operations at high bandwidth than the previous architectures.

Our device remains delicate to handle and the requirement of phase-locked inputs significantly restricts the domains of application. It is also subject to limitations in terms of the number of neurons possible, as discussed in section 6.6.1. For these reasons, the frequency multiplexed architecture proposed is realistically applicable in its present form only in some particular contexts. On a more general scale, neural networks have been proven to be theoretically able of universal computation. However, digital neural networks are used advantageously only in some specific domains. Similar restrictions should be expected for eventual future optical neural networks.

Conclusion and outlook

At the end of this three-years period of in-depth research, photonics appears to me as a still vibrant human-sized research area. Unlike some other branches of Physics, optics is still bursting with unexplored domains accessible to small teams of researchers with moderate resources. At the opposite, breakthroughs in some classical fields of fundamental research such as particle physics or astronomy presently require enormous infrastructures and involve collaboration between gigantic teams. Finding my place in scientific research as an individual has been a very rewarding experience, as well as being involved in the progress of a topic benefitting to the general public. Looking backwards, these are likely the major arguments that made this field so attractive to me at the time of my choice of a career orientation a few years ago.

This thesis took place at the interface between multiple fields and combined different concepts to develop novel applications. First of all, every result of this work relied on the **conjugate use of advanced linear and nonlinear optical processing techniques**. Fourier domain processing was a key operation at every stage, including dispersion emulation, spectral shaping of short pulses, delaying of pseudo independent signals, generation of hash codes and relative phase and amplitude control between channels. Kerr nonlinear effects were also omnipresent in signal monitoring, demultiplexing, signal comparison and design of transfer functions. Secondly, the targeted applications overlapped with both **information communication and computation**. The realization of a high bandwidth transmission link strongly relied on cuttingedge optical signal processing methods, while optical arithmetics and neural network logic required generation, conditioning and measurement of data with usual telecommunication methods. In these projects, communication and computation were even more strongly connected by the fact that similar techniques were used for both of them. Finally, this work was at the intersection of **novel platforms, principles and functionalities**. Fourier domain programmable optical processors and chalcogenide photonic chips enabled implementation of wavelength parallel operation of a dot-product to realize new functions such as low-latency optical error detection and neural network computing. The whole process of blending topics fits into a more general trend in science where many new applications arise from marrying different fields. Think of optofluidics, astrophysics, biophotonics, ...

The central motivations of the work are related to **latency**, **energy efficiency and bandwidth**. The evolution of the applications in information technology is shifting more and more towards a cloud approach where storage and processing are spread across the network. In this context, the delays due to transport between the data location, the computing nodes and the end-users directly impact the overall processing time. Some specific networks aimed at automated decisions have even more stringent latency requirements. On the other hand, the growth of information technologies is only sustainable if it is accompanied by a dramatic reduction of the energy intensity. Their environmental footprint is becoming non negligible on a global scale, with the consumption of a few percents of the total electricity produced. All optical networks could be advantageous to both aspects by removing the lag time of DSP and possibly being more power efficient in some cases.

In this discussion, comparison with existing data multiplexing and digital processing techniques is unavoidable. The functions realized here with optical systems already have well-known electronic implementations and development of optical equivalents is only relevant if it can provide a significant net advantage on the long term. As stated previously, this nurtures a long debate inside and outside the optics community. The comparison is necessarily speculative given the fact that transformative breakthroughs in optical technologies are expected in the future, particularly in nanophotonics.

7.1 Conclusion on high baudrate telecommunications

We saw that the use of light for information technology really started with the development of low loss optical fibres for telecommunication purposes. Since then, giant advances have been achieved and are still on-going. The first part of my work contributed to the investigation of ultra high baudrate signals for increasing the bandwidth of fibre links. The method which was used, optical time division multiplexing, was challenging at multiple levels. Initially, research on increasing the symbol rate was essentially motivated by the historical

observation that higher baudrate have always led to cost reduction in telecommunications. In the context of my interest for signal processing, I mainly saw advantages of single channel OTDM in terms of potential for high speed serial operations. With this approach, a single processing unit is needed, instead of multiplying the number of processors for treating every channel separately, which is the case of parallel architectures.

Chapter 4 started with an introduction to the techniques for generating OTDM signals in a laboratory and detecting them. Although we were not able to demonstrate error-free demultiplexing at 1.28 Tbaud by nonlinear sampling in a chalcogenide waveguide, other schemes have been successfully used in the literature [28, 75, 78]. Given that it is well known that such signals can be received, we executed further OTDM experiments based only on RF-spectrum, autocorrelation and eye diagram monitoring, without BER testing. Troubleshooting analysis of our setup came to the conclusion that excessive timing jitter from the laser sources was the principal origin of BER degradation, in spite of numerous attempts with different schemes.

The rest of the chapter introduced compensation techniques for dispersion of multiple orders as well as DGD, compatible with wide bandwidth signals. Feedback control from a RF-spectrum analyzer based signal monitor allowed controlling the compensators in closed loop for real time cancellation of fluctuations.

The difficulties experienced throughout these sets of experiments let me doubtful about a practical long-haul implementation of Terabaud serial data. The technique would be usable over moderate distances up to a few tens of kilometres if our automatic distortion compensation methods were applied to stabilize the transmission. Ultra high baudrate systems might be wanted for some applications where all-optical processing matters a lot. However, it is more likely that OTDM at lower rates such as 80 Gbaud or 160 Gbaud would be used in conjunction with WDM, AMF or OFDM. Pushing the clock rate over the Terabaud does not seem to fit in the industrial landscape in the near future as the difficult engineering work required to make it practical would incur prohibitive costs for today's networks.

7.2 Conclusion on analog optical computing

At their present state of development, it seems clear to me that optical technologies are not ready yet to implement digital processing with similar architectures as electronic systems. Spatial footprint, energy efficiency and difficulty to cascade optical nonlinearities would make attempts of building such alloptical processor unreasonable unless significant breakthroughs are made in integrated nonlinear optical components. This restricts the use of all-optical nonlinear processing to analog schemes, at least in a near future. In this last context, however, photonics can be very powerful. In chapters 5 and 6 we demonstrated how the parallelism of WDM optics can be exploited to carry out complex functions with minimal hardware. Analog systems are typically bound to the application that they have been designed for, although we have shown that the narrow range of use can be relaxed by exploiting the reconfigurability of Fourier domain optical processors. For instance, the field dot-product and hash code schemes introduced in this work both permit flexible use with a range of different baudrates, encoding formats and function configurations without physical change in the setup.

We pushed this approach one step further by investigating a novel paradigm for reconfigurable optical signal processing. Neural networks, by essence, are a generic model optimized to perform a specific task. The learning process adapts the system to optimally operate a function with the resources available. Given that the learning of a unique device can be done with training sequences corresponding to different tasks, it is a form of programmable computing. In theory, neural networks have been proven to be computationally complete, which means that they are in principle able to act as a Turing machine.

One remarkable property of these structures is that they do not critically depend on the exact action of their components, neither on their uniformity. This may enable exploiting optical nonlinearities in spite of their complex behaviour and inhomogeneous fabrication. From the beginning, however, analog processing has been suffering from the inherent issue of propagation and amplification of the internal noise of the components through the system, which strongly limits the possibility of cascading operations. Although internal noise remains a challenge in neural network architectures, these systems present an interesting robustness to input noise (the fluctuations in the input signal being processed) and generalization capabilities. These are also key properties in favor of this type of analog processing.

It would be extremely pretentious and illegitimate to claim that such system would revolutionize the use of optics in processing and open the door to optical CPUs. Firstly, we have shown the limits of our frequency multiplexed method in terms of bandwidth requirements and the degradation of OSNR occurring inside the dot-product for certain combinations of coefficients. This restricts the number of neurons that can be implemented onto a single device, as well as the range of coefficients accessible. Moreover, the phase-locking requirements pose a major issue for telecommunication applications with signals originating from different sources. Secondly, digital implementations of neural networks on classical electronic circuits are already well mastered. They have been proved to be incomparably efficient for certain kinds of applications, but their use in every domain is not straightforward. The learning mechanism remains a non trivial process that often involves know-how and trial-and-error search.

The optical neural network presented here could provide an interesting

all-optical platform for relatively high level tasks such as pattern recognition or decision logic that are not possible with a classical approach. The extreme simplicity of its physical implementation allows realization of circuits with high functionality using only the presently available technology of optical components. It could bridge the gap between DSP and the current simplistic all-optical schemes.

7.3 General outlook on optics in information science

As a general conclusion, I would like to consider the field of photonics in the global perspective of future information systems. Light greatly impacted long haul transmission systems with low loss and interference free transmission of dense data signals through optical fibres. In particular, this technology enabled the development of internet as a worldwide network. The impact on shorter scale interconnect nested inside the logic itself will likely be another revolution. Regarding computation, my feelings are more mitigated. The classical digital approach does not seem to fit well with the intrinsic laws of optics. Approaches such as analog computing and quantum computing might be better able to fully leverage the power of optics.

When considering photonics in active logic components, it is important to keep in mind that fundamental rules limit the use of light for highly integrated devices. The size of an elementary photonic gate is bound by the photon wavelength. Hence, the minimum possible footprint of an optical transistor has a typical length of the order of 1 μm . In comparison, the wavelength of a relativistic electron in electronic processors, is of the order of 1 nm. As a consequence, electronics has a higher potential of integration than optics. The CMOS industry is already integrating transistors with cell sizes well under 1 μm^2 . These considerations lead to believe that a large scale all-optical digital CPU is not realistic.

It is much more likely and, looking at the current evolution of the field, almost certain that future generations of computing chips will embed hybrid opto-electronic technologies. Energy consumption and heat dissipation are the main limiting factors of current CMOS chips and a large part of it is due to interconnects between internal components. Switching the state of an electric wire requires loading the corresponding capacitance, which can hardly be reduced. Here, light presents a clear advantage as a vector of information – again.

At a more speculative level, photonics may or may not lead to analog mass-computation systems. Many experimental and even commercial projects have attempted using free-space optics for fast parallel matrix operations, but without much success in real world systems up to now. Some researches are
also looking at photonics for high speed chaos-based computation, while others, as developed in this thesis, try operating simple optical functions at ultra high bandwidth and low latency. It would not be surprising to see analog optical processors executing specific tasks in domains where latency and bandwidth are key. Some specialized fields may look at exploiting different properties of optical circuits, such as their insensitivity to electromagnetic radiations in space science and extreme industrial conditions.

On a different aspect, the control of single photons starts enabling the use of light as vector of quantum information. Although many engineering challenges are still hindering large scale experiments, demonstration of principle of optical quantum gates has already been achieved. The bosonic nature of photons avoids the issue of decoherence experienced with most other quantum approaches and the fast improvement of single photon sources and handling methods promise more breakthroughs in a not-so-far future.

One sure thing, photonics is in the game more than ever.

Fourier domain programmable optical processors

Numerous optical signal processing operations use spectral filtering and phase control, classically implemented with devices such as grating filters, phase modulators or dispersion engineered components. For more flexibility and precise setting of the parameters, Fourier domain programmable optical processors (FD-POP) enable full control over the spectral phase and attenuation of the optical field.

The technique was introduced by A. M. Weiner in 1988 and first applied to spectral pulse shaping [104]. His setup was a symmetric arrangement comprising two diffraction gratings and a spatial light modulator (SLM) in the plane of symmetry of a 4-f imaging setup. The gratings were aligned to spread the input signal spectrally, then recombine it. Programming of a custom transmission and phase for different parts of the SLM allowed for spectral shaping and selective switching of the beam to multiple outputs.

A commercial device based on a similar working principle was used in our experiments, known under the name of Waveshaper, produced by Finisar. The SLM is replaced by a liquid crystal on silicon (LCoS) display reflecting the light with a programmable added relative phase. The phase pattern applied allows steering the reflected beam and defining its spectral phase.

Figure A.1 shows the working principle of this type of FD-POP. The incoming signal passes through optical elements splitting the two polarization states into separate beams of same polarization. They are reflected on a cylindrical mirror and sent to a conventional grating that spreads the beams spectrally. The optics is aligned so as to image them onto the LCoS display via the same



FIGURE A.1: Diagram of a FD-POP for an incoming signal polarized along a principal axis of the device.

cylindrical mirror. The beams reflected by the LCoS follow the same path backwards, are recombined by the grating and coupled to an output optical fibre.

Every pixel of the LCoS display matrix can be controlled to apply a custom phase offset comprised between 0 and 2π . Given that the signal is spread spectrally, every pixel column corresponds to a different frequency component, enabling wavelength selective operation. After calibration, the system can be programmed to precisely apply any spectral phase to the signal within the resolution limit of the device (10 GHz of frequency spacing). Application of a linear phase ramp also allows steering the beams towards a given output port. Direction of part of a signal to a *block* port enables the introduction of a controlled amount of attenuation.

We made an extensive use of FD-POP devices in our experiments to split, combine and apply complex spectral phase and attenuation patterns to optical signals. In particular, we used them to emulate and compensate for dispersion, introduce wavelength selective delays and create advanced interferometric circuits.

OTDM demultiplexing troubleshooting

In this appendix, we investigate the causes of the issues faced for demultiplexing terabaud OTDM signals. It relates to the setup described in section 4.1.3.

The major source of impairments for BER testing has been identified to be the fast timing jitter of the optical sources. Our reference results of 1.28 Tbaud to 10 Gbaud nonlinear demultiplexing [28] were obtained with both signal and pump pulse trains generated from a single Erbium-doped glass oscillator (ERGO) laser source delivering significantly better performances than our two separate modelocked fibre laser sources, resulting in a more regular spectrum and lower timing jitter.

The timing jitter of our sources was evaluated using root mean square (RMS) noise integration of the RF-spectrum measured after detection of their optical outputs [220, 221]. Based on the measurements on figure B.1, we found RMS jitter values of respectively $t_{\rm J,signal} = 78.8$ fs and $t_{\rm J,pump} = 117.0$ fs for the signal and pump sources. The electronic synchronization of the two modelocked lasers introduced an additional jitter $t_{\rm J,sync}$ evaluated to about 59.3 fs. Supposing a Gaussian distribution of the timing noise, combination of both sources for demultiplexing would result in an effective total jitter of

$$t_{\rm J,tot} = \sqrt{(t_{\rm J,signal})^2 + (t_{\rm J,pump})^2 + (t_{\rm J,sync})^2} = 153.0 \text{ fs}$$
 (B.1)

Compared with the pulse width of the OTDM channel of about 300 fs, the jitter was not negligible and may have significantly impaired the demultiplexing operation. No other successful attempt of demonstration of terabaud demultiplexing in chalcogenide chip with fibre laser sources is known to date¹.

¹Discussion with Dr. E. Palushani and Pr. L.K. Oxenløwe, who contributed to the benchmark results in reference [28], pointed out the fact that previous attempts at 1.28 Tbaud

The effect of the timing jitter became negligible with respect to the 510 fs pulse width of the 640 Gbaud signal.



FIGURE B.1: Left to right: timing jitter of the synchronization electronics, 40 GHz signal source and 10 GHz pump source calculated from the RMS noise figure with an electronic RF spectrum analyzer.

An additional argument confirming the hypothesis of the timing jitter issue was found by operating demultiplexing of a 640 Gbaud signal made of excessively short pulses (300 fs, comparable to the case at 1.28 Tbaud). The pump and signals, both supported by a spectrum wider than needed, had lower duty cycle and higher peak power and generated an idler of higher power level. Instead of improving the bit-error rate measured thanks to higher OSNR, this change degraded it. A likely explanation of this effect is that for such narrow pulses, the timing jitter became non-negligible regarding the pulse width and caused a decrease of the eye opening in the same manner as for the 1.28 Tbaud data.

Although error-free detection of a 640 Gbaud data could be achieved, the BER measurement could hardly have been performed in parallel with another operation such as automatic impairment compensation described later in the chapter. The brief stability period and high sensitivity of the setup to environmental fluctuations made impractical its parallel use with a system causing additional time variations of the signals.

B.1 Timing drift

On top of the jitter issues already mentioned, a slower scale timing drift effect also appeared when combining the two wide bandwidth signals. The environmental temperature fluctuations induced slight changes in the path length that were sufficient to desynchronize signal and pump over a time scale of the order of a second. At 1.28 Tbaud, the length of a symbol in a silica fibre is of 153.4 μ m. The pump and signal paths both counted several hundreds of metres of optical fibre, mostly SMF, HNLF and Erbium doped fibre in the

using modelocked fibre lasers were all unsuccessful. Error-free operation could only be achieved with the introduction of an ERGO source laser.



FIGURE B.2: Demultiplexing of 640 Gbaud (red) and 1.28 Tbaud (green) signals. Left: Optical spectrum at the output of the chip waveguide. Middle: autocorrelation traces of the pump and signals. Right: Eye diagram of the extracted idlers. Demultiplexing at 1.28 Tbaud is not error free, arguably due to jitter.

EDFA. These fibres, exposed to ambient fluctuations, varied of more than a symbol length within a few seconds.

In the laboratory conditions, the thermal drifts were mainly due to the air conditioning installation which caused oscillations of about 1°C. After deactivation of the air conditioning and a settling time of a few hours, the period of stability of the demultiplexing experiment was improved by approximately a factor 10.

In these conditions, transmission through a link of several kilometres would not have been realistically feasible without a fast delay tracking system or transmission of the clock along with the data [94]. However, in a real world system, the pump pulse source would likely be synchronized by clock recovery of the data channel, which would remove the issue of the time drift [79].

OTDM setup evolution

The realization of the OTDM test bench required multiple changes in the setup. The final configuration presented in chapter 4 is the outcome of three years of optimization. Here, the main steps of evolution are cited, essentially focused on optimizing the demultiplexing scheme.

C.1 40 GHz demultiplexing pump

The initial setup for nonlinear demultiplexing of the OTDM signal used a pump pulse train at a repetition rate of 40 GHz, generated by a 40 GHz modelocked fibre laser. The high repetition rate of the pump presented the inconvenient of reducing the peak power of the pump pulses for a given average power, limited by the power handling of the chalcogenide photonic chip. The efficiency of the FWM frequency conversion process depending on the peak power, the generated idler signal was weaker than the one obtained with a 10 GHz demultiplexing scheme. The OSNR of the demultiplexed signal after the filtering and amplification stages was excessively degraded and we gave up this method.

C.2 10 GHz demultiplexing pump from semiconductor laser source

The first 10 Ghz demultiplexing setup used a femtosecond semiconductor modelocked laser as pulse source for the pump channel. This choice was motivated by the good pulse quality and stability expected from this type of lasers. However, after a long troubleshooting period, the source was found to exhibit an excessive level of timing jitter, measured at 290 fs (see figure C.1) to 530 fs (according to the locking parameters) by the technique introduced in appendix B. Compared to the pulse width of the terabaud OTDM signal of 275 fs and the duty cycle of 781 fs, this jitter was unacceptable and caused dramatic closing of the eye diagram of the received idler.

In a second step, the pump source was changed to a 10 GHz fibre modelocked laser with a RMS timing jitter of 78 fs, as shown in appendix B.



10 GHz semiconductor laser source

FIGURE C.1: RMS jitter measurement of the 10 GHz semiconductor modelocked laser source.

C.3 Chip sample changes and damages in chalcogenide waveguides

Numerous changes in waveguides and chip samples have been made to select the best working ones and avoid the degraded ones. In section 3.5.2, we have mentioned that excessive optical power and prolonged exposure to ambient light could cause damages to the chalcogenide chip samples. This has been the origin of many issues and degradations of performance, requiring to regularly and carefully select the samples used.

The optical power threshold is an ambiguous parameter to take into account. A first discrimination has to be made between instantaneous and average power. The waveguides employed in our work could handle a peak power up to about 5 W, while the absolute average power maximum was about ten times lower.

On top of these, we observed other factors of degradation occurring at a slower timescale. Injection of a strong light beam of average power below the high threshold would not affect the ChG chip for short exposure times. However, the propagation loss was measured to slowly increase by up to 4 dB after continuous run for periods of 10 minutes to 1 hour. After resting in a dark enclosure for a couple of days, the waveguides eventually recovered to almost their initial specifications.

The transmission spectrum of the samples after exposure to intense light occasionally presented periodical structures typical of an interference process. The spectral patterns remained after the exposure was interrupted and they turned out to be either permanent or temporarily lasting for a day, according to the level of degradation. The physical origin of this effect is understood as the establishment of standing wavefronts along the waveguides due to internal back reflections at the input and output facets of the chip. The periodic maxima of power would literally write Bragg gratings by photosensitive alteration of the material. These last two effects appeared to be more severe and kicking in at lower average power in waveguides that were not anti-reflection coated. It is deduced that the absence of coating was causing a bigger part of the input light to be reflected back-and-forth inside the chips, establishing standing waves forming grating structures.

Photosensitivity of the ChG compounds can also induce deterioration of samples at rest if they are exposed to ambient light for a long period. This has been observed on chips left unprotected on coupling stages, while those that were kept in the dark when unused did not decay.

Finally we observed an irreversible destructive change in two chip samples old of a couple of years. The insertion loss abruptly increased by about 10 dB in all the waveguides and unusual features appeared on the transmission spectrum. After investigation, the origin of the issue was found to be a partial delamination of the polymer cladding covering the chips.



FIGURE C.2: Delamination and grating induced damages in a chalcogenide chip. Top: ASE noise source input into waveguides. Bottom: high insertion loss and grating patterns appear on the transmitted spectrum after two different waveguides (left: WG2_2, right: WG2_5).

D

A well known issue in quantized electronic circuits is the apparition of glitches or "ghost states" in the output of logic functions. This undetermined transient behaviour can be detrimental to the transmission and processing of information. It occurs with signals in any kind of information system and is well understood.

We consider the example of an XOR gate taking two input signals A and B which have relatively long rise and fall times t_r and t_f . The output of the XOR gate should be 0 when A and B are both in the same state ([00] or [11]) and 1 when they are not ([01] or [10]). On the transition from the first to the second bit on the state diagram of figure D.1, A and B keep opposite values, so that the output of the XOR should steadily remain at 1. However, both inputs change during the transition and pass by the middle level 0.5 at the same time. As A and B have same state at that precise moment, the XOR gate briefly outputs the misleading value 0.

The eye diagram on the right of the figure was produced in the same fashion as the optoelectronic XOR gate of section 6.4.2, but with the pulse train source replaced with a CW laser. Due to this change, the data was encoded with a non return-to-zero (NRZ) format causing glitches to occur at the transient states. The resulting eye diagram presented many irregularities.

Several solutions can get around the problem of glitches. A first technique consists in designing the system to avoid multiple signals to change state simultaneously. It is for instance the case with *Gray code*, or *reflected binary code*, widely used in error correction methods for telecommunications, position encoders, and many other domains. However, the use of such coding formats is not always possible or optimal.



FIGURE D.1: Apparition of glitches in an XOR gate. The orange arrows point at the ghost states. Right: eye diagram of an XOR output of NRZ data.

Current synchronous digital logic circuits are based on a distributed clock for gating the successive states of all their elements. At every period of the clock, a settling time allows for all the values to stabilize before they are taken into account in the next operation.

In our logic experiments, we used pulsed RZ signals, which is similar to a clocked circuit. The short optical pulses sampled the states of the slowly varying electronic components and quasi conservation of the pulse shape along the paths allowed us to almost perfectly match the respective symbols when an operation was made between several signals.

Transfer function of the difference of two neurons

This appendix presents the measurement of the transfer function of an ONN with two neurons subtracted from each other. The nonlinearity used is based on degenerate FWM and the measurement process is the same as in section 6.5.1.

The theoretical behaviour in this setting is described by equation (E.1).

$$y = |tanh(S_1) - tanh(S_2)| \tag{E.1}$$

The data are presented below in figure E.1 in the same fashion as in section 6.5.1. The symmetry regarding the two inputs is conserved.



FIGURE E.1: Transfer function with two neurons in subtraction configuration, measured experimentally (top) and theoretical simulation(bottom). The color maps in the middle represent the output power as a function of the intensity of the two neurons. Onedimensional transfer profiles sliced according to the axes are featured in the left and right series of graphs.

Acronyms

\mathbf{AMF}	advanced modulation formats
ASE	amplified spontaneous emission
ASK	amplitude shift keying
ADC	analog-to-digital converter
\mathbf{ANU}	Australian National University
B2B	back-to-back
BPSK	binary phase shift keying
\mathbf{BER}	bit-error rate
BPF	bandpass filter
ChG	chalcogenide
\mathbf{CW}	continuous wave
\mathbf{XPM}	cross phase modulation
$\mathbf{D}\mathbf{G}\mathbf{D}$	differential group delay
DCF	dispersion compensating fibre
\mathbf{ETDM}	electronic time division multiplexing
EDFA	erbium doped fibre amplifiers
ERGO	erbium-doped glass oscillator
FD-POP	fourier domain programmable optical processor
FFNN	feed-forward neural network
\mathbf{FPGA}	field-programmable gate array
\mathbf{FWM}	four wave mixing
FCA	free carrier absorption
\mathbf{FWHM}	full width at half maximum
HNLF	highly nonlinear fibre
\mathbf{LCoS}	liquid crystal on silicon
\mathbf{LSB}	least significant bit
$\mathbf{M}\mathbf{Z}\mathbf{M}$	Mach-Zehnder modulator
\mathbf{MSB}	most significant bit
NOLM	nonlinear optical loop mirror
OADM	optical add-drop multiplexer

OFDM	optical frequency division multiplexing
ONN	optical neural network
OPC	optical phase conjugation
OSNR	optical signal to noise ratio
OSA	optical spectrum analyzer
OTDM	optical time division multiplexing
\mathbf{PDE}	partial differential equation
\mathbf{PS}	phase shifter
PSK	phase shift keying
\mathbf{PhC}	photonic crystal
POTS	plain old telephone service
\mathbf{PM}	phase modulator
\mathbf{PC}	polarization controller
\mathbf{PMD}	polarization mode dispersion
\mathbf{PBS}	polarizing beam splitter
\mathbf{QAM}	quadrature and amplitude modulation
QPSK	quadrature phase shift keying
\mathbf{RFSA}	RF-spectrum analyzer
\mathbf{RMS}	root mean square
\mathbf{SPM}	self phase modulation
SOA	semiconductor optical amplifiers
\mathbf{SUT}	signal under test
\mathbf{SNR}	signal-to-noise ratio
SVEA	slowly varying envelope approximation
\mathbf{SLM}	spatial light modulator
TPA	two-photon absorption
\mathbf{WDM}	wavelength division multiplexing
\mathbf{ZDW}	zero dispersion wavelength

References

- [1] A. Pope, *Homer's Iliad* (Wildside Press LLC, 2008), 18th ed.
- [2] H. A. Giles, *The Civilization of China* (Gutenberg Project, 2006).
- [3] Aeschylus, Agamemnon (Cambridge University Press, 2004).
- [4] C. H. Sterling, Military Communications: From Ancient Times to the 21st Century (ABC-CLIO, 2008).
- [5] T. Standage, The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers (Bloomsbury Publishing, 2009).
- [6] A. R. Michaelis, *From semaphore to satellite* (International Telecommunication Union, 1965).
- [7] I. Kaminow and T. Li, Optical Fiber Telecommunications IV-B: Systems and Impairments, no. v. 2 in Optics and Photonics (Elsevier Science, 2002).
- [8] J. Hecht, *City of Light: The Story of Fiber Optics* (Oxford University Press, 2004).
- [9] K. Kao and G. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," Proceedings of Electrical Engineers pp. 1151–1158 (1966).
- [10] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications." Nature 411, 1027–30 (2001).
- [11] A. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear Shannon limit," Journal of Lightwave Technology 28, 423–433 (2010).
- [12] W. Aspray, A. Bromley, M. Campbell-Kelly, P. Ceruzzi, and M. Williams, *Computing Before Computers* (Iowa State University Press, 1990).
- [13] G. Ifrah, The Universal History of Computing: From the Abacus to Quantum Computing (John Wiley & Sons, Inc., 2000).

- [14] M. S. Mahoney, "The History of Computing in the History of Technology," IEEE Annals of the History of Computing 10, 113–125 (1988).
- [15] R. Tucker, K. Hinton, and G. Raskutti, "Energy consumption limits in high-speed optical and electronic signal processing," Electronics Letters 43, 906 (2007).
- [16] R. S. Tucker and K. Hinton, "Energy Consumption and Energy Density in Optical and Electronic Signal Processing," IEEE Photonics Journal 3, 821–833 (2011).
- [17] H. Caulfield, J. Kinser, and S. Rogers, "Optical neural networks," (1989).
- [18] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," Scientific reports 2, 287 (2012).
- [19] P. Winzer and R.-J. Essiambre, "Advanced Optical Modulation Formats," Proceedings of the IEEE 94, 952–985 (2006).
- [20] D. Qian, M.-F. Huang, E. Ip, Y.-K. Huang, Y. Shao, J. Hu, and T. Wang, "101.7-Tb/s (370294-Gb/s) PDM-128QAM-OFDM Transmission over 355-km SSMF using Pilot-based Phase Noise Mitigation," in "Optical Fiber Communication Conference," (Optical Society of America, 2011), p. PDPB5.
- [21] S. Randel, R. Ryf, A. Sierra, P. J. Winzer, A. H. Gnauck, C. A. Bolle, R.-J. Essiambre, D. W. Peckham, A. McCurdy, and R. Lingle, "656-Gb/s mode-division multiplexed transmission over 33-km few-mode fiber enabled by 66 MIMO equalization." Optics express 19, 16697–707 (2011).
- [22] J. Sakaguchi, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, T. Hayashi, T. Taru, T. Kobayashi, and M. Watanabe, "109-Tb/s (797172-Gb/s SDM/WDM/PDM) QPSK transmission through 16.8-km homogeneous multi-core fiber," (2011).
- [23] N. Mac Suibhne, Z. Li, B. Baeuerle, J. Zhao, J. P. Wooler, S.-u. Alam, F. Poletti, M. N. Petrovich, A. Heidt, I. Giles, D. J. Giles, B. Pálsdóttir, L. Grüner-Nielsen, R. Phelan, J. O'Carroll, B. Kelly, D. Murphy, A. Ellis, D. Richardson, and F. C. Garcia Gunning, "Wavelength Division Multiplexing at 2μm," in "European Conference and Exhibition on Optical Communication," (OSA, Washington, D.C., 2012), p. Th.3.A.3.
- [24] F. Gholami, S. Zlatanovic, E. Myslivets, S. Moro, B. P.-P. Kuo, C.-S. Brès, A. O. J. Wiberg, N. Alic, and S. Radic, "10Gbps Parametric Short-Wave Infrared Transmitter," in "Optical Fiber Communication

Conference/National Fiber Optic Engineers Conference 2011," (OSA, Washington, D.C., 2011), p. OThC6.

- [25] C. M. Smith, N. Venkataraman, M. T. Gallagher, D. Müller, J. A. West, N. F. Borrelli, D. C. Allan, and K. W. Koch, "Low-loss hollow-core silica/air photonic bandgap fibre." Nature 424, 657–9 (2003).
- [26] M. Nakazawa, E. Yoshida, T. Yamamoto, E. Yamada, and A. Sahara, "TDM single channel 640 Gbit/s transmission experiment over 60 km using 400 fs pulse train and walk-off free, dispersion flattened nonlinear optical loop mirror," Electronics Letters 34, 907 (1998).
- [27] J. Schröder, T. D. Vo, Y. Paquot, and B. J. Eggleton, "Breaking the Tbit/s Barrier: Higher Bandwidth Optical Processing," Optics and Photonics News 23, 32 (2012).
- [28] T. D. Vo, H. Hu, M. Galili, E. Palushani, J. Xu, L. K. Oxenløwe, S. J. Madden, D.-Y. Choi, D. A. P. Bulla, M. D. Pelusi, J. Schröder, B. Luther-Davies, and B. J. Eggleton, "Photonic chip based transmitter optimization and receiver demultiplexing of a 1.28 Tbit/s OTDM signal." Optics express 18, 17252–61 (2010).
- [29] A. Sawchuk and T. Strand, "Digital optical computing," Proceedings of the IEEE 72, 758–779 (1984).
- [30] R. A. Athale, "Optical matrix algebraic processors: a survey," in "10th International Optical Computing Conference,", S. Horvitz, ed. (International Society for Optics and Photonics, 1983), pp. 24–31.
- [31] S. Watanabe, F. Futami, R. Okabe, Y. Takita, S. Ferber, R. Ludwig, C. Schubert, C. Schmidt, and H. G. Weber, "160 Gbit/s optical 3Rregenerator in a fiber transmission experiment," (2003).
- [32] S. Watanabe and M. Shirasaki, "Exact compensation for both chromatic dispersion and Kerr effect in a transmission fiber using optical phase conjugation," Journal of Lightwave Technology 14, 243–248 (1996).
- [33] L. Lei, J. Dong, Y. Zhang, H. He, Y. Yu, and X. Zhang, "Reconfigurable photonic full-adder and full-subtractor based on three-input XOR gate and logic minterms," Electronics Letters 48, 399 (2012).
- [34] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, "Plasmonics for extreme light concentration and manipulation." Nature materials 9, 193–204 (2010).

- [35] D. E. Tamir, N. T. Shaked, P. J. Wilson, and S. Dolev, "High-speed and low-power electro-optical DSP coprocessor," Journal of the Optical Society of America A 26, 11–20 (2009).
- [36] G. Kurian, J. E. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, L. C. Kimerling, and A. Agarwal, "ATAC," in "Proceedings of the 19th international conference on Parallel architectures and compilation techniques PACT '10," (ACM Press, New York, New York, USA, 2010), p. 477.
- [37] J. V. Stone, Information Theory: A Tutorial Introduction (Sebtel Press, 2014).
- [38] H. Nyquist, "Certain topics in telegraph transmission theory," American Institute of Electrical Engineers, Transactions of the **47**, 617–644 (1928).
- [39] C. E. Shannon, "A mathematical theory of communication, Bell Telephone System Tech," Publ. Monograph B-1598 pp. 1–80 (1948).
- [40] D. J. C. MacKay, Information Theory, Inference and Learning Algorithms (Cambridge University Press, 2003).
- [41] G. Forney, T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," IEEE Communications Letters 5, 58–60 (2001).
- [42] I. P. Kaminow, T. Li, and A. E. Willner, Optical Fiber Telecommunications V A: Components and Subsystems, Volume 1 (Academic Press, 2008).
- [43] P. J. Winzer and R.-J. Essiambre, "Advanced Modulation Formats for High-Capacity Optical Transport Networks," Journal of Lightwave Technology 24, 4711–4728 (2006).
- [44] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," Journal of Lightwave Technology 28, 662–701 (2010).
- [45] J. Tang, "A Comparison Study of the Shannon Channel Capacity of Various Nonlinear Optical Fibers," J. Lightwave Technol. 24, 2070– (2006).
- [46] R. Essiambre and G. Kramer, "Capacity limits of optical fiber networks," Journal of Lightwave Technology 28, 662–701 (2010).
- [47] X. Liu, A. R. Chraplyvy, P. J. Winzer, R. W. Tkach, and S. Chandrasekhar, "Phase-conjugated twin waves for communication beyond the Kerr nonlinearity limit," Nature Photonics 7, 560–568 (2013).

- [48] H. Takara, A. Sano, T. Kobayashi, H. Kubota, H. Kawakami, A. Matsuura, Y. Miyamoto, Y. Abe, H. Ono, K. Shikama, Y. Goto, K. Tsujikawa, Y. Sasaki, I. Ishida, K. Takenaga, S. Matsuo, K. Saitoh, M. Koshiba, and T. Morioka, "1.01-Pb/s (12 SDM/222 WDM/456 Gb/s) Crosstalk-managed Transmission with 91.4-b/s/Hz Aggregate Spectral Efficiency," in "European Conference and Exhibition on Optical Communication," (OSA, Washington, D.C., 2012), p. Th.3.C.1.
- [49] Cisco, "The Zettabyte Era," Tech. rep. (2012).
- [50] D. Schneider, "Trading at the speed of light," IEEE Spectrum 48, 11–12 (2011).
- [51] J. Hasbrouck and G. Saar, "Low-latency trading," Journal of Financial Markets (2013).
- [52] N. V. Wheeler, M. N. Petrovich, R. Slavik, N. K. Baddela, E. R. Numkam Fokoua, J. R. Hayes, D. Gray, F. Poletti, and D. Richardson, "Widebandwidth, low-loss, 19-cell hollow core photonic band gap fiber and its potential for low latency data transmission," in "National Fiber Optic Engineers Conference," (OSA, Washington, D.C., 2012), p. PDP5A.2.
- [53] L. Schares, D. M. Kuchta, and A. F. Benner, "Optics in Future Data Center Networks," in "2010 18th IEEE Symposium on High Performance Interconnects," (IEEE, 2010), pp. 104–108.
- [54] E. G. Sauter, *Nonlinear Optics* (John Wiley & Sons, 1996).
- [55] P. Franken, A. Hill, C. Peters, and G. Weinreich, "Generation of optical harmonics," Physical Review Letters 7, 118–119 (1961).
- [56] J. D. Reis and A. L. Teixeira, "Unveiling nonlinear effects in dense coherent optical WDM systems with Volterra series." Optics express 18, 8660–70 (2010).
- [57] G. P. Agrawal, Nonlinear fiber optics (Academic Press, 2001).
- [58] T. Inoue and S. Namiki, "Pulse compression techniques using highly nonlinear fibers," Laser & Photonics Review 2, 83–99 (2008).
- [59] M. D. Pelusi and B. J. Eggleton, "Optically tunable compensation of nonlinear signal distortion in optical fiber by end-span optical phase conjugation." Optics express 20, 8015–23 (2012).
- [60] D. Wang, T.-H. Cheng, Y.-K. Yeo, Z. Xu, Y. Wang, G. Xiao, and J. Liu, "Performance Comparison of Using SOA and HNLF as FWM Medium in a Wavelength Multicasting Scheme With Reduced Polarization Sensitivity," Journal of Lightwave Technology 28, 3497–3505 (2010).

- [61] B. Eggleton, B. Luther-Davies, and K. Richardson, "Chalcogenide photonics," Nature Photonics 5, 141–148 (2011).
- [62] T. Simoyama, H. Kuwatsuka, B. Little, M. Matsuda, Y. Kotaki, and H. Ishikawa, "High-efficiency wavelength conversion using FWM in an SOA integrated DFB laser," IEEE Photonics Technology Letters 12, 31–33 (2000).
- [63] C. Koos, P. Vorreau, P. Dumon, R. Baets, B. Esembeson, I. Biaggio, T. Michinobu, F. Diederich, W. Freude, and J. Leuthold, "Highly-Nonlinear Silicon Photonic Slot Waveguide - OSA Technical Digest (CD)," in "Optical Fiber Communication Conference/National Fiber Optic Engineers Conference," (Optical Society of America, San Diego, California, 2008), p. PDP25.
- [64] A. C. Turner, C. Manolatou, B. S. Schmidt, M. Lipson, M. A. Foster, J. E. Sharping, and A. L. Gaeta, "Tailored anomalous group-velocity dispersion in silicon channel waveguides," Optics Express 14, 4357 (2006).
- [65] C. Monat, B. Corcoran, M. Ebnali-Heidari, C. Grillet, B. J. Eggleton, T. P. White, L. O'Faolain, and T. F. Krauss, "Slow light enhancement of nonlinear effects in silicon engineered photonic crystal waveguides," Optics Express 17, 2944 (2009).
- [66] L. O'Faolain, X. Yuan, D. McIntyre, S. Thoms, H. Chong, R. De La Rue, and T. Krauss, "Low-loss propagation in photonic crystal waveguides," Electronics Letters 42, 1454 (2006).
- [67] M. Pelusi, F. Luan, T. D. Vo, M. R. E. Lamont, S. J. Madden, D. A. Bulla, D.-Y. Choi, B. Luther-Davies, and B. J. Eggleton, "Photonic-chip-based radio-frequency spectrum analyser with terahertz bandwidth," Nature Photonics 3, 139–143 (2009).
- [68] F. Luan, M. Pelusi, M. R. E. Lamont, D.-Y. Choi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Dispersion engineered As2S3 planar waveguides for broadband four-wave mixing based wavelength conversion of 40 Gb/s signals," Optics Express 17, 3514–3520 (2009).
- [69] A. Gnauck, G. Charlet, P. Tran, P. Winzer, C. Doerr, J. Centanni, E. Burrows, T. Kawanishi, T. Sakamoto, and K. Higuma, "25.6-Tb/s WDM Transmission of Polarization-Multiplexed RZ-DQPSK Signals," Journal of Lightwave Technology 26, 79–84 (2008).
- [70] P. Winzer and R. Essiambre, "Advanced optical modulation formats," Proceedings of the IEEE **94**, 952–985 (2006).

- [71] Ivan Kaminow, T. Li, and A. E. Willner, Optical Fiber Telecommunications Volume VIA: Components and Subsystems (Academic Press, 2013), academic p ed.
- [72] Z. Zhu, M. Funabashi, Z. Pan, L. Paraschis, and S. Yoo, "10 000-hop cascaded in-line all-optical 3R regeneration to achieve 1 250 000-km 10-Gb/s transmission," IEEE Photonics Technology Letters 18, 718–720 (2006).
- [73] A. Bogoni, X. Wu, Z. Bakhtiari, S. Nuccio, and A. E. Willner, "640 Gbits/s photonic logic gates." Optics letters 35, 3955–7 (2010).
- [74] H. Hansen Mulvad, L. Oxenløwe, M. Galili, A. Clausen, L. Gruner-Nielsen, and P. Jeppesen, "1.28 Tbit/s single-polarisation serial OOK optical data generation and demultiplexing," Electronics Letters 45, 280–281 (2009).
- [75] T. Richter, E. Palushani, C. Schmidt-Langhorst, M. Nölle, R. Ludwig, and C. Schubert, "Single Wavelength Channel 10.2 Tb/s TDM-Data Capacity using 16-QAM and coherent detection - OSA Technical Digest (CD)," in "Optical Fiber Communication Conference," (Optical Society of America, 2011), p. PDPA9.
- [76] J. Van Erps, J. Schroeder, T. Vo, M. Pelusi, S. Madden, D. Choi, D. Bulla, B. Luther-Davies, and B. Eggleton, "Automatic dispersion compensation for 1 . 28Tb/s OTDM signal transmission using photonicchip-based dispersion monitoring," Optics Express 18, 25415–25421 (2010).
- [77] P. Lazaridis, G. Debarge, and P. Gallion, "Time-bandwidth product of chirped sech² pulses: application to phase-amplitude-coupling factor measurement," Optics Letters 20, 1160 (1995).
- [78] E. Palushani, H. C. H. Mulvad, M. Galili, L. K. Oxenløwe, A. T. Clausen, and P. Jeppesen, "OTDM-to-WDM Conversion Based on Time-to-Frequency Mapping by Time-Domain Optical Fourier Transformation," IEEE Journal of Selected Topics in Quantum Electronics 18, 681–688 (2012).
- [79] E. Tangdiongga, H. C. H. Mulvad, H. de Waardt, G. D. Khoe, A. M. J. Koonen, and H. J. S. Dorren, "SOA-based Clock Recovery and Demultiplexing in a Lab Trial of 640- Gb/s OTDM Transmission over 50-km Fibre Link," (2007).
- [80] J. H. Lee, S. Ohara, T. Nagashima, T. Hasegawa, N. Sugimoto, K. Igarashi, K. Katoh, and K. Kikuchi, "Clock recovery and demultiplexing of high-speed OTDM signal through combined use of bismuth oxide

nonlinear fiber and erbium-doped bismuth oxide fiber," IEEE Photonics Technology Letters **17**, 2658–2660 (2005).

- [81] T. Yamamoto, L. Oxenløwe, C. Schmidt, C. Schubert, E. Hilliger, U. Feiste, J. Berger, R. Ludwig, and H. Weber, "Clock recovery from 160 Gbit/s data signals using phase-locked loop with interferometric optical switch based on semiconductor optical amplifier," Electronics Letters 37, 509 (2001).
- [82] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, and A. E. Willner, "Optical Performance Monitoring," J. Lightwave Technol. 22, 294– (2004).
- [83] I. Shake, W. Takara, S. Kawanishi, and Y. Yamabayashi, "Optical signal quality monitoring method based on optical sampling," Electronics Letters 34, 2152–2154 (1998).
- [84] J. Van Erps, F. Luan, M. Pelusi, T. Iredale, S. Madden, D. Choi, D. Bulla, B. Luther-Davies, H. Thienpont, and B. Eggleton, "High-Resolution Optical Sampling of 640-Gb/s Data Using Four-Wave Mixing in Dispersion-Engineered Highly Nonlinear As2S3 Planar Waveguides," Journal of Lightwave Technology 28, 209–215 (2010).
- [85] J. Curtis and J. Carroll, "Autocorrelation systems for the measurement of picosecond pulses from injection lasers," International journal of electronics 60, 87–111 (1986).
- [86] P. Westbrook, S. Hunsche, G. Raybon, T. Her, and B. Eggleton, "Measurement of pulse degradation using all-optical 2R regenerator," Electronics Letters 38, 1193 (2002).
- [87] P. Westbrook, B. Eggleton, G. Raybon, and S. Hunsche, "Measurement of residual chromatic dispersion of a 40-Gb/s RZ signal via spectral broadening," IEEE Photonics Technology Letters 14, 346–348 (2002).
- [88] Y. K. Lizé, L. Christen, J.-Y. Yang, P. Saghari, S. Nuccio, A. E. Willner, and R. Kashyap, "Independent and simultaneous monitoring of chromatic and polarization-mode dispersion in OOK and DPSK transmission," IEEE Photonics Technology Letters 19, 2006–2008 (2007).
- [89] C. Dorrer and D. Maywar, "RF Spectrum Analysis of Optical Signals Using Nonlinear Optics," Journal of Lightwave Technology 22, 266–274 (2004).

- [90] T. D. Vo, J. Schröder, M. D. Pelusi, S. J. Madden, D.-Y. Choi, D. A. P. Bulla, B. Luther-Davies, and B. J. Eggleton, "Photonic Chip-Based Simultaneous Multi-Impairment Monitoring for Phase-Modulated Optical Signals," Journal of Lightwave Technology 28, 3176–3183 (2010).
- [91] D.-Y. Choi, S. Madden, D. A. Bulla, R. Wang, A. Rode, and B. Luther-Davies, "Submicrometer-Thick Low-Loss As2S3 Planar Waveguides for Nonlinear Optical Devices," IEEE Photonics Technology Letters 22, 495–497 (2010).
- [92] T. D. Vo, M. D. Pelusi, J. Schröder, F. Luan, S. J. Madden, D.-Y. Choi, D. A. P. Bulla, B. Luther-Davies, and B. J. Eggleton, "Simultaneous multi-impairment monitoring of 640 Gb/s signals using photonic chip based RF spectrum analyzer." Optics express 18, 3938–45 (2010).
- [93] C. Poole, J. Wiesenfeld, D. Digiovanni, and A. Vengsarkar, "Optical fiber-based dispersion compensation using higher order modes near cutoff," Journal of Lightwave Technology 12, 1746–1758 (1994).
- [94] M. Nakazawa, T. Yamamoto, and K. R. Tamura, "1.28Tbit/s 70 km OTDM transmission using third- and fourth-order simultaneous dispersion compensation with a phase modulator," Electronics Letters 36, 2027–2029 (2000).
- [95] B. Eggleton, B. Mikkelsen, G. Raybon, A. Ahuja, J. Rogers, P. Westbrook, T. Nielsen, S. Stulz, and K. Dreyer, "Tunable dispersion compensation in a 160-Gb/s TDM system by a voltage controlled chirped fiber Bragg grating," Photonics Technology Letters, IEEE 12, 1022–1024 (2000).
- [96] M. Durkin, M. Ibsen, M. Cole, and R. Laming, "1 m long continuouslywritten fibre Bragg gratings for combined second-and third-order dispersion compensation," Electronics Letters 33, 1891–1893 (1997).
- [97] S. Wielandy, P. Westbrook, M. Fishteyn, P. Reyes, W. Schairer, H. Rohde, and G. Lehmann, "Demonstration of automatic dispersion control for 160 Gbits transmission over 275 km of deployed fibre," Electronics Letters 40, 690 (2004).
- [98] D. T. Neilson, R. Ryf, F. Pardo, V. A. Aksyuk, M.-E. Simon, D. O. Lopez, D. M. Marom, and S. Chandrasekhar, "MEMS-Based Channelized Dispersion Compensator With Flat Passbands," Journal of Lightwave Technology 22, 101– (2004).
- [99] G.-H. Lee, S. Xiao, and A. Weiner, "Optical Dispersion Compensator With 4000-ps/nm Tuning Range Using a Virtually Imaged Phased Array

(VIPA) and Spatial Light Modulator (SLM)," IEEE Photonics Technology Letters 18, 1819–1821 (2006).

- [100] T. Kurosu, K. Tanizawa, S. Petit, and S. Namiki, "Parametric tunable dispersion compensation for the transmission of sub-picosecond pulses," Optics Express 19, 15549–15559 (2011).
- [101] T. Kato, Y. Koyano, and M. Nishimura, "Temperature dependence of chromatic dispersion in various types of optical fiber," Optics Letters 25, 1156 (2000).
- [102] M. Hamp, J. Wright, M. Hubbard, and B. Brimacombe, "Investigation into the temperature dependence of chromatic dispersion in optical fiber," IEEE Photonics Technology Letters 14, 1524–1526 (2002).
- [103] M. A. F. Roelens, S. Frisken, J. A. Bolger, D. Abakoumov, G. Baxter, S. Poole, and B. J. Eggleton, "Dispersion trimming in a reconfigurable wavelength selective switch," Journal of Lightwave Technology 26, 73–78 (2008).
- [104] A. M. Weiner, J. P. Heritage, and E. M. Kirschner, "High-resolution femtosecond pulse shaping," Journal of the Optical Society of America B 5, 1563 (1988).
- [105] J. Gordon, "PMD fundamentals: Polarization mode dispersion in optical fibers," Proceedings of the National 97, 4541–50 (2000).
- [106] M. Wegmuller, S. Demma, C. Vinegoni, and N. Gisin, "Emulator of first- and second-order polarization-mode dispersion," IEEE Photonics Technology Letters 14, 630–632 (2002).
- [107] H. Sunnerud, C. Xie, M. Karlsson, R. Samuelsson, and P. A. Andrekson, "A comparison between different PMD compensation techniques," Journal of Lightwave Technology 20, 368–378 (2002).
- [108] L. M. Adams and J. L. Nazareth, Linear and Nonlinear Conjugate Gradient-related Methods (SIAM, 1996).
- [109] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, 2007).
- [110] F. Heismann, D. Fishman, and D. Wilson, "Automatic compensation of first order polarization mode dispersion in a 10 Gb/s transmission system," in "24th European Conference on Optical Communication. ECOC '98 (IEEE Cat. No.98TH8398),", vol. 1 (1998), vol. 1, pp. 529–530.

- [111] M. Karlsson, J. Brentel, and P. A. Andrekson, "Long-term measurement of PMD and polarization drift in installed fibers," Journal of Lightwave Technology 18, 941–951 (2000).
- [112] T. Takahashi, T. Imai, and M. Aiki, "Automatic compensation technique for timewise fluctuating polarisation mode dispersion in in-line amplifier systems," Electronics Letters **30**, 348 (1994).
- [113] H. Sunnerud, M. Westlund, J. Li, J. Hansryd, M. Karlsson, P.-O. Hedekvist, and P. Andrekson, "Long-term 160 Gb/s-TDM, RZ transmission with automatic PMD compensation and system monitoring using an optical sampling system," in "Proceedings 27th European Conference on Optical Communication (Cat. No.01TH8551),", vol. 6 (IEEE, 2001), vol. 6, pp. 18–19.
- [114] K. Harako, D. Seya, T. Hirooka, and M. Nakazawa, "640 Gbaud (1.28 Tbit/s/ch) optical Nyquist pulse transmission over 525 km with substantial PMD tolerance," Optics Express 21, 21063 (2013).
- [115] S. J. Savory, G. Gavioli, R. I. Killey, and P. Bayvel, "Transmission of 42.8Gbit/s Polarization Multiplexed NRZ-QPSK over 6400km of Standard Fiber with no Optical Dispersion Compensation," in "OFC/NFOEC 2007 - 2007 Conference on Optical Fiber Communication and the National Fiber Optic Engineers Conference," (IEEE, 2007), pp. 1–3.
- [116] F. Buchali and H. Bulow, "Adaptive PMD Compensation by Electrical and Optical Techniques," Journal of Lightwave Technology 22, 1116– 1126 (2004).
- [117] D. Hillerkuss, R. Schmogrow, T. Schellinger, M. Jordan, M. Winter, G. Huber, T. Vallaitis, R. Bonk, P. Kleinow, F. Frey, M. Roeger, S. Koenig, A. Ludwig, A. Marculescu, J. Li, M. Hoh, M. Dreschmann, J. Meyer, S. Ben Ezra, N. Narkiss, B. Nebendahl, F. Parmigiani, P. Petropoulos, B. Resan, A. Oehler, K. Weingarten, T. Ellermeyer, J. Lutz, M. Moeller, M. Huebner, J. Becker, C. Koos, W. Freude, and J. Leuthold, "26 Tbit s1 line-rate super-channel transmission utilizing all-optical fast Fourier transform processing," Nature Photonics 5, 364– 371 (2011).
- [118] P. Bower and I. Dedic, "High speed converters and DSP for 100G and beyond," Optical Fiber Technology 17, 464–471 (2011).
- [119] T. Fujisawa, S. Kanazawa, K. Takahata, W. Kobayashi, T. Tadokoro,

H. Ishii, and F. Kano, "1.3- μ m, 4 25-Gbit/s, EADFB laser array module with large-output-power and low-driving-voltage for energy-efficient 100GbE transmitter." Optics express **20**, 614–20 (2012).

- [120] R. G. Hunsperger, Integrated Optics: Theory and Technology (Springer, 2002).
- [121] S. Aleksić, "Analysis of Power Consumption in Future High-Capacity Network Nodes," Journal of Optical Communications and Networking 1, 245 (2009).
- [122] P. Pepeljugoski, J. Kash, F. Doany, D. Kuchta, L. Schares, C. Schow, M. Taubenblatt, B. Offrein, and A. Benner, "Low power and high density optical interconnects for future supercomputers," (2010).
- [123] Y. Zhang, P. Chowdhury, M. Tornatore, and B. Mukherjee, "Energy Efficiency in Telecom Optical Networks," IEEE Communications Surveys & Tutorials 12, 441–458 (2010).
- [124] R. S. Tucker, "Green Optical Communications Part II: Energy Limitations in Networks," IEEE Journal of Selected Topics in Quantum Electronics 17, 261–274 (2011).
- [125] D. A. B. Miller, "Are optical transistors the logical next step?" Nature Photonics 4, 3–5 (2010).
- [126] R. S. Tucker and K. Hinton, "Energy Consumption and Energy Density in Optical and Electronic Signal Processing," IEEE Photonics Journal 3, 821–833 (2011).
- [127] L. Yan, A. Willner, X. Wu, and A. Yi, "All-Optical Signal Processing for UltraHigh Speed Optical Systems and Networks," Journal of Lightwave Technology 30, 3760–3770 (2012).
- [128] J. Kurumida and S. J. B. Yoo, "Nonlinear Optical Signal Processing in Optical Packet Switching Systems," IEEE Journal of Selected Topics in Quantum Electronics 18, 978–987 (2012).
- [129] S. Singh and L. Lovkesh, "Ultrahigh speed optical signal processing logic based on an SOA-MZI," IEEE Journal of Selected Topics in Quantum Electronics 18, 970–977 (2012).
- [130] A. Willner, O. Yilmaz, J. Wang, X. Wu, A. Bogoni, L. Zhang, and S. R. Nuccio, "Optically efficient nonlinear signal processing," IEEE Journal of Selected Topics in Quantum Electronics 17, 320–332 (2011).

- [131] C. DeCusatis, Fiber Optic Data Communication: Technological Trends and Advances (Academic Press, 2002).
- [132] P. Mamyshev, "All-optical data regeneration based on self-phase modulation effect," in "24th European Conference on Optical Communication. ECOC '98 (IEEE Cat. No.98TH8398),", vol. 1 (Telefonica, 1998), vol. 1, pp. 475–476.
- [133] C. J. McKinstrie and S. Radic, "Phase-sensitive amplification in a fiber," Optics Express 12, 4973 (2004).
- [134] R. Neo, J. Schröder, Y. Paquot, D.-Y. Choi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Phase-sensitive amplification of light in a $\chi(3)$ photonic chip using a dispersion engineered chalcogenide ridge waveguide." Optics express **21**, 7926–33 (2013).
- [135] J. Martinez, F. Ramos, and J. Marti, "All-optical packet header processor based on cascaded SOA-MZIs," Electronics Letters 40, 894 (2004).
- [136] F. Tong, C. Chan, and L. Chen, "An all-optical packet header recognition scheme for self-routing packet networks," in "Optical Fiber Communication Conference and Exhibit," (Opt Soc. America, 2002), pp. 284–285.
- [137] J. E. McGeehan, S. Kumar, and A. E. Willner, "Simultaneous optical digital half-subtraction and -addition using SOAs and a PPLN waveguide." Optics express 15, 5543–5549 (2007).
- [138] K. Azadet, E. Haratsch, and H. Kim, "Equalization and FEC techniques for optical transceivers," IEEE Journal of Solid-State Circuits 37, 317– 327 (2002).
- [139] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems." Optics express 18, 17075–88 (2010).
- [140] X. Wang, J. Peng, M. Li, Z. Shen, and O. Shan, "Carry-free vectormatrix multiplication on a dynamically reconfigurable optical platform," Applied Optics 49, 2352–2362 (2010).
- [141] P. Yeh and A. E. T. Chiou, "Optical matrix-vector multiplication through four-wave mixing in photorefractive media," Optics Letters 12, 138–140 (1987).
- [142] H. J. Caulfield and S. Dolev, "Why future supercomputing requires optics," Nature Photonics 4, 261–263 (2010).

- [143] J. Wang, Q. Sun, and J. Sun, "All-optical 40 Gbit/s CSRZ-DPSK logic XOR gate and format conversion using four-wave mixing," Optics Express 17, 12555–12563 (2009).
- [144] J. Schröder, M. A. F. Roelens, L. B. Du, A. J. Lowery, S. Frisken, and B. J. Eggleton, "An optical FPGA: reconfigurable simultaneous multioutput spectral pulse-shaping for linear optical processing," Optics Express 21, 690–697 (2013).
- [145] T. Gyselings, G. Morthier, and R. Baets, "Crosstalk Analysis of Multiwavelength Optical Cross Connects," J. Lightwave Technol. 17, 1273– 1283 (1999).
- [146] J. Zhou, M. O'Mahony, and S. Walker, "Analysis of optical crosstalk effects in multi-wavelength switched networks," IEEE Photonics Technology Letters 6, 302–305 (1994).
- [147] G. Hill, P. Chidgey, F. Kaufhold, T. Lynch, O. Sahlen, M. Gustavsson, M. Janson, B. Lagerstrom, G. Grasso, F. Meli, S. Johansson, J. Ingers, L. Fernandez, S. Rotolo, A. Antonielli, S. Tebaldini, E. Vezzoni, R. Caddedu, N. Caponio, F. Testa, A. Scavennec, M. O'Mahony, J. Zhou, A. Yu, W. Sohler, U. Rust, and H. Herrmann, "A transport network layer based on optical network elements," Journal of Lightwave Technology 11, 667–679 (1993).
- [148] A. Hill and D. Payne, "Linear crosstalk in wavelength-divisionmultiplexed optical-fiber transmission systems," Lightwave Technology, Journal of 3 (1985).
- [149] J. G. Proakis and M. Salehi, *Digital Communications* (McGraw-Hill Higher Education, 2008).
- [150] A. A. Bruen and M. A. Forcinito, Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century (John Wiley & Sons, 2011).
- [151] D. Russell, The Principles of Computer Networking, vol. 1989 (Cambridge University Press, 1989).
- [152] T. Mizuochi, "Recent progress in forward error correction and its interplay with transmission impairments," IEEE Journal of Selected Topics in Quantum Electronics 12, 544–554 (2006).
- [153] F. Chang, K. Onohara, and T. Mizuochi, "Forward error correction for 100 G transport networks," IEEE Communications Magazine pp. 48–55 (2010).

- [154] C. Xie, Y. Zhao, Z. Xiao, D. Chang, and F. Yu, "FEC for high speed optical transmission," Proc. of SPIE-OSA-IEEE 8309, 83091 R-1 (2011).
- [155] M. Freiberger, D. Templeton, and E. Mercado, "Low latency optical services," in "National Fiber Optic Engineers Conference," (OSA, Washington, D.C., 2012), p. NTu2E.1.
- [156] Y. Miyata, K. Sugihara, W. Matsumoto, K. Onohara, T. Sugihara, K. Kubo, H. Yoshida, and T. Mizuochi, "A triple-concatenated FEC using soft-decision decoding for 100 Gb/s optical transmission," in "Optical Fiber Communication Conference," (OSA, Washington, D.C., 2010), p. OThL3.
- [157] M. Suzuki and H. Uenohara, "Investigation of all-optical error detection circuit using SOA-MZI-based XOR gates at 10 Gbit/s," Electronics letters 45, 224 – 225 (2009).
- [158] Y. Aikawa, S. Shimizu, and H. Uenohara, "Investigation of all-optical division processing using a SOA-MZI-based XOR gate for all-optical FEC with cyclic code," Photonics in Switching 1, 3–5 (2010).
- [159] J. Schröder, O. Brasier, J. VanErps, M. A. F. Roelens, S. Frisken, and B. J. Eggleton, "OSNR monitoring of a 1.28 Tbaud signal by interferometry inside a wavelength-selective switch," Journal of Lightwave Technology 29, 1542–1546 (2011).
- [160] O. Almomani, S. Hassan, and S. A. Nor, "Effects of Packet Size on FEC Performance," 2008, 21–22 (2008).
- [161] "Internet Header Format, Internet Protocol DARPA Internet program protocol specification," Tech. rep., IETF (1981).
- [162] S. B. Wicker and V. K. Bhargava, Reed-Solomon codes and their applications (John Wiley & Sons, 1999).
- [163] J. Yu and X. Zhou, "Multilevel modulations and digital coherent detection," Optical Fiber Technology 15, 197–208 (2009).
- [164] A. Szczepanek, "FEC Latency Considerations," in "IEEE P802.3bm 40 Gb/s and 100 Gb/s Fiber Optic Task Force," (2012), pp. 1–9.
- [165] A. Ellis, J. Lucek, D. Pitcher, D. Moodie, and D. Cotter, "Full 10 x 10 Gbit/s OTDM data generation and demultiplexing using electroabsorption modulators," Electronics Letters 34, 1766 (1998).
- [166] P. Andrekson, "Picosecond optical sampling using four-wave mixing in fibre," Electronics Letters 27, 1440–1441 (1991).

- [167] Y. Paquot, J. Schröder, J. Van Erps, T. D. Vo, M. D. Pelusi, S. Madden, B. Luther-Davies, and B. J. Eggleton, "Single parameter optimization for simultaneous automatic compensation of multiple orders of dispersion for a 1.28 Tbaud signal," Optics Express 19, 25512–25520 (2011).
- [168] F. E. Doany, B. Lee, A. Rylyakov, D. M. Kuchta, C. Baks, C. Jahnes, F. Libsch, and C. Schow, "Terabit/sec VCSEL-Based Parallel Optical Module Based on Holey CMOS Transceiver IC," in "National Fiber Optic Engineers Conference," (OSA, Washington, D.C., 2012), p. PDP5D.9.
- [169] C. L. Schow, F. E. Doany, A. V. Rylyakov, B. G. Lee, C. V. Jahnes, Y. H. Kwark, C. W. Baks, D. M. Kuchta, and J. A. Kash, "A 24-Channel, 300 Gb/s, 8.2 pJ/bit, Full-Duplex Fiber-Coupled Optical Transceiver Module Based on a Single Holey CMOS IC," J. Lightwave Technol. 29, 542–553 (2011).
- [170] X. Zheng, J. Lexau, Y. Luo, H. Thacker, T. Pinguet, A. Mekis, G. Li, J. Shi, P. Amberg, N. Pinckney, K. Raj, R. Ho, J. E. Cunningham, and A. V. Krishnamoorthy, "Ultra-low-energy all-CMOS modulator integrated with driver." Optics express 18, 3059–70 (2010).
- [171] X. Zheng, D. Patil, J. Lexau, F. Liu, G. Li, H. Thacker, Y. Luo, I. Shubin, J. Li, J. Yao, P. Dong, D. Feng, M. Asghari, T. Pinguet, A. Mekis, P. Amberg, M. Dayringer, J. Gainsley, H. F. Moghadam, E. Alon, K. Raj, R. Ho, J. E. Cunningham, and A. V. Krishnamoorthy, "Ultra-efficient 10 Gb/s hybrid integrated silicon photonic transmitter and receiver." Optics express **19**, 5172–86 (2011).
- [172] Y. Vlasov, "Silicon CMOS-integrated nano-photonics for computer and data communications beyond 100G," IEEE Communications Magazine 50, s67–s72 (2012).
- [173] M. Djurfeldt, M. Lundqvist, C. Johansson, M. Rehn, O. Ekeberg, and A. Lansner, "Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer," IBM Journal of Research and Development 52, 31–41 (2008).
- [174] W. J. Freeman, Neurodynamics: An Exploration in Mesoscopic Brain Dynamics (Springer, 2000).
- [175] M. T. Hagan, H. B. Demuth, M. H. Beale, and Others, Neural network design (Pws Pub. Boston, 1996).
- [176] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. thesis, Harvard University, Cambridge, MA (1974).

- [177] H. Jaeger and H. Haas, "Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication." Science (New York, N.Y.) **304**, 78–80 (2004).
- [178] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: a new framework for neural computation based on perturbations." Neural computation 14, 2531–60 (2002).
- [179] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," Chemometrics and Intelligent Laboratory Systems 39, 43–62 (1997).
- [180] R. Hecht-Nielsen, "Theory of the backpropagation neural network," (1989).
- [181] P. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," (1974).
- [182] P. D. Wasserman, Advanced Methods in Neural Computing (John Wiley & Sons, Inc., 1993).
- [183] R. C. Frye, E. A. Rietman, and C. C. Wong, "Back-propagation learning and nonidealities in analog neural network hardware." IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council 2, 110–7 (1991).
- [184] H. de Garis, B. Goertzel, J. Misra, and I. Saha, "Artificial neural networks in hardware: A survey of two decades of progress," Neurocomputing 74, 239–255 (2010).
- [185] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in "Numerical analysis," (Springer, 1978), pp. 105–116.
- [186] C.-T. Lin and C. Lee, "Neural-network-based fuzzy logic control and decision system," IEEE Transactions on Computers 40, 1320–1336 (1991).
- [187] C. M. Bishop, Neural Networks for Pattern Recognition, vol. 1995 (Oxford University Press, 1995).
- [188] R. P. Lippmann, "Review of Neural Networks for Speech Recognition," Neural Computation 1, 1–38 (1989).
- [189] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme recognition using time-delay neural networks," (1989).
- [190] S. R. Bhatikar, R. L. Mahajan, K. Wipke, and V. Johnson, "Artificial Neural Network Based Energy Storage System Modeling for Hybrid Electric Vehicles," Society of Automotive Engineers (2000).

- [191] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," (1998).
- [192] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in "1990 IJCNN International Joint Conference on Neural Networks," (IEEE, 1990), pp. 1–6 vol.1.
- [193] J. Kakande, R. Slavik, F. Parmigiani, P. Petropoulos, and D. Richardson, "All-optical processing of multi-level phase shift keyed signals," (2012).
- [194] B. Widrow, D. E. Rumelhart, and M. A. Lehr, "Neural networks: applications in industry, business and science," Communications of the ACM 37, 93–105 (1994).
- [195] A. R. Omondi and J. C. Rajapakse, eds., FPGA Implementations of Neural Networks (Springer US, 2006).
- [196] S. Jung and S. su Kim, "Hardware Implementation of a Real-Time Neural Network Controller With a DSP and an FPGA for Nonlinear Systems," IEEE Transactions on Industrial Electronics 54, 265–271 (2007).
- [197] S. Eberhardt, T. Duong, and A. Thakoor, "Design of parallel hardware neural network systems from custom analog VLSI 'building block' chips," in "International Joint Conference on Neural Networks," (IEEE, 1989), pp. 183–190 vol.2.
- [198] B. Boser, E. Sackinger, J. Bromley, Y. Le Cun, and L. Jackel, "An analog neural network processor with programmable topology," IEEE Journal of Solid-State Circuits 26, 2017–2025 (1991).
- [199] D. Schwartz, R. Howard, and W. Hubbard, "A programmable analog neural network chip," IEEE Journal of Solid-State Circuits 24, 313–319 (1989).
- [200] C. Fernando and S. Sojakka, "Pattern recognition in a bucket," in "Advances in Artificial Life," (Springer, 2003), pp. 588–597.
- [201] D. Nikolić, S. Haeusler, W. Singer, and W. Maass, "Temporal dynamics of information content carried by neurons in the primary visual cortex," in "Advances in neural information processing systems," (2006), pp. 1041–1048.
- [202] D. Woods and T. J. Naughton, "Optical computing: Photonic neural networks," 8, 257–259 (2012).

- [203] T. Lu, S. Wu, X. Xu, and F. T. Yu, "Two-dimensional programmable optical neural network." Applied optics 28, 4908–13 (1989).
- [204] D. Psallis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals." Applied optics 27, 1752–9 (1988).
- [205] L.-S. Lee, H. M. Stoll, and M. C. Tackitt, "Continuous-time optical neural network associative memory," Opt. Lett. 14, 162–164 (1989).
- [206] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical Reservoir Computing," (2012).
- [207] O. Leclerc, B. Lavigne, E. Balmefrezol, P. Brindel, L. Pierre, D. Rouvillain, and F. Seguineau, "Optical Regeneration at 40 Gb/s and Beyond," Journal of Lightwave Technology 21, 2779 (2003).
- [208] J. D. Yuh and R. W. Newcomb, "A multilevel neural network for A/D conversion." IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council 4, 470–83 (1993).
- [209] S. Oda and A. Maruta, "A novel quantization scheme by slicing supercontinuum spectrum for all-optical analog-to-digital conversion," (2005).
- [210] S. Oda and A. Maruta, "Two-bit all-optical analog-to-digital conversion by filtering broadened and split spectrum induced by soliton effect or self-phase modulation in fiber," (2006).
- [211] A. Daryoush, X. Hou, and W. Rosen, "All-optical ADC and its applications in future communication satellites," in "2004 IEEE International Topical Meeting on Microwave Photonics (IEEE Cat. No.04EX859)," (IEEE, 2004), pp. 182–185.
- [212] I. A. Goncharenko, A. K. Esman, V. K. Kuleshov, and V. A. Pilipovich, "Optical broadband analogdigital conversion on the base of microring resonator," Optics Communications 257, 54–61 (2006).
- [213] T. Nishitani, T. Konishi, and K. Itoh, "Resolution Improvement of All-Optical Analog-to-Digital Conversion Employing Self-frequency Shift and Self-Phase-Modulation-Induced Spectral Compression," IEEE Journal of Selected Topics in Quantum Electronics 14, 724–732 (2008).
- [214] Y. Han, O. Boyraz, and B. Jalali, "Ultrawide-band photonic time-stretch A/D converter employing phase diversity," IEEE Transactions on Microwave Theory and Techniques 53, 1404–1408 (2005).
- [215] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," (1990).
- [216] R. M. Rodriguez-Dagnino, J. A. Faz-Gomez, and I. Larraga-OroZco, "Raman crosstalk cancellation in wavelength-division multiplexing by using neural networks," Optical Engineering 42, 1843–1849 (2003).
- [217] T. Chattopadhyay and J. N. Roy, "Polarization-encoded all-optical quaternary multiplexer and demultiplexer A proposal," Optik - International Journal for Light and Electron Optics 120, 941–946 (2009).
- [218] T. Chattopadhyay, "All-optical symmetric ternary logic gate," Optics & Laser Technology 42, 1014–1021 (2010).
- [219] R. P. Webb, A. J. Waddie, K. J. Symington, M. R. Taghizadeh, and J. F. Snowdon, "Optoelectronic Neural-Network Scheduler for Packet Switches," Appl. Opt. 39, 788–795 (2000).
- [220] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 47, 655–674 (2000).
- [221] A. Demir, "Computing Timing Jitter From Phase Noise Spectra for Oscillators and Phase-Locked Loops With White and 1/f Noise," IEEE Transactions on Circuits and Systems I: Regular Papers 53, 1869–1884 (2006).