Digital Techniques for Ultra-High Data Rate Optical Fibre Transmission

Gabriele LIGA

A thesis submitted to the University College London (UCL) for the
degree of Doctor of Philosophy (Ph.D.)

Optical Networks Group
Department of Electronic and Electrical Engineering
University College London (UCL)

I, Gabriele Liga, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated.

Abstract

THE exponential growth of the demand for higher data rates is pushing scientists to find ways to improve the internet infrastructure, which crucially relies on optical fibres. The main obstacle to increasing transmission rates of optical fibre systems is presented by the fibre Kerr nonlinear effect, which impairs signal transmission as the transmitted power is increased. Fortunately, optical coherent detection, in combination with digital signal processing techniques, have enabled more sophisticated digital receivers, tailored to the optical fibre channel.

This thesis describes a comprehensive study on the performance of two digital receiver-side techniques: digital back-propagation (DBP) and maximum likelihood sequence detection (MLSD).

DBP is the most widespread digital technique to mitigate fibre nonlinearity at the receiver. The performance of DBP, is assessed for long-haul, wide-bandwidth systems, highlighting theoretical gains and practical limitations. Analytical models to predict DBP performance are discussed and compared to numerical results. The impact of polarisation-mode dispersion on the capability of DBP to remove nonlinear impairments is investigated.

The principles of detection theory are discussed in the context of the optical fibre nonlinear channel. Following such principles, MLSD strategies are studied and their performance analysed for unrepeatered systems. A close to optimum receiver scheme, using the Viterbi algorithm, is proposed and investigated for the first time in a single-span fibre system.

Finally, information-theoretic tools are used to predict achievable information rates of both receiver schemes, when employed in combination with forward error correction codes. In particular, pragmatic coded modulation schemes were examined to assess the potential of off-the-shelf channel codes.

Both receiving strategies analysed were demonstrated to significantly outperform conventional receivers optimised for the additive white Gaussian noise channel. The results of this thesis provide a useful insight on the properties of the optical fibre channel and on the design of receivers aiming to maximise information rates through it.

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Polina Bayvel for her experienced guidance and for her generosity. Without her trust and perseverance I would not be writing these words. I equally acknowledge my second supervisor Dr Robert Killey whose patience and kindness are second to none.

I would like to say thank you to the Optical Networks Group (ONG), whose members, over the years, helped to make this PhD journey extremely pleasant. In particular, Dr Hou-Man Chin with whom I enjoyed countless coffees and bottles of wine, Dr Tianhua Xu whose surreal kindness is almost enlightening, and Dr David Ives for our interminable discussions on the GN-model. I would also like to mention the new members of group. Observing their young and fresh enthusiasm has reminded me of how I was in the past and probably how I should be in the future.

I am hugely indebted to former ONG member Dr Alex Alvarado with whom I share a true passion for communication theory which has led us into so many stimulating discussions. I owe him a great deal for showing me, through his rigorous method, how to produce high-quality research and for reminding me how beautiful is what we do.

I owe my deepest gratitude to Prof Erik Agrell from Chalmers University for being such an exemplary and inspiring figure for my research. But most of all, for sharing with me truly powerful and soothing words that helped me change the course of my PhD and, maybe, of my life. I will never forget that.

I also want to thank Cristian Bogdan Czegledi who visited ONG for a few months. He has been a wonderful collaborator and good friend. I hope to have other opportunities to work with him in the future.

I would like to pay tribute to my family: I thank my mom Maria and my sister Serena for their continuous support over the years through pure and unconditional love. Thanks to their presence, I never felt entirely alone, albeit far from home.

Finally, I acknowledge my father, whose efforts made me the man I am today. So many times daily life tricked me into thinking he never existed. I was just not seeing he turned into a part of me, rejoicing now for this achievement as I do.

Contents

List of Figures

List of Tables

1

Introduction

1.1 The *capacity crunch* problem

Internet applications and services have become a pervasive presence in our lives. Due to this unstoppable growth in popularity, internet devices constantly require higher throughputs to sustain such services which, in turn, are becoming ever more data-hungry in order to provide higher quality (e.g. virtual reality, video contents, IP-TV etc.). Furthermore, the number of devices per user is increasing at an unprecedented rate, pushed by the *Internet of Things* concept. Indeed, it is predicted that by 2020 the number of devices connected to IP networks will be three times the world population in that year [1]. All these factors contribute to the growth of the global traffic demand. Recent estimates indicate that the internet traffic is growing exponentially, at a rate of 22% per year (compound annual growth rate) over the next five years [1]. The internet infrastructure is, thus, under pressure to deliver higher throughputs to accommodate for the expected growth in data demand in the near future.

The internet backbone is fundamentally underpinned by optical fibre networks which carry 99% of the internet data traffic. The map in Fig. 1.1 shows how optical fibres have become necessary to move large quantity of data across continents by means of undersea cables. Fibre cables have also made their way into land-line links, becoming by far the first technological option to aggregate and distribute data within wide geographical areas.

The choice of optical fibres as a transmission medium is due to their extremely

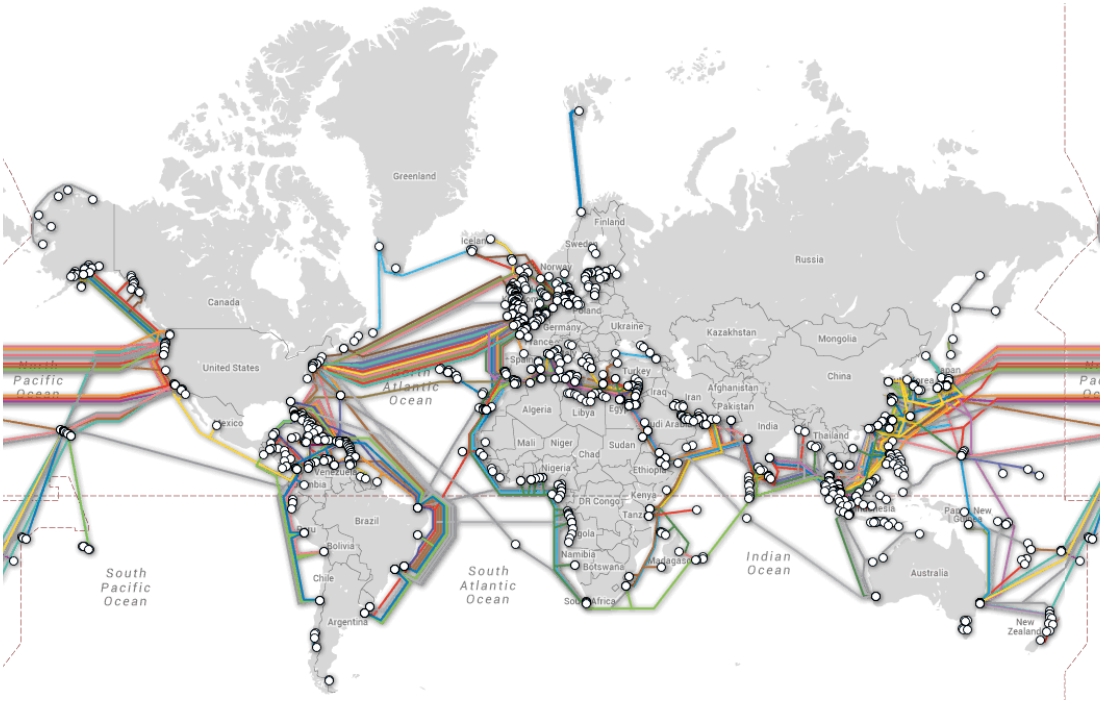


Figure 1.1: Geographical map of the submarine optical fibre links underpinning the backbone of the internet [2].

large bandwidth and low attenuation which allows to transmit extremely high data rates over very long distances. Such transmission performance is currently unrivalled. An example of large data rate, long-distance transmission is given by the FASTER undersea fibre cable [2], partly owned by Google, which is among the highest-speed submarine systems deployed to date: it transmits up to 60 Tb/s (10 Tb/s per fibre pair) over about 9000 km connecting the US with Japan across the Pacific ocean with no electrical regeneration.

Undersea cables and medium/long-haul terrestrial links form what is typically referred to as a *transport network*. However, metropolitan networks also strongly rely on optical fibres to *backhaul* mobile networks, i.e. to link mobile networks to the core network. Finally, access networks are transitioning to optical fibres to provide higher-speed broadband services to domestic users. There is then no doubt that the data traffic demand growth will have to be met by optical fibre systems.

When optical fibres started to be deployed for long-haul systems in 1980's, they were considered to have an unlimited transmission bandwidth, at least for any foreseeable data rate requirement. Thirty years later, the throughputs demonstrated in record optical fibre transmission experiments are showing clear signs of saturation. This saturation is caused by the limitations on reliable data transmission imposed by the Kerr nonlinear effects that characterise the optical fibre channel. As commercial systems approach the data rates of the state-of-the-art laboratory transmission experiments, the scientific community is concentrating its efforts on protracting the growth of data rates for the years to come. This data rate exhaust is sometimes referred to as *capacity crunch* [3–5], and has been the main argument underpinning research efforts on optical fibre transmission systems in recent years.

The struggle in sustaining the growth of optical fibre transmission throughput is highlighted by the recent so-called *hero* experiments, i.e. transmission experiments, carried out in research laboratories, which set a new performance record. Fig. 1.2a shows the total transmission throughputs attained in these experiments over the last 16 years. The modulation formats used to achieve these records are also shown. A first thing to note is that the currently holding record transmission throughput on a single-mode fibre was achieved in 2012, and corresponds to 102.3 Tb/s transmission over 240 km [6]. From 2012 to date, no new experiments with higher throughputs have been reported (for single-mode fibre transmission), indicating a clear saturation. The early 2000s saw a slow growth of the data rate records as direct-detection technology reached its maturity and the optical bandwidth was already fully utilised thanks to the adoption of wavelength-division multiplexing (WDM). However, 2007-2011 saw an exponential growth coincident with the revived application of coherent detection techniques in combination with digital signal processing (DSP), which enabled modulation formats able to encode higher number of bits for each transmitted symbol.

As state-of-the-art optical transmission demonstrations are entirely exhausting the available optical bandwidth, it has become necessary for future optical systems to increase the amount of bit/s transmitted over a given portion of bandwidth, a quantity known as spectral-efficiency (SE). The evolution of the SE in optical transmission experiments is shown in Fig. 1.2b. Until 2006 the SE of fibre transmission was below 2 bit/(s·Hz), due to the use of direct-detection systems and coarse spacing between WDM channels. In 2006–2010, coherent detection lead to a dramatic SE increase thanks to the use of polarisation-multiplexed (PM), high-cardinality modulation formats such as quadrature-amplitude modulation (QAM) with up to 512 points, and denser WDM channel packing. More recently, in 2014, a new record was set at 15.3 bit/(s·Hz) using PM-2048QAM [7]. However, saturation of the SE also appears evident. Moreover, it must be mentioned that record SE experiments are, in most cases, conducted with a single transmitted channel. The reason is the greater difficulty of achieving high SE with multiple WDM channels, due to the increased nonlinear interference compared to the case of a reduced number of channels. In other words, SE does not stay constant as the total optical bandwidth increases (as in a linear channel) but rather tends to decrease. As a result, different SE records should be compared only for a fixed transmitted bandwidth.

To give a better idea on how different experiments compare with each other, a list including all the most relevant recent experiments in optical fibre transmission is shown in table 1.1. It can be seen that the typical SE of state-of-the-art experiments performed over a fully populated C- and L-band is around 6 bit/(s·Hz) for long-haul

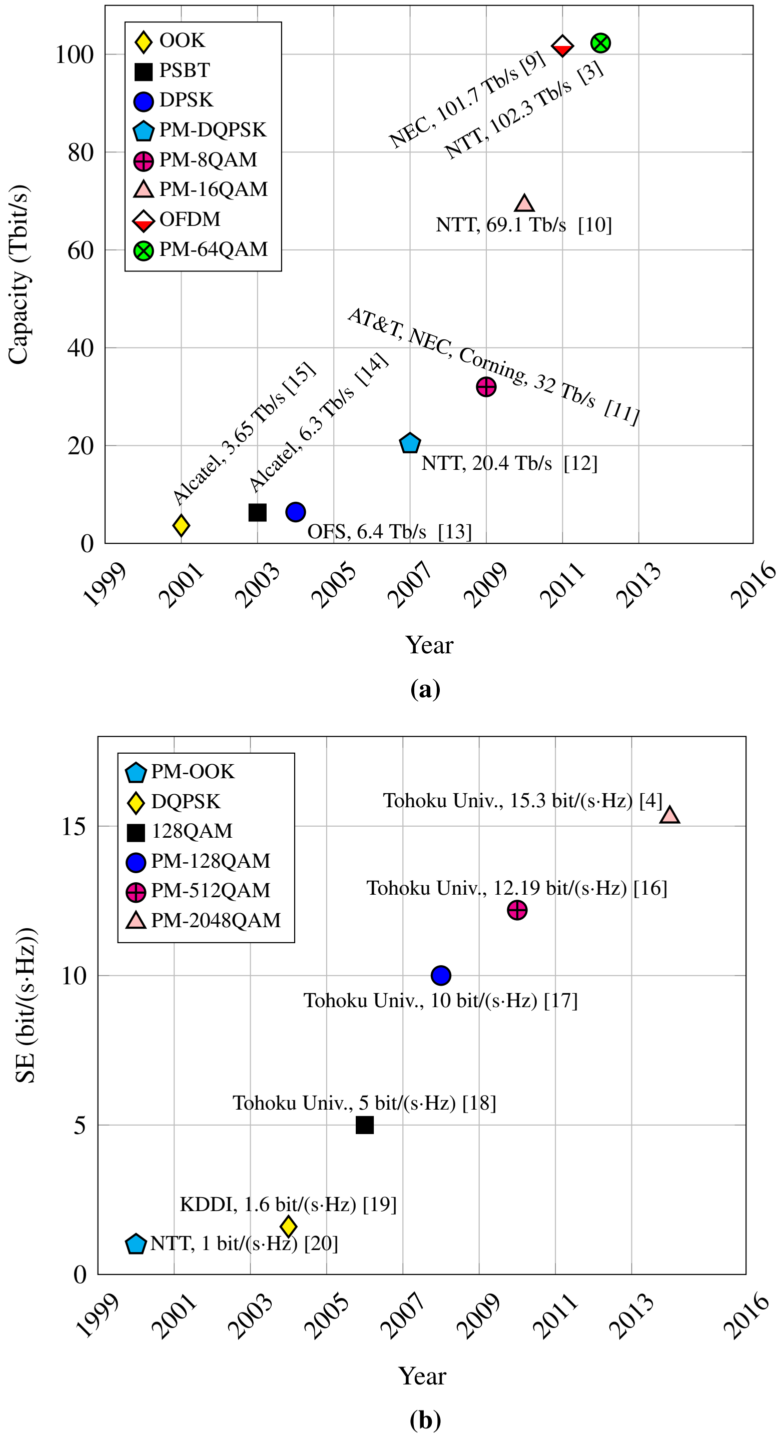


Figure 1.2: Evolution of (a) the record throughputs and (b) record SE of optical fibre transmission experiments over the last two decades.

Data Rate (D)	SE	Opt. BW	Distance (L)	Year	Record	Ref.
102.3 Tb/s	9.1 bit/(s·Hz)	C+L	240 km	2012	D	[6]
66 Gb/s	15.3 bit/(s·Hz)	3.6 GHz	150 km	2014	SE	[7]
52.2 Tb/s	5.82 bit/(s·Hz)	C+L	10230 km	2015	D×L	[9]
21.2 Tb/s	6 bit/(s·Hz)	C	10290 km	2013	SE×L	[10]
71.64 Tb/s	7.36 bit/(s·Hz)	C+L	6970 km	2017	LR×L	[8]

Table 1.1: State-of-the-art experiments in optical fibre transmission.

transmission, whereas it increases up to 9.1 bit/(s·Hz) for much shorter distances (240 km). As mentioned before, very high SE such as 15.3 bit/(s·Hz) has been demonstrated only over a very small bandwidth (3.6 GHz single-channel) and short distance (150 km). Finally, the most recent experiment over transpacific distance is shown [8]. In this experimental demonstration, the highest overall data rate of 71.64 Tb/s and 70.2 Tb/s was achieved over 6970 km and 7600 km, respectively.

It is evident how, in order to overcome the capacity crunch, it will be inevitable to leverage between higher SE and larger bandwidth: an increase in the optical bandwidth is subject to the development of new optical amplification devices which can currently reach 100 nm bandwidth; further increasing the SE (while keeping transmission reliable) will instead require to tackle the current bottleneck imposed by fibre nonlinearity which will be, for a significant part, down to the development of new transceivers' schemes. Next-generation optical fibre systems will be designed according to both requirements.

1.2 Next-generation optical fibre systems

As mentioned in the previous section, coherent detection and DSP has lead to a significant increase in the SE of optical fibre transmission by enabling the adoption of PM, in-phase/quadrature modulation formats such as QAM. The tendency to adopt higher order modulation formats in optical fibre communication is highlighted by the plot in Fig. 1.3. This figure shows the number of papers considering a given MQAM modulation in the proceedings of the *Optical Fibre Communication Conference* (OFC) over the last years. While quaternary phase-shift keying (QPSK) is shown to be the most utilised format, starting from 2008 16-QAM has shown the same growth rate as QPSK. The use of higher order QAM has recently become more prevalent, with, for instance, 64QAM demonstrations going from 9% of the total number of works in 2010 to 12% in 2015. To date, QAM formats with higher cardinality than 16QAM represent together more than 20% of the works presented in OFC, demonstrating the trend to adopt high-SE modulation formats.

However, increasing the modulation order also requires higher values of the received optical signal-to-noise ratio (OSNR) in order to provide reliable transmission, which,

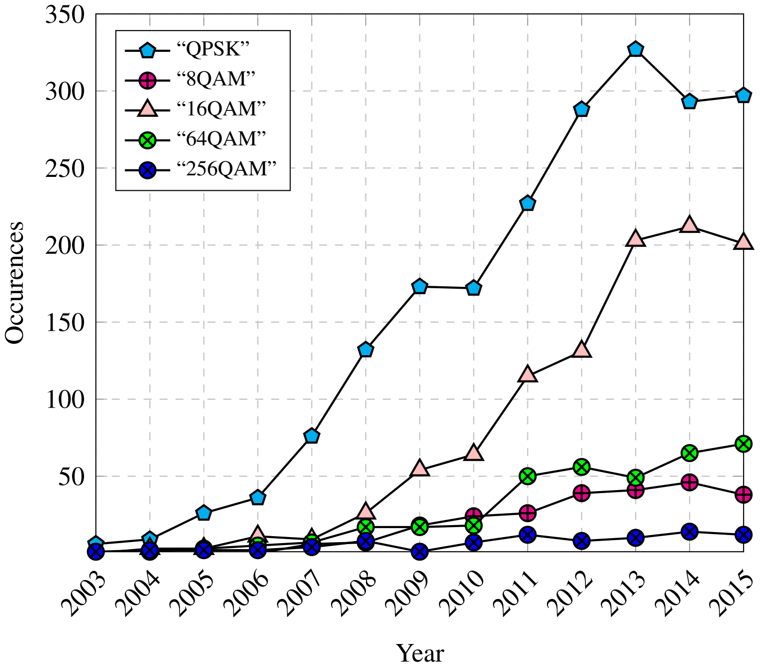


Figure 1.3: Number of papers in OFC proceedings considering MQAM formats.

in turn, means operating at higher optical transmission powers per channel. Beyond a certain power (although the optical fibre is itself only weakly nonlinear) a nonlinear regime is entered which leads to transmission impairments (as explained in sec. 2.1.2). Some of the fibre nonlinear distortions are deterministic since they can be predicted using a deterministic differential equation with known initial condition (transmitted signal). As a result, such distortions can be in principle reversed and cancelled from the detected signal. However, other nonlinear effects, such as the nonlinear interaction between signal and amplified spontaneous emission (ASE) noise originating from the optical amplifiers, cannot clearly be considered deterministic due to the stochastic nature of the ASE. These impairments cannot be compensated for (see chapter 3), hence imposing a major limitation on the achievable data rate, at least when conventional optical receivers are used [11–13]. However, the ultimate transmission limit, i.e. the channel *capacity* in Shannon's definition [14], of the optical fibre nonlinear channel remains unknown [13]. Particularly, only (many) lower-bounds, e.g. in [11, 12, 15–17], and one upperbound [18] on the capacity are known. Although these bounds are tight at relatively low powers, where the fibre is quasi-linear, it is not known how close these bounds are to the actual channel capacity in the nonlinear power regime.

Most of the current optical fibre receivers neglect nonlinearity as a source of transmission impairments, and optical fibre systems are operated at powers that are low enough to consider this to be true. Optical receivers are, therefore, optimised to operate in such a linear regime which can be modelled as the well-known additive white Gaussian noise (AWGN) channel [19]. Using these receiver schemes, reliable data rates in optical communication systems achieve a peak at a given power and decrease beyond that point [13].

This clearly indicates that in order to push optical fibre transmission to its fundamental limit it is critical to develop new and more sophisticated receiver techniques. In particular, digital receivers should be designed adapting communication theory principles to the nonlinear fibre channel. When a channel is affected by interference and distortion, such as in the optical fibre case, two main approaches are possible: equalisation or optimised detection [20, Ch. 7,8]. Although in some cases the distinction between these two approaches is not necessarily sharp, in general, important differences can be found: equalisation techniques aim to undo channel distortions from the received signal, producing an *equalised.* i.e. distortion-free, signal which is impaired only by a *true* stochastic process; optimised detection strategies instead do not modify the channel detrimental effects, but rather attempt to produce the best estimate (in the minimum error probability sense) of the transmitted data based on the observation of the distorted output signal.

The most popular algorithm implementing nonlinear equalisation in the context of the optical fibre channel is called digital backpropagation (DBP) [21]. DBP aims to compensate for deterministic nonlinear effects using a nonlinear *zero-forcing* equalisation strategy [20, Ch. 8], i.e. inverting the channel response, based on the knowledge of the fibre propagation equations. However, like all *zero-forcing* equalisation techniques, in some cases the effect of the noise can be enhanced by the equalisation process, which can be therefore detrimental for the error rate performance of the receiver.

The alternative is to resort to statistical receivers that aim to minimise the error probability, based on the knowledge of the channel probabilistic description. As the optical fibre channel has memory, optimised statistical receivers are expected to account for this memory when estimating the transmitted data. A way to do that is applying a maximum likelihood sequence detection (MLSD) strategy. Receivers implementing an optimum detection strategy are also by definition optimal receivers in the error probability sense. However, complexity issues often prevent a realistic realisation of such receivers and more pragmatic schemes must be considered.

In this thesis, both DBP and MLSD techniques are analysed and their performance are assessed, in particular in comparison to receivers optimised for an AWGN channel, which represent the standard approach for fibre transmission systems. Benefits and shortcomings of both receiver schemes are highlighted, providing insights on the receiver design issues.

Due to the markedly different features of the two receiver strategies analysed, each scheme was studied in different transmission scenarios. As a result, the work in this thesis should not be interpreted as a comparative study between DBP- and MLSD-based receivers. Instead, one of the aims of this thesis is to help understanding the potential of the two receiver schemes, and the suitability of one scheme or the other for a specific transmission scenario. For the first time, all the main transmission performance metrics are used in the characterisation of the two schemes, such as: bit error rate, signal-to-noise ratio and achievable information rates. The combination of the three metrics allows a comprehensive analysis of each receiver scheme in both the uncoded and coded regimes.

1.3 Thesis outline

The remainder of this thesis is structured as follows.

In **Chapter 2**, the fundamental theoretical tools to understand the engineering problems underlying optical fibre communications are introduced. The chapter first highlights the physics of optical fibre transmission discussing the main pulse propagation effects. Then, the reference optical fibre transmission systems and subsystems used in this thesis are described, such as: multi-span fibre systems and coherent receivers. The most widespread analytical models and numerical methods for optical fibre propagation are also discussed.

Chapter 3 analyses the performance of DBP in multi-span EDFA-amplified optical fibre systems for a wide variety of system parameters. Particularly, DBP effectiveness is evaluated in the context of wideband, long-haul transmission scenarios. Firstly, an analytical study is performed based on the analytical tools described in chapter 2. Secondly, numerical results, used as validation of the analytical approach, are presented. Ideal gains and practical limitations of the DBP algorithm are discussed.

Chapter 4 tackles the problem of optimising the detection strategy in the presence of nonlinear distortions with memory which is a salient feature of the optical fibre channel. The properties of a single-span fibre channel are outlined and design of maximum likelihood sequence detection receivers tailored to this channel is described. The uncoded bit-error rate performance of such receivers is then assessed through numerical simulations and compared to conventional receiver approaches optimised for the linear AWGN channel.

In **Chapter 5**, information-theoretic quantities are used to describe the performance of the receiver schemes studied in chapters 3 and 4 when channel coding is adopted. First, achievable information rates are quantified for long-haul, high-SE optical fibre systems. A comparative study was performed for different receiver schemes including DBP and three-different PM-MQAM modulation formats. In particular, achievable information rates, when pragmatic encoder/decoder pairs are adopted, are presented and compared across a wide range of receiver architectures.

Chapter 6 draws conclusions on the performance of the two receiver schemes analysed in chapters 3, 4 and 5 and outlines new research ideas to expand the potential of such receivers and further pushing the limits of optical fibre communication.

Finally, the **Appendix** complements some of the concepts discussed in chapters 3 and 4 adding an analytical derivation of the signal-ASE nonlinear interference power and the mathematical description of the Viterbi algorithm adopted for the results in chapter 4.

1.4 Key contributions

i) In section 3.4, closed-form expressions from the available channel models are derived to facilitate predictions on the performance of DBP. These results were partly included in [22, Sec. 3], [23] and were used to produce some of the results in [23–25].

ii) In section 3.5, DBP performance is studied with respect to the constraints imposed by pragmatic DSP implementations. This study was published in [26].

iii) In section 3.5, DBP performance is numerically evaluated for the first time for large compensation bandwidths and wideband optical transmission (1 THz). This study was included in [22, Sec. 3].

iv) Polarisation-mode dispersion (PMD) impact on the performance of multi-channel DBP is assessed as a function of the PMD parameter and compensation bandwidth. This study was published in [26–30].

v) In section 3.7.1, a novel Monte-Carlo approach to efficiently simulate wideband optical fibre systems in the presence of PMD is introduced. The method is used to evaluate the effect of PMD on the performance of DBP and led to the results published in [31].

vi) In chapter 4, the assessment of an optimum detection receiver for the single-span optical fibre channel is performed for the first time. This work resulted in the paper [32].

vii) A comprehensive study of the achievable information rates for optical fibre transmission systems employing a wide variety of receiver architectures and FEC schemes is presented in chapter 5. This study was published in [30].

1.5 List of publications

The research work carried out for this thesis has also lead to the following publications:

Journal papers

1. J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, “Digital signal processing for fiber nonlinearities,” OSA Optics Express, vol. 25, no. 3, pp. 1916–1936, 2017.
2. C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, “Polarization-mode dispersion aware digital backpropagation,” OSA Optics Express, vol. 25, no. 3, pp. 1903–1915, 2017.
3. D. Lavery, R. Maher, G. Liga, D. Semrau, L. Galdino, and P. Bayvel, "On the bandwidth dependent performance of split transmitter-receiver optical fiber nonlinearity compensation," OSA Optics Express, vol. 25, no. 4, 4554–4563, 2017.
4. T. Xu, N. Shevchenko, D. Lavery, D. Semrau, G. Liga, A. Alvarado, R. Killey, and P. Bayvel, "Modulation format dependence of digital nonlinearity compensation performance in optical fibre communication systems," OSA Optics Express, vol. 25, no. 4, 3311–3326, 2017.
5. G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, “Information rates of next-generation long-haul optical fiber systems using coded modulation,” IEEE Journal of Lightwave Technology, vol. 35, no. 1, pp. 113–123, 2017.
6. D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, “The benefit of split nonlinearity compensation for single-channel optical fiber communications,” IEEE Photonics Technology Letters, vol. 28, no. 17, pp. 1803–1806, 2016.
7. P. Bayvel, R. Maher, T. Xu, G. Liga, N. A. Shevchenko, D. Lavery, A. Alvarado, and R. I. Killey, “Maximizing the optical network capacity,” Philosophical Transactions of the Royal Society, A. Mathematical, Physical and Engineering Sciences, vol. 374, no. 2062, 2016.
8. A. D. Ellis, M. Tan, A. Iqbal, M. A. Z. Al-khateeb, V. Gordienko, G. S. Mondaca, S. Fabbri, M. F. C. Stephens, M. E. Mccarthy, A. Perentos, I. D. Phillips, G. Liga, R. Maher, P. Harper, N. Doran, S. K. Turitsyn, S. Sygletos, and P. Bayvel, “4 Tb/s transmission reach enhancement using 10×400 Gbit/s dual band optical phase conjugation,” IEEE Journal of Lightwave Technology, vol. 34, no. 8, pp. 1717–1723, 2016.
9. T. Xu, G. Liga, D. Lavery, B. C. Thomsen, S. J. Savory, R. I. Killey, and P. Bayvel, “Equalization enhanced phase noise in Nyquist-spaced superchannel transmission

systems using multi-channel digital back-propagation.," Nature Scientific Reports, vol. 5, pp. 13990, 2015.

10. G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," OSA Optics Express, vol. 22, no. 24, pp. 30053–30062, 2014.

Conference papers

11. A. Alvarado, G. Liga, T. Fehenberger, and L. Schmalen, "On the design of coded modulation for fiber optical communications," Signal Processing in Photonics Communications (SPPComm), New Orleans, USA, 2017.
12. C. B. Czegledi, G. Liga, M. Karlsson, and E. Agrell, "Modified digital backpropagation accounting for polarization-mode dispersion," Optical Fiber Communication Conference (OFC), Los Angeles, USA, 2017.
13. G. Liga, A. Alvarado, P. Bayvel, and E. Agrell, "Achievable information rates of nonbinary codes for optical fiber transmission," in IEEE Photonics Conference (IPC), Waikoloa, Hawaii USA, 2016.
14. C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, "Polarization-mode dispersion aware digital backpropagation," European Conference on Optical Communications (ECOC), Dusseldorf, Germany, 2016.
15. G. Liga, C. Czegledi, T. Xu, E. Agrell, R. I. Killey, and P. Bayvel, "Ultra-wideband nonlinearity compensation performance in the presence of PMD," European Conference on Optical Communications (ECOC), Dusseldorf, Germany, 2016.
16. R. Maher, G. Liga, M. Paskov, and A. Alvarado, "Capacity approaching transmission using probabilistic shaping and DBP for PFE constrained submarine optical links," European Conference on Optical Communications (ECOC), Dusseldorf, Germany, 2016.
17. N. A. Shevchenko, T. Xu, D. Semrau, G. Saavedra, G. Liga, M. Paskov, L. Galdino, A. Alvarado, R. I. Killey, and P. Bayvel, "Achievable information rates estimation for 100-nm Raman-amplified optical transmission system," European Conference on Optical Communications (ECOC), Dusseldorf, Germany, 2016.
18. L. Galdino, G. Liga, G. Saavedra, D. Ives, R. Maher, A. Alvarado, S. Savory, R. I. Killey and P. Bayvel, "Experimental demonstration of modulation-dependent nonlinear Interference in Optical Fibre Communication," European Conference on Optical Communications (ECOC), Dusseldorf, Germany, 2016.

19. G. Liga, A. Alvarado, E. Agrell, M. Secondini, R. I. Killey, and P. Bayvel, "Optimum detection in presence of nonlinear distortions with memory," European Conference on Optical Communications (ECOC), Valencia, Spain, 2015.

20. A. D. Ellis, I. D. Phillips, M. Tan, M. F. C. Stephens, M. E. Mccarthy, M. A. Z. Al Kahteeb, M. . A. Iqbal, A. Perentos, S. Fabbri, V. Gordienko, D. Lavery, G. Liga, G. S. M., R. Maher, S. Sygletos, P. Harper, N. J. Doran, P. Bayvel, and S. K. Turitsyn,"Enhanced superchannel transmission using phase conjugation," European Conference on Optical Communications (ECOC), Valencia, Spain, 2015.

21. A. D. Ellis, S. T. Le, M. A. Z. Al-Khateeb, S. K. Turitsyn, G. Liga, D. Lavery, T. Xu, and P. Bayvel, "The impact of phase conjugation on the nonlinear-Shannon limit: the difference between optical and electrical phase conjugation," IEEE Summer Topicals Meeting (SUM), Nassau, Bahamas, 2015.

22. L. Galdino, G. Liga, D. Lavery, R. Maher, T. Xu, M. Sato, R. I. Killey, S. J. Savory, B. C. Thomsen, and P. Bayvel, "Unrepeatered transmission over 253.4 km ultra low loss fibre achieving 6.95 b/s/Hz SE using EDFA-only pre-amplifier," European Conference on Optical Communications (ECOC), Cannes, France, 2014.

23. G. Liga, T. Xu, L. Galdino, R. Killey, and P. Bayvel, "Digital back-propagation for high spectral efficiency Terabit/s superchannels," Optical Fiber Communication Conference (OFC), San Francisco, California USA, 2014.

References

[1] Cisco, "White paper: Cisco VNI forecast and methodology, 2015–2020," 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

[2] [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html

[3] A. Chralyvy, "Plenary paper: the coming capacity crunch," in *Proc. European Conference on Optical Communication (ECOC)*, Vienna, Austria, 2009.

[4] R. W. Tkach, "Scaling optical communications for the next decade and beyond," *Bell Syst. Tech. J.*, vol. 14, no. 4, pp. 3–9, 2010.

[5] A. D. Ellis, N. M. Suibhne, D. Saad, and D. N. Payne, "Communication networks beyond the capacity crunch," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2062, 2016.

[6] A. Sano, T. Kobayashi, S. Yamanaka, A. Matsuura, H. Kawakami, Y. Miyamoto, K. Ishihara, and H. Masuda, "102.3 Tb/s (224 x 548-Gb/s) C- and extended L-band all-Raman transmission over 240 km using PDM-64QAM single carrier FDM with digital pilot tone," in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, USA, 2012.

[7] S. Beppu, K. Kasai, M. Yoshida, and M. Nakazawa, "2048 QAM (66 Gbit/s) single-carrier coherent optical transmission over 150 km with a potential SE of 15.3 bit/s/Hz," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, CA, USA, 2014.

[8] J.-X. Cai, H. G. Batshon, M. Mazurczyk, O. V. Sinkin, D. Wang, M. Paskov, W. Patterson, C. Davidson, P. Corbett, G. Wolter, T. Hammon, M. A. Bolshtyansky, D. Foursa, and A. Pilipetskii, "70.4 Tb/s capacity over 7,600 km in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation," in *Proc. Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA, 2017, p. Th5B.2.

[9] J.-X. Cai, Y. Sun, H. Zhang, H. G. Batshon, M. V. Mazurczyk, O. V. Sinkin, D. G. Foursa, and A. Pilipetskii, "49.3 Tb/s transmission over 9100 km using C+L EDFA and 54 Tb/s transmission over 9150 km using hybrid-Raman EDFA," *J. Lightw. Technol.*, vol. 33, no. 13, pp. 2724–2734, 2015.

[10] H. Zhang, J. Cai, H. G. Batshon, M. Mazurczyk, O. V. Sinkin, D. G. Foursa, A. Pilipetskii, G. Mohs, and N. Bergano, “200 Gb/s and dual-wavelength 400 Gb/s transmission over transpacific distance at 6 b/s/Hz spectral efficiency,” in *Proc. Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA, 2013.

[11] P. P. Mitra and J. B. Stark, “Nonlinear limits to the information capacity of optical fibre communications.” *Nature*, vol. 411, no. 6841, pp. 1027–30, 2001.

[12] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks,” *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, 2010.

[13] M. Secondini and E. Forestieri, “Scope and limitations of the nonlinear Shannon limit,” *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Feb 2017.

[14] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[15] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, “Analytical results on channel capacity in uncompensated optical links with coherent detection,” *Opt. Express*, vol. 19, no. 26, pp. B438–B449, 2011.

[16] M. Secondini, E. Forestieri, and G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, 2013.

[17] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “New bounds on the capacity of the nonlinear fiber-optic channel,” *Opt. Letters*, vol. 39, no. 2, pp. 398–401, 2014.

[18] G. Kramer, M. I. Yousefi, and F. R. Kschischang, “Upper bound on the capacity of a cascade of nonlinear and noisy channels,” in *IEEE Information Theory Workshop (ITW)*, Jerusalem, Israel, 2015.

[19] E. Agrell, A. Alvarado, and F. R. Kschischang, “Implications of information theory in optical fibre communications,” *Philosophical Transactions of The Royal Society*, vol. 374, no. 2062, 2015.

[20] J. Barry, E. Lee, and D. Messerschmitt, *Digital Communication*. Springer US, 2004.

[21] E. Ip and J. M. Kahn, “Compensation of dispersion and nonlinear impairments using digital backpropagation,” *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, 2008.

[22] J. C. Cartledge, F. P. Guiomar, F. R. Kschischang, G. Liga, and M. P. Yankov, "Digital signal processing for fiber nonlinearities," *Opt. Express*, vol. 25, no. 3, pp. 1916–1936, 2017.

[23] D. Lavery, D. Ives, G. Liga, A. Alvarado, S. J. Savory, and P. Bayvel, "The benefit of split nonlinearity compensation for single-channel optical fiber communications," *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1803–1806, 2016.

[24] R. Maher, D. Lavery, G. Liga, M. Paskov, A. Alvarado, T. Fehenberger, and P. Bayvel, "Capacity approaching transmission using probabilistic shaping and DBP for PFE constrained submarine optical links," in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[25] N. A. Shevchenko, T. Xu, D. Semrau, G. Saavedra, G. Liga, M. Paskov, L. Galdino, A. Alvarado, R. I. Killey, and P. Bayvel, "Achievable information rates estimation for 100-nm Raman-amplified optical transmission system," in *Proc. European Conference on Optical Communication (ECOC)*, Dussledorf, Germany, 2016.

[26] G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, "On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission," *Opt. Express*, vol. 22, no. 24, pp. 30 053–30 062, 2014.

[27] A. D. Ellis, I. D. Phillips, M. Tan, M. F. C. Stephens, M. E. Mccarthy, M. A. Z. A. Kahteeb, M. A. Iqbal, A. Perentos, S. Fabbri, V. Gordienko, D. Lavery, G. Liga, G. S. M., R. Maher, S. Sygletos, P. Harper, N. J. Doran, P. Bayvel, and S. K. Turitsyn, "Enhanced superchannel transmission using phase conjugation," in *Proc. European Conference on Optical Communication (ECOC)*, Valencia, Spain, 2015.

[28] A. D. Ellis, S. T. Le, M. A. Z. Al-Khateeb, S. K. Turitsyn, G. Liga, D. Lavery, T. Xu, and P. Bayvel, "The impact of phase conjugation on the nonlinear-Shannon limit: the difference between optical and electrical phase conjugation," in *Summer Topicals Meeting Series (SUM)*, Nassau, Bahamas, 2015.

[29] A. D. Ellis, M. Tan, A. Iqbal, M. A. Z. Al-khateeb, V. Gordienko, G. S. Mondaca, S. Fabbri, M. F. C. Stephens, M. E. Mccarthy, A. Perentos, I. D. Phillips, G. Liga, R. Maher, P. Harper, N. Doran, S. K. Turitsyn, S. Sygletos, and P. Bayvel, "4 Tb/s transmission reach enhancement using 10×dual band optical phase conjugation," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1717–1723, 2016.

[30] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, "Information rates of next-generation long-haul optical fiber systems using coded modulation," *J. Lightw. Technol.*, vol. 35, no. 1, pp. 113–123, 2017.

[31] G. Liga, C. Czegledi, T. Xu, E. Agrell, R. I. Killey, and P. Bayvel, "Ultra-wideband nonlinearity compensation performance in the presence of PMD," in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[32] G. Liga, A. Alvarado, E. Agrell, M. Secondini, R. I. Killey, and P. Bayvel, "Optimum detection in presence of nonlinear distortions with memory," in *Proc. European Conference on Optical Communication (ECOC)*, Valencia, Spain, 2015.

2

Theoretical tools for optical fibre communication

2.1 Theory of optical fibre propagation

The aim of this chapter is to introduce the basic concepts in the analysis and physical understanding of fibre transmission. The main challenges involved in the design of optical fibre communication systems will also be discussed.

2.1.1 From the wave equation to the NLSE

The propagation of optical pulses through an optical fibre can be described by the wave equation

$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = \mu_0\frac{\partial^2\mathbf{P}_L}{\partial t^2} + \mu_0\frac{\partial^2\mathbf{P}_{NL}}{\partial t^2}, \tag{2.1}$$

where $\mathbf{E}(\mathbf{r},t)$ represents the electric field as a function of position $\mathbf{r}$ and time t, c is the speed of light, μ_0 is the vacuum magnetic permeability, while $\mathbf{P}_L$ and $\mathbf{P}_{NL}$ are the linear part and the nonlinear part of the induced electric polarisation vector $\mathbf{P}$, respectively. Assume that each of the vectors above are aligned along the generic direction $\hat{\mathbf{x}}$, such that $\mathbf{E}(\mathbf{r},t) = E(\mathbf{r},t)\hat{\mathbf{x}}$ and $\mathbf{P}(\mathbf{r},t) = P(\mathbf{r},t)\hat{\mathbf{x}}$. This corresponds to an assumption of *isotropy* of the optical fibre medium, which can be considered valid only locally [1]. Effects arising when this assumption is not verified, such as fibre birefringence, is tackled later on in this chapter.

Under this assumption, Eq. (2.1) can be reduced to a scalar equation given by:

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = \mu_0 \frac{\partial^2 P_L}{\partial t^2} + \mu_0 \frac{\partial^2 P_{NL}}{\partial t^2}. \tag{2.2}$$

The linear component $P_L(\mathbf{r},t)$ and nonlinear component $P_L(\mathbf{r},t)$ of $\mathbf{P}(\mathbf{r},t)$ are related to $E(\mathbf{r},t)$ by

$$P_L(\mathbf{r},t) = \varepsilon_0 \int_{-\infty}^{\infty} \chi_{xx}^{(1)}(t-t')E(\mathbf{r},t')dt' \tag{2.3}$$

$$P_{NL}(\mathbf{r},t) = \varepsilon_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi_{xxxx}^{(3)}(t-t_1,t-t_2,t-t_3)E(\mathbf{r},t_1)E(\mathbf{r},t_2)E(\mathbf{r},t_3)dt_1dt_2dt_3, \tag{2.4}$$

where ε_0 is the vacuum dielectric constant while $\chi_{xx}^{(1)}$ and $\chi_{xxxx}^{(3)}$ are components of the first-order and third-order susceptibility tensors, respectively. Several different nonlinear optical effects arising in an optical fibre can be described by the third-order tensor component $\chi_{xxxx}^{(3)}$. In the following, it is assumed that $\chi_{xxxx}^{(3)}$ is instantaneous and purely imaginary. The physical phenomenon associated with this specific form of $\chi_{xxxx}^{(3)}$ is the well-known Kerr effect [1]. The main physical manifestation of the Kerr effect is the local change of the fibre refractive index due to the optical field intensity.

Eq. (2.1) cannot be easily treated without some further simplifying assumptions. A common approach takes into account the nonlinear part of the equation P_{NL} as a perturbation of the solution for the linear case (P_{NL}=0). Furthermore, $E(\mathbf{r},t)$ is assumed to be a quasi-monochromatic signal. This means that $E(\mathbf{r},t)$ (and $P(\mathbf{r},t)$) can be expressed as

$$E(\mathbf{r},t) = \frac{1}{2}[E(\mathbf{r},t)\exp(-j\omega_0 t) + E^*(\mathbf{r},t)\exp(j\omega_0 t)], \tag{2.5}$$

where $E(\mathbf{r},t)$ is a slowly-varying complex envelope (compared to the frequency of oscillation ω_0). Under the above conditions, the method of the separation of the variables can be used to solve Eq. (2.1). According to this method, the generic solution can be expressed in the form

$$E(\mathbf{r},t) = F(x,y)A(z,t)\exp(j\beta_0 z) \tag{2.6}$$

where β_0 is the *wave number* at frequency $\omega = 0$.

Substituting Eq. (2.6) into Eq. (2.1) and switching to the Fourier domain, we obtain a pair of equations with decoupled variables [1, Sec. 2.3]

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + [\varepsilon(\omega)k_0^2 - \tilde{\beta}]F = 0, \tag{2.7}$$

$$2i\beta_0 \frac{\partial \tilde{A}}{\partial z} + (\tilde{\beta}^2 - \beta_0^2)\tilde{A} = 0 \tag{2.8}$$

where $\tilde{A} \triangleq \tilde{A}(z, \omega - \omega_0)$ is the Fourier transform of $A(z,t)$, and $\tilde{\beta}(\omega)$ is the wave number as a function of frequency. In order to obtain Eq. (2.8), a partial second derivative in z is neglected as a result of the slowly-varying envelope assumption. This set of equations admits solutions only for specific values of $\tilde{\beta}$ referred to as *eigenvalues*. Such solutions are called *eigenfunctions*. The pair eigenvalue and eigenfuction is referred to as a *propagation mode*.

The wave number $\tilde{\beta}$ can be obtained solving the eigenvalue equation in Eq. (2.8) and contains two terms

$$\tilde{\beta}(\omega) = \beta(\omega) + \Delta\beta. \tag{2.9}$$

The term $\Delta\beta$ includes the attenuation effect and the nonlinearity induced by P_{NL} in Eq. (2.1), and it is typically expressed as

$$\Delta\beta = j\frac{\alpha}{2} + \gamma|A|^2 \tag{2.10}$$

where α and γ are referred to as *attenuation coefficient* and *nonlinearity coefficient*, respectively. Separating the two components of the wave number in Eq. (2.8) and rearranging, we obtain

$$\frac{\partial \tilde{A}}{\partial z} = j[\beta(\omega) + \Delta\beta - \beta_0]\tilde{A}, \tag{2.11}$$

where $\tilde{\beta}^2 - \beta_0^2$ was approximated by $2\beta_0(\tilde{\beta} - \beta_0)$. The *linear* wave number $\beta(\omega)$ is not in general an analytical known function, as it depends on the fibre index profile. However, it can be expanded in Taylor series around ω_0 as

$$\beta(\omega) = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \frac{1}{6}\beta_3(\omega - \omega_0)^3 + \cdots \tag{2.12}$$

where β_i for $i = 0, 1, 2, \ldots$ represents the i-th derivative of $\beta(\omega)$ evaluated in $\omega = \omega_0$. Recalling the assumption on the quasi-monochromatic envelope, Eq. (2.12) can be truncated to the second-order without significant loss in accuracy. Substituting the truncated version of Eq. (2.12) into Eq. (2.11) and then switching back to the time domain[1], Eq. (2.11) becomes

---

[1] Going from the frequency domain to the time domain the operator $(\omega - \omega_0)$ is replaced by $j\frac{\partial}{\partial t}$.

$$\frac{\partial A}{\partial z} = -\beta_1 \frac{\partial A}{\partial t} - \frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{\alpha}{2}A + j\gamma|A|^2 A \tag{2.13}$$

where the right-hand side of Eq. (2.10) was used for $\Delta\beta$. Eq. (2.13) is commonly referred to as nonlinear Schrödinger equation (NLSE)[2] and it is considered to be the key equation to understand fibre propagation effects.

2.1.2 Pulse propagation effects

Based on Eq. (2.13) it is possible to recognise three main propagation effects: signal attenuation, chromatic dispersion and nonlinear phase shift. These effects are crucial in the design of an optical fibre system and particularly in the design of ad-hoc receivers for the optical fibre channel, as discussed in chapters 3 and 4.

Each of these effects can be studied separately by isolating, in turn, one of the three terms on the right-hand side of Eq. (2.13) and forcing the residual two terms to zero. Although the combined action of these three terms leads to a qualitatively different solution, solving Eq. (2.13) in these three cases gives a strong physical insight on how pulse propagation through an optical fibre works.

The three particular cases are described in the following:

1. *Attenuation*

   $$-\frac{\alpha}{2}A \neq 0, \;\; -\frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t^2} = 0, \;\; j\gamma|A|^2 A = 0.$$

   In this case the solution of Eq. (2.13) (with an initial condition at $z = 0$ given by $A(0,t)$) is trivial and is given by

   $$A(z,t) = A(0,t)\exp\left(-\frac{\alpha}{2}z\right). \tag{2.14}$$

   Since α is a real-valued coefficient, Eq. (2.14) indicates an attenuation of the optical field and, thus, of the transmitted optical power along the propagation path. The attenuation term on the right-hand side of Eq. (2.14) can be easily obtained even when either is non-zero.

2. *Chromatic dispersion*

   $$-\frac{\alpha}{2}A = 0, \;\; -\frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t^2} \neq 0, \;\; j\gamma|A|^2 A = 0.$$

   In this case Eq. (2.13) can be easily solved by switching to the Fourier domain where it can be rewritten as

---

[2] A mathematically alike equation was derived by Schrödinger, although for the description of the well-known quantum wave-function, hence the name.

$$\frac{\partial \tilde{A}}{\partial z} = \left(-j\beta_1\omega + \frac{j\beta_2\omega^2}{2}\right)\tilde{A} \tag{2.15}$$

whose solution is given by

$$\tilde{A}(z,\omega) = \tilde{A}(0,\omega)\exp\left[\left(-j\beta_1\omega + \frac{j\beta_2\omega^2}{2}\right)z\right]. \tag{2.16}$$

The solution in Eq. (2.16) indicates a frequency-dependent phase-shift which is referred to as *chromatic dispersion* (CD). The physical meaning of β_1 and β_2 can be understood from Eq. (2.16). Indeed, β_1 induces a delay in the propagation of the wave group around the frequency ω_0 (phase shift linearly varying with ω). Particularly, β_1 specifies the velocity at which a wave group propagates through the fibre, hence the name *group velocity* (GV). The β_2 coefficient instead indicates the velocity at which different wave groups around ω_0 travel compared to each other (quadratically dependent phase shift). For this reason, β_2 is commonly referred to as group velocity dispersion (GVD). The propagation delay between different groups results in a broadening of the pulse amplitude profile (see [1, Ch. 3]). This broadening can be shown to accumulate linearly with the transmission distance and quadratically with the signal bandwidth. CD is a key phenomenon in optical fibre transmission as it causes different pulses to interact and interfere with each other. This effect is called inter-symbol interference (ISI) and severely impairs transmission particularly when pulses are transmitted at high rate.

ISI also leads to what it is typically referred to as a channel with *memory*. Channel memory can be defined as the amount of time (or symbol periods) over which a pulse effectively extends due to the action of the channel, and its interference on other pulses is still significant. A more rigorous definition of channel memory is given in section 4.2.

Although, as discussed in 2.4.2, CD can be easily compensated through DSP, CD still interacts with nonlinear effects producing ISI that needs to be taken into account at the receiver (see chapter 4).

An example of how transmitted pulses are broadened by the effect of CD is illustrated in Fig. 2.1. The propagation of a Gaussian pulse is numerically simulated through a variable length fibre section. The duration of the pulse was set to $\sigma = 16$ ps, corresponding to a symbol rate of 32 GBaud which is the one adopted for the performance studies in chapter 3 and 4. Each pulse is normalised to unitary energy and fibre γ parameter was set to zero. It can be seen that the profile of the transmitted pulse is substantially broadened already

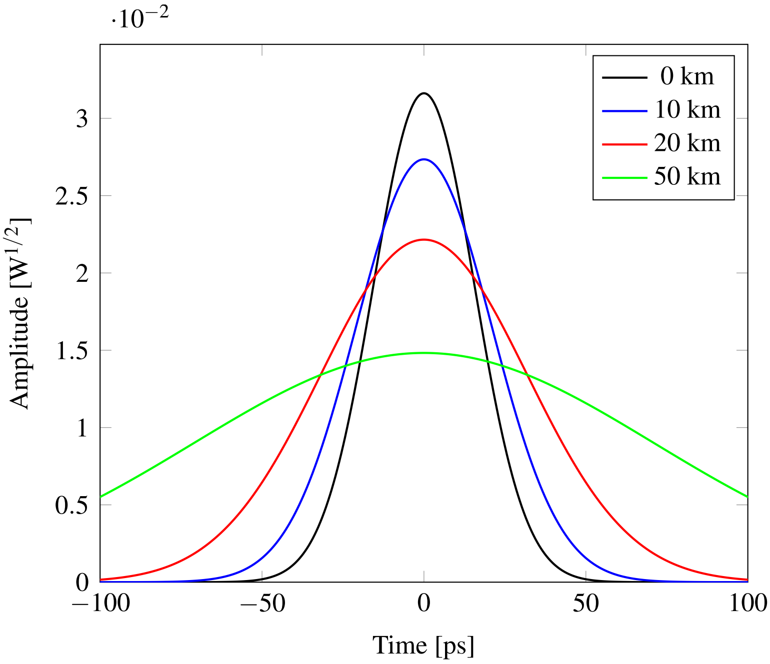


Figure 2.1: Dispersion of a (normalised) Gaussian pulse (σ=16 ps) during propagation through an optical fibre with α=0.2 dB/km, D=17 ps/(nm·km) and γ=0.

after propagating a few tenths of kilometers. The β_1 coefficient produces a delay between the transmission instant and the arrival time of a pulse travelling between transmitter and receiver. Therefore it is customary to replace in Eq. (2.13) the time variable t with $t' = t - \beta_1 z$, such that, in the new time reference, the propagating pulse is always centered around the time instant $t' = 0$. Thus, more conveniently, Eq. (2.13) can be rewritten as

$$\frac{\partial A}{\partial z} = -\frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t'^2} - \frac{\alpha}{2}A + j\gamma|A|^2 A \tag{2.17}$$

which is the most typical form of the NLSE.

3. *Nonlinear phase shift*

$$-\frac{\alpha}{2}A = 0, \quad -\frac{j\beta_2}{2}\frac{\partial^2 A}{\partial t^2} = 0, \quad j\gamma|A|^2 A \neq 0.$$

In this case the NLSE reduces to

$$\frac{\partial A}{\partial z} = j\gamma|A|^2 A \tag{2.18}$$

which, differently from the previous cases, is still a nonlinear differential equation. In can be shown that ([1, Ch. 3]) a solution of Eq. (2.18) can be found in the time domain in the form

$$A(z,t) = U(z,t)\exp\left[j\phi_{\mathrm{NL}}(z,t)\right] \tag{2.19}$$

where $U(z,t) \triangleq |A(z,t)|$ and $\phi_{\mathrm{NL}}(z,t) \triangleq \mathrm{Arg}\,(A(z,t))$. Substituting Eq. (2.19) into Eq. (2.18) yields the following set of differential equations

$$\frac{\partial U}{\partial z} = 0, \tag{2.20}$$

$$\frac{\partial \phi_{\mathrm{NL}}}{\partial z} = \frac{U^2}{L_{\mathrm{NL}}}, \tag{2.21}$$

where L_{NL} is the so-called nonlinear length and is defined as

$$L_{\mathrm{NL}} \triangleq \frac{1}{\gamma P_0} \tag{2.22}$$

where $P_0 = \max(|A(0,t)|^2)$ is the transmitted pulse peak power.

The set of Eqs. (2.20) and (2.21) can be solved analytically and the solution is given by

$$A(z,t) = |A(0,t)| \exp\left[j(\phi_{\mathrm{NL}}(z,t)\right], \tag{2.23}$$

$$\phi_{\mathrm{NL}}(z,t) = \frac{|A(0,t)|^2}{L_{\mathrm{NL}}} z + \phi_{\mathrm{NL}}(0,t). \tag{2.24}$$

The solution shows that the amplitude of the pulse does not change as the pulse propagates[3] but a time-varying phase-shift is induced on the pulse, which is dependent on the pulse instantaneous transmitted power $|A(0,t)|^2$. This nonlinear phenomenon is sometimes referred to as *self-phase modulation* (SPM) due to this action of the signal instantaneous power of modulating its phase.

As mentioned before, the three propagation effects discussed above combine to produce a more complicated evolution of the transmitted pulse. Particularly, the interplay between chromatic dispersion and nonlinearity generates a joint phase/amplitude distortion which cannot in general be analytically described.

An analytical solution of NLSE is in general not known. In order to theoretically address the analysis of fibre propagation effects it is possible to: i) find approximated analytical solution of NLSE using perturbation methods (see section 2.5); ii) solve Eq. (2.17) using numerical methods, such as the split-step Fourier method (SSFM), which is discussed in section 2.6.

2.1.3 Classification of nonlinear Kerr effects

The nonlinear term in Eq. (2.17) is responsible for the so-called Kerr effect [1]. High-data rate optical transmission uses WDM channels, i.e. parallel data streams are carried

---

[3] When attenuation is also present, its only effect on the pulse amplitude is given by Eq. (2.14).

by different wavelengths, spaced apart by a given frequency spacing to avoid channel crosstalk. Thus, traditionally, it has been of great interest to classify the effects of the Kerr term in Eq. (2.17) based on which channels are involved in the generation of such effects. A first broad categorisation can be made by separating nonlinear effects based on whether or not channels other than the one of interest are involved in their generation. We therefore refer to either *inter-channel* nonlinearity and *intra-channel* nonlinearity. Furthermore, different kind of nonlinear interactions (still within the scope of the Kerr effect) are possible when two or more channels co-propagate in a nonlinear optical fibre.

In order to identify these different effects, let us assume that the transmitted optical field is given by

$$A = A_0 + A_1 + A_2 \tag{2.25}$$

where A_0, A_1, and A_2 are the (scalar) complex envelopes of the transmitted channels at three different wavelengths and A_0 is assumed to be the channel of interest. Replacing Eq. (2.25) into Eq. (2.17) we find

$$\frac{\partial(A_0 + A_1 + A_2)}{\partial z} = -\frac{j\beta_2}{2}\frac{\partial^2(A_0 + A_1 + A_2)}{\partial t^2} + j\gamma|A_0 + A_1 + A_2|^2(A_0 + A_1 + A_2) \tag{2.26}$$

where the attenuation term has been dropped for simplicity of notation. A standard approach consists in assuming a small nonlinearity (perturbative approach) [1], hence Eq. (2.26) can be rewritten as a set of three differential equations. The equation relative to the channel of interest reads as

$$\frac{\partial A_0}{\partial z} = -\frac{j\beta_2}{2}\frac{\partial^2 A_0}{\partial t^2} + \underbrace{j\gamma|A_0|^2 A_0}_{\text{SPM}} + \underbrace{2j\gamma(|A_1|^2 + |A_2|^2)A_0}_{\text{XPM}} + \underbrace{j\gamma(A_1^2 A_2^*)}_{\text{FWM}}. \tag{2.27}$$

As shown in Eq. (2.27), the original nonlinear phase shift acting on the channel of interest is split into three components: i) a term dependent only on the power of the channel of interest itself, hence called self-phase modulation (SPM); ii) a term dependent only on the power of the the interfering channels called cross-phase modulation (XPM); iii) a term dependent on a cross-product between the two interfering channels called four-wave mixing (FWM). The derivation above can be repeated for three or more channels. However, the arising terms are similar to the ones in Eq. (2.27). The only difference consists of the XPM term adding up multiple channels, and additional pairwise FWM cross-products arising with more channels. While both SPM and XPM depend on the power of the channels, the FWM effect is dependent on the actual optical field. Due to this feature, the phase relationship between the co-propagating channels is crucial for the accumulation of the FWM.

The derivation above can be repeated on a frequency component basis, i.e. assum-

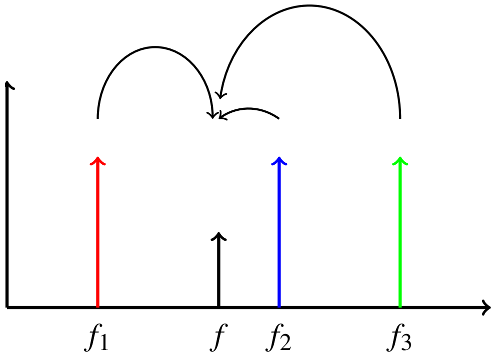


Figure 2.2: Illustration of the generation of FWM at frequency f, due to the CW optical fields at frequencies f_1, f_2 and f_3.

ing the propagating field is composed of four continuous-wave (CW) components at frequencies f, f_1, f_2 and f_3. As illustrated in Fig. 2.2, CW optical fields at frequency f_1, f_2 and f_3 interact with each other to form a new wave at frequency f. The phenomenon is possible whenever the relationship between the four frequencies is [2]

$$f = f_1 + f_2 - f_3. \tag{2.28}$$

Defining $A(z) \triangleq A(f,z)$ and $A_j(z) \triangleq A(f_j,z)$ and solving Eq. (2.17) in the Fourier domain for the field component at frequency f it can be shown that[4]

$$A(z) = jd\frac{(2\pi)^2 f}{nc}A_1^* A_2 A_3 \exp\left(-\frac{\alpha}{2}z\right)\frac{1-\exp\left(j\Delta\beta z - \alpha z\right)}{\alpha - j\Delta\beta} \tag{2.29}$$

where c is the speed of light, d is a degeneracy coefficient either equal to 3 for $f_1 = f_2 \neq f_3$ or to 6 for $f_1 \neq f_2 \neq f_3$, n is the refractive index (as a function of the frequency f), and

$$\Delta\beta = \beta(f_1) + \beta(f_2) - \beta(f_3) - \beta(f). \tag{2.30}$$

From Eq. (2.29) it can be observed that the amplitude of the FWM strongly depends on $\Delta\beta$. For $\Delta\beta \ll 1$, the rightmost fraction in Eq. (2.29) significantly increases in magnitude and, as a result, the FWM product is amplified. The condition where $\Delta\beta$ is close to zero is commonly referred to as *phase matching condition* and it is typically met in systems where the dispersion parameter is low. As a result, in scenarios where transmission is performed far from the zero-dispersion point, and standard single mode fibre is used, FWM can be considered as a minor nonlinear effect compared to SPM and XPM. In chapters 3 and 5, the performance of receivers compensating for intra-channel nonlinearity (SPM) is analysed, as well as compensation of inter-channel nonlinearity, including both XPM and FWM.

---

[4] In order to obtain Eq. 2.29, a perturbative approximation must be adopted, which in the FWM literature is commonly referred to as *undepleted pump* assumption [1, Ch. 10].

2.1.4 Polarisation effects

In the section above, the optical field was assumed to stay polarised during propagation, i.e. aligned to a given polarisation vector. Also, the fibre was assumed to be perfectly isotropic, meaning that no preferred polarisation orientation exists. As a result of that, the wave equation in Eq. (2.1) can be reduced from a vector equation to a scalar one.

However, a real optical fibre can be considered only isotropic at a local level. Perfect isotropy is preserved when the fibre is an ideal cylindrical waveguide, in which case each mode of polarisation is a degenerate solution of Eq. (2.1), i.e. they show the same transversal distribution $F(x,y)$ and wave number $\beta(\omega)$.

In reality, due to the fabrication process and stress applied along a fibre cable, the fibre itself shows a slight asymmetry of its transversal geometry, which can be approximated as an ellipse (instead of a circle). As a result, two *non-degenerate* modes of polarisation, corresponding to the two ellipse axes, are found as the solution of Eq. (2.1). Because of the difference between the wave numbers experienced, these two *preferred* modes of polarisation show two different effective refractive indices[5] n_x and n_y. This phenomenon is thus commonly referred to as fibre *birefringence*. Fibre birefringence is typically measured by $|n_x - n_y|$, and it is typically very small. However, over a sufficiently long distance this effect accumulates and can become a significant source of impairment for light-wave transmission. Luckily, the orientation of the birefringence axes also varies randomly along the transversal direction z, so reducing the overall accumulated birefringence.

The main effects due to fibre birefringence on the propagating signal are: i) random polarisation evolution of a continuous-wave (CW) optical field; ii) polarisation mode dispersion (PMD). The latter is discussed in depth in section 2.2. The former effect results in random evolution of the polarisation during propagation which statistically repeats itself after a given distance. Such a spatial period is called *beat length* and it is related to the strength of the fibre birefringence and to the evolution of the fibre birefringence axes.

Clearly, in the presence of birefringence, the propagation problem cannot be reduced to a scalar problem. The NLSE can be therefore re-derived from Eq. (2.1) under the assumptions discussed in 2.1 exception made for the assumption of a polarised signal. Instead, two preferred modes of polarisation must be considered coupled at any given fibre section. In the fixed birefringence case, this leads to the so-called coupled NLSE (CNLSE) [1, Ch. 6], [3], [4], [5] which is given by

---

[5] The axis with a lower index is called *slow axis* while the one with a higher index is called *fast axis*.

$$\begin{aligned} \frac{\partial A_x}{\partial z} - \frac{j\beta_2}{2}\frac{\partial^2 A_x}{\partial t^2} - j\gamma\left(|A_x|^2 + \frac{2}{3}|A_y|^2\right)A_x = \frac{j\gamma}{3}|A_y|^2 A_x^* \exp\left(-2j\Delta\beta z\right) \\ \frac{\partial A_y}{\partial z} - \frac{j\beta_2}{2}\frac{\partial^2 A_y}{\partial t^2} - j\gamma\left(|A_y|^2 + \frac{2}{3}|A_x|^2\right)A_y = \frac{j\gamma}{3}|A_x|^2 A_y^* \exp\left(2j\Delta\beta z\right) \end{aligned} \tag{2.31}$$

where the functions $A_x(z,t)$ and $A_y(z,t)$ are defined by

$$\mathbf{E}(z,t) = F(x,y)\left[A_x(z,t)\hat{\mathbf{x}} + A_y(z,t)\hat{\mathbf{y}}\right] \tag{2.32}$$

and

$$\Delta\beta = \beta_1^x - \beta_1^y, \tag{2.33}$$

while β_1^x and β_1^y are the GVs of x and y polarisations, respectively.

Note that here $\hat{\mathbf{x}}$, and $\hat{\mathbf{y}}$ represent a generic pair of states of polarisation (SOPs), i.e. they are complex vectors describing the time-domain oscillation of the optical field along the fibre axes (which are in turn real vectors) [6]. It is also customary to represent a SOP as a 3D real vector moving along a sphere called Poincaré sphere [6].

Under the assumption of fast birefringence, each of the two polarisations quickly moves around the Poincaré sphere. In this case, it can be shown that [5], [7] the nonlinear terms on the right-hand side of the two equations in Eq. (2.31) quickly averages to zero. As a result of the averaging of the nonlinear term in Eq. (2.31) one obtains

$$\frac{\partial \mathbf{A}}{\partial z} = -\frac{\alpha}{2}\mathbf{A} + \frac{j\beta_2}{2}\frac{\partial^2 \mathbf{A}}{\partial t^2} + j\frac{8}{9}\gamma|\mathbf{A}|^2 \tag{2.34}$$

which is commonly referred to as the Manakov equation [5]. The Manakov equation describes the average effect of the birefringence axes rotation on the nonlinear propagation phenomena. This allows to obtain an equivalent description of the interaction birefringence/nonlinearity, avoiding stochastic descriptions of the fibre polarisation evolution. Both Manakov and CNLSE equations were shown to return essentially identical results when SOPs move on the Poincare' sphere faster (in space) than the nonlinear Kerr effect. Also, in such case Manakov equation can be numerically solved with a coarser integration step than the CNLSE, thus enabling a more efficient simulation of the optical fibre propagation. For this reason, in this thesis the Manakov equation in (2.34) is used instead of the CNLSE.

However, when PMD is included, the description given by Eq. (2.34) is not sufficient and additional terms are required. This leads to the so-called Manakov-PMD equation, as shown in section 2.2.

2.2 PMD theory

In this section a short theory of PMD is developed. Illustrating the physical mechanism underlying PMD and its stochastic nature will help to understand its detrimental impact on nonlinearity compensation schemes which are studied in section 3.7.

In the previous section, we discussed constant birefringent fibres, for which propagation of lightwave pulses can be described using Eq. (2.31). When the fibre birefringence axes are rapidly rotating but with a birefringence which is constant for all frequencies of the transmitted signal, Eq. (2.34) can be used to capture the average nonlinearity accumulated by a CW signal whose SOP uniformly evolves on the Poincare' sphere. It is worth noticing that the Manakov equation assumes that the transmitted signal does not *depolarise*, i.e. all signal frequencies have SOPs which are assumed aligned in their movement on the Poincare' sphere.

On the other hand, when the GVs β_1^x and β_1^y over the two fibre local axes of birefringence assume different values, and such axes rotate along the fibre, PMD arises. This can be understood by analysing the 2x2 matrix describing the polarisation transformations of signals propagating through an optical fibre, which is referred to as Jones matrix [1, Ch. 2]. Let us assume that a fibre section is made up of many sections with different local birefringence axes with different GVs. The transformation between the input optical field of each section and its output can be described by the 2x2 frequency dependent matrix

$$\mathbf{T}_k = \mathbf{R}_k \cdot \mathbf{D}_k(\omega) = \begin{pmatrix} e^{-j\frac{\varphi_k}{2}}\cos\theta_k & e^{-j\frac{\varphi_k}{2}}\sin\theta_k \\ e^{j\frac{\varphi_k}{2}} - \sin\theta_k & e^{j\frac{\varphi_k}{2}}\cos\theta_k \end{pmatrix} \cdot \begin{pmatrix} e^{-j\frac{\tau_k}{2}\omega} & 0 \\ 0 & e^{j\frac{\tau_k}{2}\omega} \end{pmatrix} \tag{2.35}$$

where φ_k and θ_k denote a random phase shift and angle difference between the birefringence axes of the $(k-1)$-th and the k-th section, respectively, and

$$\tau_k \triangleq \frac{|\beta_{1k}^x - \beta_{1k}^y|}{2}\Delta z \tag{2.36}$$

represents the delay between two pulses travelling over the local birefringence axes of the k-th section of length Δz. The random processes φ_k and θ_k define the stochastic evolution of the SOP over the Poincare' sphere at frequency $\omega = 0$.

Over N sections we then obtain, in the absence of any other propagation effect (CD, nonlinearity, etc.),

$$\mathbf{J}(\omega)\mathbf{A}_{in}(\omega) = \mathbf{A}_{out}(\omega) \tag{2.37}$$

where

$$\mathbf{J}(\omega) = \prod_{k=1}^{N} \mathbf{T}_k(\omega) = \begin{pmatrix} u_1(\omega) & u_2(\omega) \\ -u_2^*(\omega) & u_1^*(\omega) \end{pmatrix} \tag{2.38}$$

is the Jones matrix of the cascade of fibre sections, and u_1 and u_2 are generic functions of frequency representing a frequency-dependent random transformation. In the absence of polarisation-dependent loss, such transformation is unitary and therefore u_1 and u_2 are such that [8]

$$|u_1|^2+|u_2|^2=1. \tag{2.39}$$

Eqs. (2.37), (2.38) therefore indicate a different (random) polarisation rotation of the frequency components of the signal that hence depolarises. This effect is schematically illustrated in Fig. 2.3 which shows the random evolution of the SOPs of three frequency components of a modulated signal under the effect of PMD. At section z=0 of the optical fibre the signal is linearly polarised and, thus, all frequency components show the same SOP. During propagation PMD acts as a frequency dependent Jones matrix at each fibre section. As a result, the SOPs of each frequency component will evolve independently yielding a random output SOPs configuration ($z=L$).

Alternatively, [8] showed that the transformation in Eq. (2.37) can be more intuitively understood by choosing a particular pair of orthogonal polarisation states called *principal states of polarisation* (PSP). Signals propagating over the two PSPs, to the first-order, is mapped onto a different pair of PSPs (or output PSPs) without any depolarisation but only with a relative delay. In other words, $\mathbf{J}(\omega)$ allows two orthogonal eigen-vectors $\mathbf{A}_{\text{in}}^{+}$, $\mathbf{A}_{\text{in}}^{-}$ (and relative eigen-values), such that the corresponding output SOPs $\mathbf{A}_{\text{out}}^{+}$, $\mathbf{A}_{\text{out}}^{-}$ verify

$$\frac{\partial \mathbf{A}_{\text{out}}^{\pm}}{\partial \omega}=0. \tag{2.40}$$

For the two PSPs we then have

$$\mathbf{A}_{\text{out}}^{\pm} \approx \begin{pmatrix} e^{-j\frac{\varphi}{2}}\cos\theta & e^{-j\frac{\varphi}{2}}\sin\theta \\ -e^{j\frac{\varphi}{2}}\sin\theta & e^{j\frac{\varphi}{2}}\cos\theta \end{pmatrix} \cdot \begin{pmatrix} e^{-j\frac{\tau}{2}\omega} & 0 \\ 0 & e^{j\frac{\tau}{2}\omega} \end{pmatrix} \mathbf{A}_{\text{in}}^{\pm}. \tag{2.41}$$

θ here reflects the relative angle between the input and the output PSPs, whereas τ

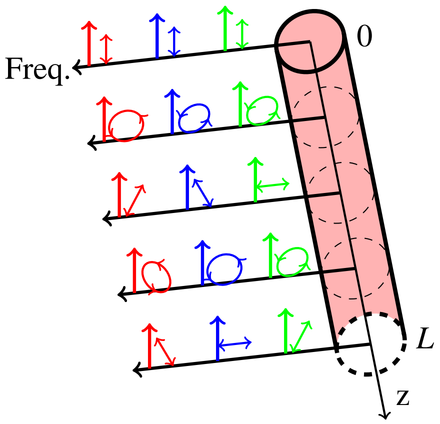


Figure 2.3: Evolution of the SOPs of different signal frequency components in the presence of PMD.

is the overall delay experienced by two pulses transmitted over the input PSPs, and is referred to as differential group delay (DGD). Although the dependency on the distance z was dropped for simplicity of notation in Eq. (2.41), the overall DGD is clearly a stochastic process of the variable z which can be modelled as a Wiener process [9]. At a given fibre section, τ is then a random variable resulting from the combination of multiple τ_k random variables and subsequent random rotations. It can be shown that [9, 10], as long as many τ_k add up into τ then, regardless of the stochastic behaviour of φ_k and θ_k, the probability density function (pdf) of τ at the end of a fibre section of length z is given by

$$p(\tau,z)=\frac{2\tau^2}{\sqrt{2\pi}q^3}\exp\left[-\left(\frac{\tau^2}{2q^2}\right)\right] \tag{2.42}$$

where $q^2 =<\tau_k>^2 \frac{z}{L_c}$, and L_c is the length over which the PSPs can be considered correlated and corresponds to the length of a PMD section.

The pdf in Eq. (2.42) is a so-called Maxwellian distribution, whose parameters are given by

$$<\tau> = \sqrt{\frac{8}{\pi}} <\tau_k>^2, \tag{2.43}$$

$$\sigma = \sqrt{\frac{3\pi-8}{\pi}}\sqrt{\frac{z}{L_c}}. \tag{2.44}$$

The validity of this pdf to describe the statistical behaviour of the DGDs can be confirmed numerically. In Fig. 2.4, a histogram of multiple occurrences of the DGD at the output of a fibre section of length $z=1000$ km is shown. The results are obtained using the numerical method described in section 2.6.3, with L_c =100 m which yields 10^4 PMD sections. The histogram, obtained from 2×10^4 random realisations, shows very good agreement with the pdf in Eq. (2.42) which serves as a validation of the theory above. Furthermore, the Maxwellian distribution of the DGDs and the statistical models for PMD presented in [9, 10] have also been experimentally validated in several works, e.g. [10, 11]. The power of this description of the DGD lies in the fact that it allows to treat a random concatenation of birefringence sections as an equivalent single section. However, it is worth recalling that Eq. (2.41) does not mean that two signals aligned over the two input PSPs will not change SOP during propagation. Instead, it means that such signals, to the first-order, are aligned to the output PSPs of that fibre section with a relative delay. Also, as mentioned above, the approximation in Eq. (2.41) is true only to the first order. Thus, a signal launched on one of the PSP will also in general experience distortion due to second-order PMD effects, which are in turn reflected by the dependence of τ on ω. The Jones matrix in Eq. (2.38) indicates that, in general, PMD is present to all orders. However, second-order and higher-order PMD terms

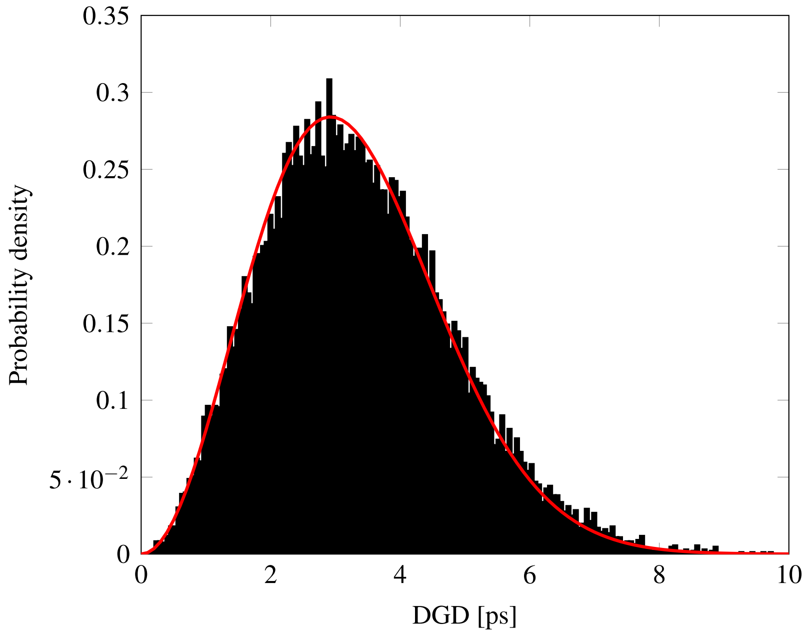


Figure 2.4: Histogram of the DGD values for z=1000 km and L_c =100 m and $2 \cdot 10^4$ realisations accumulated. Eq. (2.42) is shown in red.

can be neglected depending on how large the bandwidth of the propagating signal is compared to the PMD characteristic bandwidth that can be defined as $1/\tau$.

In the presence of PMD, the Manakov equation needs to be extended to account for the independent SOPs evolution for the frequency components of the propagating signal. This can be done starting from the CNLSE set of equations and dropping the assumptions of constant birefringence. Averaging again over the Poincare' sphere, the following equation can be obtained [5, 7]

$$\frac{\partial \mathbf{A}}{\partial z} + \frac{\alpha}{2}\mathbf{A} - \frac{j\beta_2}{2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - j\frac{8}{9}\gamma|\mathbf{A}|^2 = -b'\overline{\sigma}\frac{\partial \mathbf{A}}{\partial t} + \frac{1}{3}j\gamma(\hat{\mathbf{N}} - < \hat{\mathbf{N}} >) \tag{2.45}$$

where

$$b' = \frac{|\beta^x_{1k} - \beta^y_{1k}|}{2} \tag{2.46}$$

is the delay per unit length,

$$\overline{\sigma} = \mathbf{T^H}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\mathbf{T} \tag{2.47}$$

and $\mathbf{T}$ is a matrix describing the transformation of the SOPs as a function of the distance z (see [7]). Finally, $\hat{\mathbf{N}} - < \hat{\mathbf{N}} >$ is a nonlinear transformation defined in [7]. Eq. (2.45) is referred to as Manakov-PMD equation. The difference between this equation and the Manakov equation in (2.34) is represented by the two terms on the right-hand side of Eq. (2.45): the first term accounts for the so-called *linear* PMD effect which was described earlier; the second term represents a residual interaction between PMD and nonlinearity after the averaging process over the Poincare' sphere is performed. For this reason, such term is typically referred to as *nonlinear* PMD. However, if the SOP evolves quicker (in space) than the characteristic length over which nonlinear fibre effects occur, then nonlinear PMD averages to zero over the fibre nonlinear length and therefore can be discarded, leading to

$$\frac{\partial \mathbf{A}}{\partial z}+\frac{\alpha}{2}\mathbf{A}-\frac{j\beta_2}{2}\frac{\partial^2 \mathbf{A}}{\partial t^2}-j\frac{8}{9}\gamma|\mathbf{A}|^2=-b'\overline{\sigma}\frac{\partial \mathbf{A}}{\partial t}. \tag{2.48}$$

In this thesis, Eq. (2.48) is used for the description of fibre propagation in the presence of PMD. In particular, it is used in section 3.7 for the evaluation of the effectiveness of nonlinearity compensation receiver schemes in the presence of PMD.

2.3 Structure of an optical fibre communication system

The architecture of an optical fibre link can vary depending on the transmission distance to be covered and throughput requirements. A major impact on the optical link structure comes from the selected amplification scheme and the fibres used in each span. The span losses, the amplification scheme and the specific fibre make a huge difference to the performance of the communication system. Indeed, different link architectures will result in markedly different channel properties and, as a result, the transceiver design will need to be tailored to that specific *optical channel*, i.e. to a specific link configuration.

In the following, a description of two different kinds of optical fibre link studied in this thesis is provided, along with a discussion of their main features from an optical transmission perspective.

2.3.1 Single-span fibre systems

Single-span systems are optical links connecting two nodes over a relatively extended reach without optical amplification in between. They are generally used for access network links or to connect offshore islands to the mainland and typically span distances in the range between 100 to 400 km. A typical example of a single-span system is illustrated in Fig. 2.5. It consists of a high-power transmitter including an optical amplifier, referred to as *booster* amplifier. No optical amplifiers are usually placed along the link. However, in the cases where the span length is considerable (e.g. 200≈400 km), optical amplifiers are placed at about 100 km before the receiver, albeit using a remote feeding. For this reason such amplifiers are referred to as remote optical pre-amplifiers. At the receiver, an optical *pre-amplifier* is used at the receiver to recover from the link attenuation. The attenuation experienced by the optical signal in these systems is usually

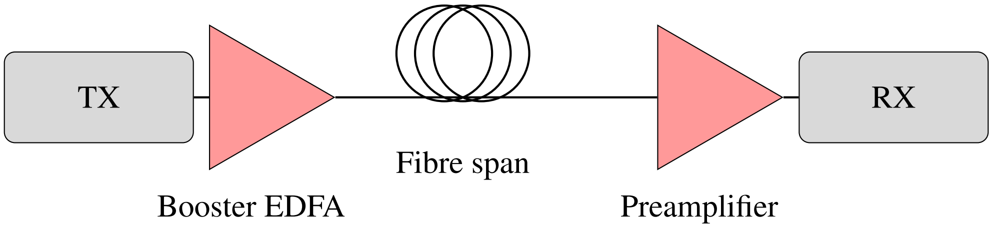


Figure 2.5: Schematic diagram of a typical single-span optical fibre system.

considerable (e.g. 80 dB for a 400 km span with 0.2 dB/km attenuation coefficient), and therefore represents the main limitation to transmission. Special fibres such as *ultra low-loss* fibres (attenuation coefficient below 0.17 dB/km) are usually adopted to mitigate this issue and to maximise the receiver optical signal-to-noise ratio (OSNR). Optical amplification is in most cases performed using an erbium-doped fibre amplifier (EDFA). EDFA amplifiers are based on the principle of *stimulated emission* [12, Ch. 6] of light by the dopants population which is achieved through an external *pump* laser. In EDFAs, the dopant material consists of erbium ions. The optical amplification process is accompanied by the generation of optical noise which is due to the *amplified spontaneous emission* (ASE) phenomenon [12, Ch. 6]. The power spectral density (PSD) of the ASE noise generated by an EDFA and coherently detected is given by [13]

$$N(\nu) = N_p(G-1)n_{\text{sp}}h\nu \tag{2.49}$$

where N_p is the number of detected polarisations, G is the amplifier gain, h is the Planck constant, ν is the frequency of operation and n_{sp} is the spontaneous emission factor which indicates the amount of spontaneous emission noise produced by an EDFA and the factor of 2 is due to the noise being in two polarisations. The *noisiness* of the EDFA is often indicated by a parameter called the *noise figure* (NF), which is defined as

$$\text{NF} = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} \tag{2.50}$$

and therefore represents the signal-to-noise ratio (SNR) degradation of an optical signal going through an EDFA. Spontaneous emission factor and noise figure are related by [14]

$$\text{NF} = 2n_{\text{sp}}\frac{G-1}{G} \approx 2n_{\text{sp}} \tag{2.51}$$

where the approximation in the right-hand side of Eq. (2.51) follows from the fact that $G \gg 1$.

If we neglect the effect of fibre nonlinearities and the noise introduced by the *booster* EDFA, the SNR of single span systems is then given by

$$\mathrm{SNR} = \frac{G \cdot P_{RX}}{P_{\mathrm{ASE}}} = \frac{P\exp(-\alpha L)G}{B(G-1)\mathrm{NF}\cdot h\nu} \approx \frac{P\exp(-\alpha L)}{B\cdot\mathrm{NF}\cdot h\nu} \tag{2.52}$$

where P_{RX} is the received optical power, P_{ASE} is the power of the ASE noise added by the EDFA over the channel bandwidth, P is the transmitted optical power, B is the signal bandwidth and L is the span length. From Eq. (2.52) it can be inferred that for a fixed transmitted power the SNR degrades quite dramatically (exponential decay) with the transmission distance, regardless of the pre-amplifier gain. When high-sensitivity receivers are adopted [6], the pre-amplifier can be avoided and the optical signal is detected at very low optical powers. In this case, the noise limitation is given by the *shot-noise*, which represents the unavoidable quantum uncertainty inherent to the photo-detection process [15]. It can be shown that [12], the shot-noise PSD is equivalent to the PSD introduced by an ideal EDFA with $n_{\mathrm{sp}} = 1$ (NF=3 dB) and thus the shot-noise limited SNR in single-span systems is given by

$$\mathrm{SNR} = \frac{P\exp(-\alpha L)}{2h\nu B}. \tag{2.53}$$

In conclusion, in order to operate a single span system at an SNR value that guarantees error-free transmission for a given net data rate, the transmitted power is required to increase exponentially with the transmission distance (and linearly increasing with the transmitted bandwidth). This requirement, in turn, forces to operate transmission in a highly nonlinear regime. As a result, single-span systems are impaired by nonlinear distortions as the transmission distance increases. Because of the high powers required (in excess of 15 dBm per channel), nonlinear effects in single-span systems are even more prominent than in other fibre systems where the optical signal is repeated, and therefore receiver schemes need to be properly adapted to this condition. This topic is tackled in chapter 4.

2.3.2 Multi-span fibre systems

Multi-span EDFA-amplified systems represent the vast majority of optical fibre communication systems. As illustrated in Fig. 2.6, such systems comprise multiple optical amplification nodes before the signal is electrically regenerated (receiver node). The amplification length is therefore much shorter than of a typical single-span system and it varies between 40 and 120 km [13]. This amplification length guarantees a much higher receiver SNR values than in a single span case (for the same transmission distance), at the expense, of course, of the deployment of additional EDFAs. The gain of each EDFA is commonly chosen to guarantee a perfect compensation of the span losses

---

[6] Such are receivers where the impact of the electrical front-end noise can be considered small due to extremely sensitive photo-diodes, e.g. the *avalanche* photo-diode.

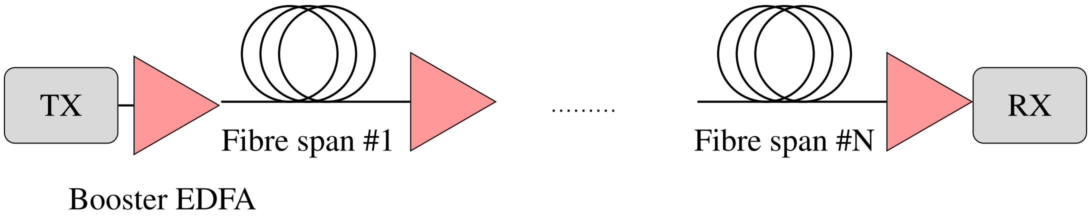


Figure 2.6: Schematic diagram of a typical multi-span EDFA-amplified optical fibre system.

(*transparent link* condition). Under this assumption, the receiver SNR (neglecting again nonlinear effects) for a multispan EDFA system can be written as

$$\mathrm{SNR} = \frac{P}{N_s P_{\mathrm{ASE}}} \approx \frac{P}{2N_s G n_{\mathrm{sp}} h\nu} = \frac{P}{2N_s \exp(\alpha L_s) n_{\mathrm{sp}} h\nu} \tag{2.54}$$

where N_s is the number of fibre spans in the link and L_s is the span length. As can be observed from Eq. (2.54), in a multi-span system the SNR decreases only linearly with distance, as opposed to the previously discussed single-span case. However, this assumes that L_s is kept fixed, i.e. the number of amplifiers must increase linearly with the transmission distance. Adding more amplifiers for a given transmission distance (i.e. decreasing the span length), therefore, improves the received SNR. However, when nonlinear effects are taken into account through some specific nonlinearity management criterion an optimum span length can be identified [16, 17].

Legacy multi-span systems employ in-line compensation of CD through dispersion compensating fibres. After the introduction of coherent detection in the mid-2000s, in-line CD compensation was gradually abandoned because of the availability of DSP techniques able to compensate CD electronically. Also, dropping in-line optical CD was demonstrated to improve coherent transmission performance since dispersion helps to mitigate nonlinear fibre effects [13, 18, 19].

The lower amount of power required to guarantee a certain SNR level compared to the single-span case, is beneficial for the performance of multi-span systems which can operate in a *pseudo-linear* transmission regime. However, a specific characteristic of multi-span systems is the significant generation of signal-ASE nonlinear interactions, due to the accumulation of ASE noise along the link. This is addressed in detail in section 2.5.3.

Dispersion-unmanaged, multi-span systems using EDFA amplification represent the main solution for medium to long-haul optical fibre links and, thus, their transmission performance is studied in chapters 3 and 5.

2.4 High-spectral efficiency optical fibre communication systems

In the previous sections, an overview of the general physical properties of an optical fibre system was given. This section is focused on the techniques that can enable higher SE optical transmission compared to the older generation of intensity-modulation, direct-detection (IMDD) systems. These techniques include optical generation of 4D modulation formats, spectral shaping and coherent detection.

2.4.1 High-SE optical transmitters

The advent of the new coherent detection era has opened the door for new formats of modulation able to encode information on the four available degrees of freedom of an optical field: in-phase (I), and quadrature (Q) components, and two orthogonal SOPs. This resulted in a major SE improvement, as compared to IMDD systems, where only one degree of freedom (the field intensity) is exploited. Several attempts were made in the past to increase SE using IMDD systems [20] and research on this is still ongoing. However, increasing the cardinality M of the modulation format through a higher number of dimensions significantly improves the sensitivity in the uncoded regime (minimum Euclidean distance) for the same nominal SE. Moreover, for the classical additive white Gaussian noise (AWGN) channel [13, Sec. III], this is also true for coded systems. Indeed, in this case, the net rate of coded systems using 2D modulation formats is in general higher than 1D formats with the same cardinality and same SNR [13, Sec. IV].

Let us assume that the baseband equivalent of the transmitted signal is in the form (linear modulation)

$$s(t) = \sum_{n=-\infty}^{\infty} a_n p(t - nT_s) \tag{2.55}$$

where $a_n \in \mathcal{S} = \{s_1, s_2, ..., s_M\}$ is the symbol transmitted at time slot n and it is complex number drawn from the constellation $\mathcal{S}$, $p(t)$ is the transmitted pulse and T_s is the symbol period. The nominal SE of such a modulated signal can be defined as

$$\mathrm{SE} \triangleq \frac{R_s}{B} \log_2 M \qquad [\mathrm{bit}/(\mathrm{s} \cdot \mathrm{Hz})] \tag{2.56}$$

where $R_s = 1/T_s$ is the symbol rate, B is the bandwidth of the transmitted pulse $p(t)$ (typically defined at 3 dB). The nominal SE represents the *maximum* transmission rate per unit bandwidth at which information can be transferred using a specific modulation format and a given pulse. It is worth noting that the nominal SE does not always correspond to the actual information per unit bandwidth that is transferred through the channel. The latter quantity is instead referred to as *net* SE, and depends on the coding scheme adopted and on the quality of the channel.

The nominal SE can be increased using two parameters of the modulator: the constellation cardinality M and the ratio $\rho \triangleq R_s/B$. When a pulse $p(t)$ is selected, the ratio ρ is a fixed number that indicates its spectral confinement. The Nyquist criterion [21, Ch. 2] sets an upper limit on ρ which, in the absence of ISI, cannot be greater than one[7]. Transmission at a symbol rate $R_s = B$ is said to be at Nyquist rate. In general, ρ can be quite a small number. In order to increase such number the transmitted signal must be carefully shaped in order for its spectrum to meet specific requirements. Techniques aiming to contain the spectral shape of $p(t)$ for a given R_s and thus to maximise ρ are referred to *spectral shaping*. The most spectrally-efficient pulses are members of the well-known family of Nyquist pulses, also called root-raised cosine (RRC) pulses [21, Ch. 7].

For an AWGN channel, the optimal signal detection is represented by a *matched-filter* followed by a sampler at 1 Sa/sym [22, Ch. 7]. Therefore, in order for the sampler output to be ISI-free, the spectrum of the matched filter output should be given by

$$Q(f) = \begin{cases} T_s, & |f| \geq \frac{1-\alpha}{2T_s} \\ \frac{T_s}{2}\left\{1 - \cos\left[\frac{\pi T_s}{\alpha}\left(f - \frac{1+\alpha}{2T_s}\right)\right]\right\}, & \frac{1-\alpha}{2T_s} \leq |f| \geq \frac{1+\alpha}{2T_s} \\ 0, & |f| \geq \frac{1+\alpha}{2T_s} \end{cases} \tag{2.57}$$

where α is called *roll-off* of the pulse and can vary between 0 and 1. The spectral shape in Eq. (2.57) is referred to as *raised-cosine*, and since

$$Q(f) = P(f)P(f)^* \tag{2.58}$$

hence the name RRC for a pulse $p(t)$ whose normalised spectrum is given by $P(f)$.

The roll-off is an indicator of spectral confinement of $p(t)$. Small values of the roll-off are related to higher values of ρ and, therefore, higher SE, e.g. when $\alpha = 0$ then $\rho = 1$.

High-SE optical transmitters are, therefore, based on a combination of both an increased number of signal dimensions and spectral shaping. The modulation of an optical carrier in both I and Q components is performed by optical IQ modulators. The structure of an IQ modulator is shown in Fig. 2.7. The optical CW carrier is split over two branches each containing a so-called Mach-Zehnder modulator (MZM). The MZM, shown in Fig. 2.7, is composed of two arms introducing a phase shift ϕ controlled by an electrical driving signal. The two optical paths recombined with different phases give rise to an interference with an amplitude that varies based on the phase ϕ, thus allowing

---

[7] B here represents the double-sided bandwidth.

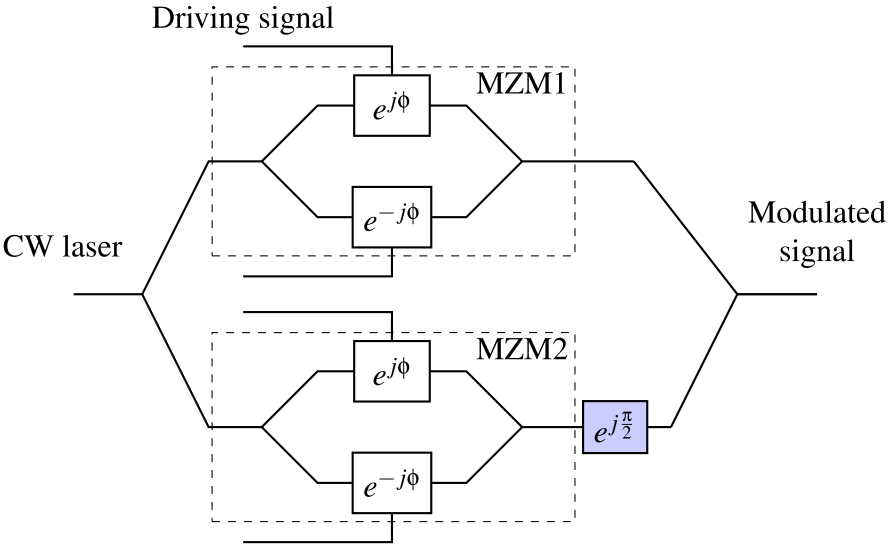


Figure 2.7: Schematic diagram of an IQ modulator.

to modulate the amplitude of the input CW laser. In an IQ modulator, the two MZMs modulate the I and Q channels. The Q channel is then subsequently phase-shifted by 90 degrees and the two optical path are finally recombined to generate a 2D modulated signal. In order to generate a polarisation-multiplexed (PM) signal, the IQ modulator structure is replicated over two distinct branches in which the two orthogonal SOPs are modulated, as illustrated in Fig. 2.8. The SOP in one of the two arms is then rotated onto an orthogonal SOP using a polarisation rotator. Such a device is referred to as a dual-polarisation (DP) optical modulator.

The spectral shaping of the pulse can be performed either optically [23–25] or electrically [26, 27]. The optical approach can be based on the design of sharp optical filters [23, 24], parametric amplification techniques [25], or cascaded MZM structures [28]. The advantage of such techniques is that high-quality Nyquist pulses can be

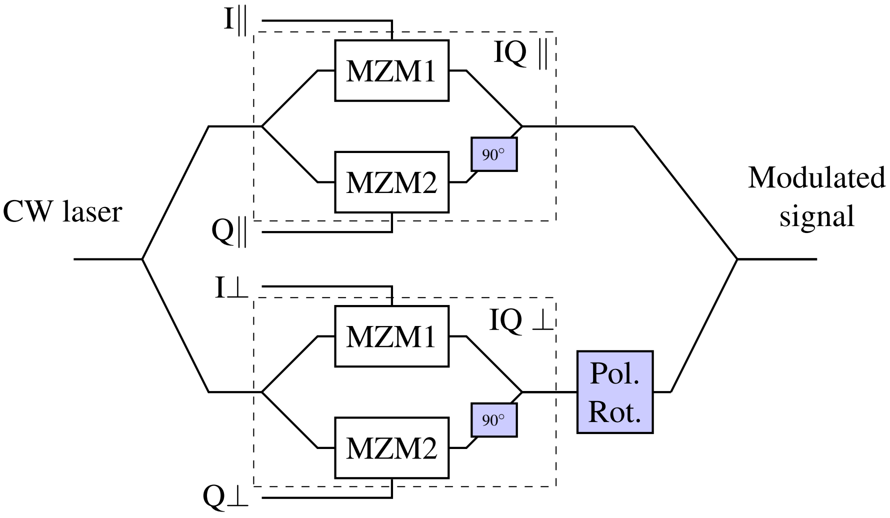


Figure 2.8: Schematic diagram of a DP IQ modulator.

generated even down to a zero roll-off value. The electrical approach is more flexible, as it typically relies on a digital-domain design of the pulse [28]. However, the main drawback is given by the increasing resolution required as the pulse roll-off decreases (increasing peak-to-average ratio), which is usually a demanding requirement for digital-to-analogue converters.

2.4.2 Coherent detection

A major step ahead in the endeavour for spectrally efficient optical fibre transmission was marked by the re-discovery of coherent optical detection and digital signal processing techniques [29–31]. Coherent detection allows access to the entire information of the transmitted optical field, i.e. its amplitude and phase components, as opposed to IMDD systems, in which only the field intensity can be detected. As a result, coherent detection enables the use of high-SE modulation formats, where the information is encoded jointly on the amplitude and phase of the optical field.

The photo-detection process, which is typically performed by a photo-diode, is inherently sensitive to the intensity of an incoming field. In order to detect the optical field the process must follow two steps, which are illustrated in Fig. 2.9: i) mixing of the incoming optical signal with a CW signal at the same nominal frequency; ii) photo-detection.

The mixing consists of coupling the optical field to be detected with an optical carrier (typically available at the receiver) with the same nominal frequency and phase of the transmitted one. Such an optical carrier is called local oscillator (LO). The summation of these two fields is then photo-detected by mean of a photo-diode. Assuming that the polarisation state of the incoming optical field and the one of the LO are aligned and the LO frequency and phase are ideally locked to the signal carrier, then the electrical signal at the output of the photodiode is proportional to

$$|E_{\mathrm{sig}}(t)+E_{\mathrm{LO}}|^2=(E_{\mathrm{sig}}(t)+E_{\mathrm{LO}})(E_{\mathrm{sig}}+E_{\mathrm{LO}})^*=|E_{\mathrm{sig}}(t)|^2+|E_{\mathrm{LO}}|^2+2\,\mathrm{Re}(E_{\mathrm{sig}}(t)E_{\mathrm{LO}}^*) \tag{2.59}$$

where $E_{\mathrm{sig}}(t)$ and E_{LO} are the complex amplitudes of the transmitted optical field and the LO, respectively. The time dependency is made explicit only for $E_{\mathrm{sig}}(t)$ to stress the fact that the incoming signal is modulated, whereas E_{LO} is a CW optical field. The right-most equality in Eq. (2.59) shows that by mixing and subsequently detecting the field intensity, a coherent term proportional to the I component of E_{sig} $(2\,\mathrm{Re}(E_{\mathrm{sig}}(t)E_{\mathrm{LO}}^*))$ arises. Likewise, by coupling E_{sig} with a 180 degrees phase-shifted version of E_{LO} (-E_{LO}), a term proportional to the Q component of E_{sig} is obtained. The direct-detection (d.d.) components, i.e. the terms that are proportional to the field intensity ($|E_{\mathrm{sig}}(t)|^2$

and $|E_{\mathrm{LO}}|^2$), can be removed using different approaches. The most straightforward one consists in filtering out the direct current component that arises from $|E_{\mathrm{LO}}|^2$, and increasing the ratio between the LO power and the signal power, such that the residual term $|E_{\mathrm{sig}}|^2$ is small compared to the coherent term.

The polarisation state of the incoming signal is in general unknown at the receiver. In order to reconstruct it, polarisation-diverse receivers are adopted. Fig. 2.9 shows the structure of such a coherent receiver. The incoming signal is first split into two orthogonal polarisations through a polarisation beam splitter (PBS). These orthogonal polarisations are then mixed with two phase-shifted versions of the LO provided by IQ couplers. The resulting four optical signals are then sent to four photo-diodes whose currents are proportional to

$$\begin{aligned} I_1 &\propto \mathrm{Re}(E_x) + \text{d.d. terms} \\ I_2 &\propto \mathrm{Im}(E_x) + \text{d.d. terms} \\ I_3 &\propto \mathrm{Re}(E_y) + \text{d.d. terms} \\ I_4 &\propto \mathrm{Im}(E_y) + \text{d.d. terms} \end{aligned} \tag{2.60}$$

where E_x and E_y are the signal polarisation components selected by the PBS.

An alternative coherent receiver structure is the balanced architecture, where both signal and LO direct-detection terms are coherently cancelled through the generation of identical copies of the same direct-detection component. Such an architecture is schematically shown in Fig. 2.10. This architecture features a pair of four photodiodes per polarisation. Each signal polarisation coming out of the PBS is mixed with the LO using a 90 hybrid which is 4×4 optical coupler producing the four linear combinations between signal and LO shown in Fig. 2.10. The photodiodes outputs are pairwise fed into a differential amplifier which amplifies the difference between the two signal, so

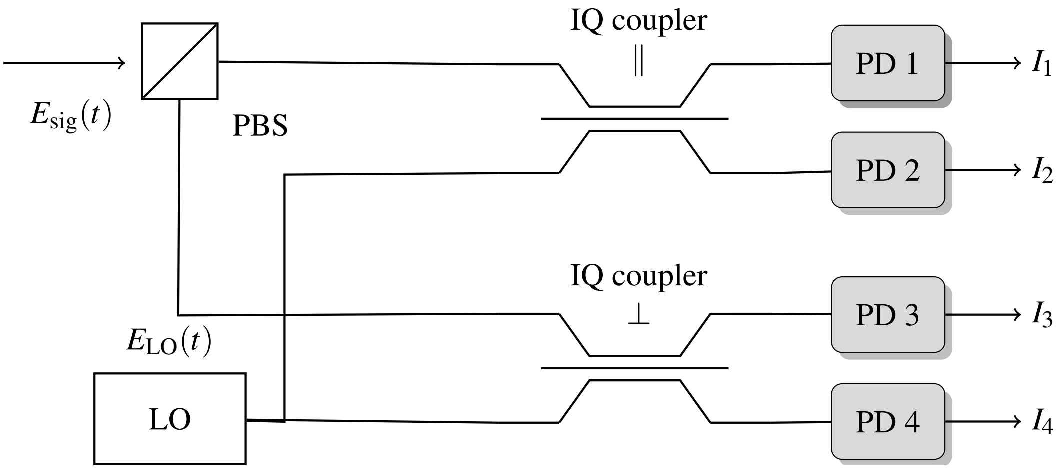


Figure 2.9: Structure of a single-ended polarisation-diverse coherent receiver.

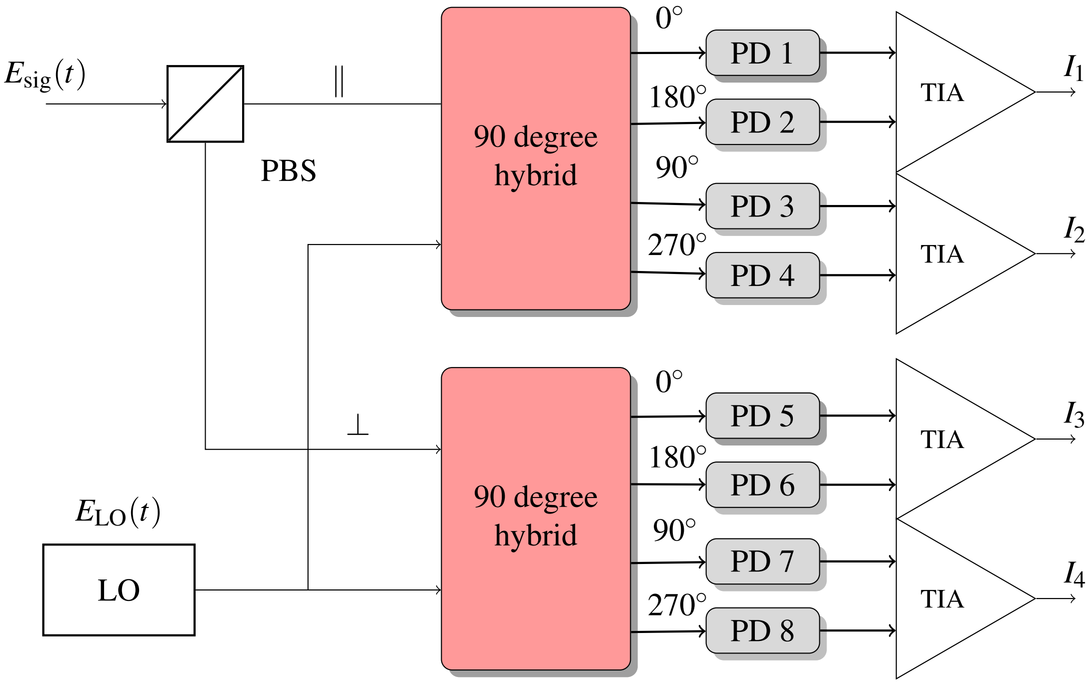


Figure 2.10: Balanced polarisation-diverse coherent receiver.

performing the coherent suppression of the direct-detection terms. Altough a better sensitivity (due to a better direct-detection term suppression) can be achieved using this structure [15], this comes at the cost of double the number of photodiodes required for a single-ended structure.

Separating the received field into four orthogonal components allows for a full reconstruction of the transmitted field which is performed in the digital domain by the DSP subsystems [30]. The typical DSP chain of a coherent receiver is displayed in Fig. 2.11. The four outputs of a coherent receiver are sampled and a first deskew and orthonormalisation is performed to compensate for a non-ideal receiver front-end. The compensation of static linear fibre impairments, such as CD, can then be performed using digital filters [29], thanks to the linear mapping between the optical field and the electrical signal at the output of the coherent receiver. Specifically, electronic CD compensation (EDC) digitally reverses the CD of the fibre performing the following operation in the digital domain

$$\exp\left(-\frac{j\beta_2\omega^2}{2}z\right). \tag{2.61}$$

As discussed in [29], the implementation of this *all-pass* filtering in the digital domain can vary from a time-domain finite impulse response (FIR) filter to a frequency domain FFT-based filter, depending on the amount of CD to be compensated and the frequency resolution adopted.

When a PM signal is transmitted, a DSP block is required to separate the two data

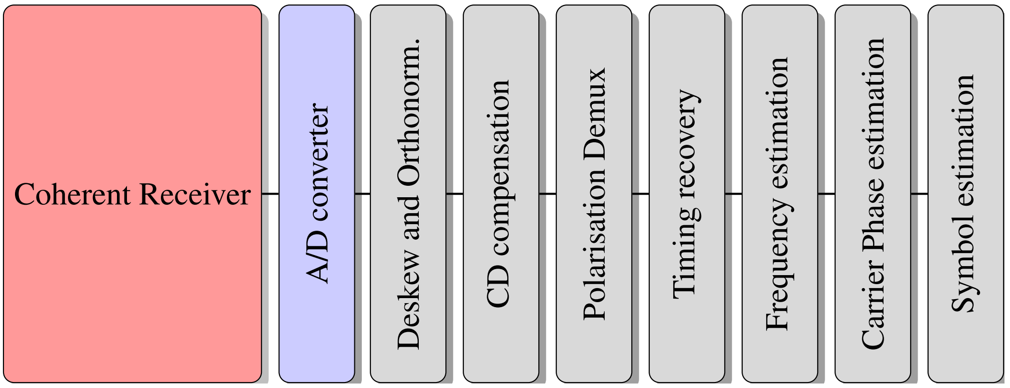


Figure 2.11: Typical DSP chain in a coherent receiver.

channels encoded on the two orthogonal polarisations. Indeed, the orthogonal SOPs provided by the coherent receiver are obtained through the PBS, and contain a generic linear superposition of the transmitted PM channels. Separating the two transmitted data channels requires finding a 2×2 matrix that inverts this superposition. This operation is referred to as *polarisation demultiplexing* and it is performed using blind equalisation algorithms such as the constant-modulus algorithm (CMA) [30]. The CMA exploits particular features of specific PM modulation formats, such as PM-QPSK, where the signal always features a constant modulus on each of the two polarisations (hence the name *constant modulus* algorithm). The algorithm then aims to minimise the error of the received signal to a constant modulus signal. In addition to separating the two polarisation channels, the CMA[8] also performs a compensation for other dynamic fibre linear impairments such as PMD and residual CD. This is achieved by adapting the taps of a set of real FIR filters for each polarisation [9].

Finally, frequency- and phase-syncronisation are required to compensate for a frequency shift between signal carrier and LO, and phase noise on both transmitted laser and LO laser. Feedforward techniques for frequency estimation are generally modulation-format dependent, and are based, for instance for the QPSK case, on taking the fourth power of the received symbols. This quantity can be used as an estimator for either maximum likelihood estimation techniques [33] or spectral techniques [34].

The phase noise of lasers typically utilised in optical communications varies between 1 KHz and 1 MHz [30]. This means that the phase noise process can be tracked over hundreds or tens of symbols, depending on the cardinality of the constellation and on its tolerance to phase fluctuations. Popular feedforward techniques for phase-noise tracking are also based on fourth power estimators which are averaged over a given time window. The most widespread of such techniques is referred to as Viterbi &

---

[8] The CMA algorithm is not the only adopted solution, and more sophisticated algorithms are required for modulation formats with higher cardinalities than PM-QPSK, such as PM-16QAM, PM-64QAM etc.

[9] Alternatively, separate FIR filters can be used for I and Q components on each of the two polarisation channels to adaptively compensate for residual skew between the I and Q channels [32].

Viterbi algorithm [35]. Main alternatives to feedforward techniques are represented by *decision-directed* phase estimation or digital phase-locked loops [34, 36].

This thesis exclusively analyses coherent systems. Although of paramount importance for real-world optical coherent receivers, the focus of thesis work is on transmission impairments yielded by fibre nonlinearity rather than on the implementation and on the performance of the DSP blocks discussed above. A suboptimal choice of the DSP parameters (e.g. window size of the Viterbi& Viterbi algorithm) can result in a penalty in the receiver performance. For all the results presented in the remainder of the thesis, phase noise and frequency shifts are neglected and full knowledge on the signal polarisation evolution is made available to the receiver. This approach arises from the assumption that polarisation demultiplexing, frequency- and phase-syncronisation are performed ideally at the receiver and no penalty is, thus, incurred.

2.5 Modelling of the optical fibre channel

It section 2.1.4 and 2.2 it was explained why Manakov or Manakov-PMD equations were chosen instead of NLSE or CNLSE for a comprehensive and effective description of fibre propagation in the presence of random polarisation phenomena. However, Manakov equation (see Eq. (2.34)) has no general analytical solution. Because of this, numerical methods are often used to derive arbitrarily accurate solutions, as discussed in section 2.6. However, due to the computational complexity of these methods, in many applications it is often important to derive simple analytical relationships that allow for a quick estimation of the performance of an optical fibre communication system. In particular, it is of great interest researching for approximated closed-form expressions which describe the impact of fibre nonlinearity on the transmission performance. Furthermore, looking for analytical relationships between the transmitted signal and the received one is of paramount importance for solving the question on the ultimate limits of optical fibre transmission, i.e. the optical fibre channel *capacity* [13, 37, 38].

The problem of finding (relatively simple) explicit analytical relationships for the performance metrics of interest, satisfying Eq. (2.34) with acceptable accuracy, is usually referred to as *channel modelling*. In recent years this has been the focus of a large part of research in optical fibre communications and it still represents a central topic in this area. In the following, some of the recent advances on this topic are discussed, with particular focus on the most widely used analytical model, i.e. the Gaussian noise (GN) model [39]. Because of its mathematical simplicity, this model is used in the rest of this thesis to gain a deeper insight on the results obtained through numerical simulations.

2.5.1 The GN-model of fibre propagation

The GN-model was first introduced by Carena et al. in [40], and further discussed in [39, 41, 42]. The GN-model is arguably the most popular and widespread among the available models of nonlinear fibre propagation, mainly due to the simplicity of the expressions resulting from the model.

According to GN-model distortions arising from fibre nonlinearity can be accounted for as an additive source of (Gaussian) noise which is independent from both the signal and the ASE noise. As a result this noise source adds in power to the conventional ASE noise accumulating along an optical fibre link, and it is usually referred to as nonlinear interference (NLI). The additive characteristic of the NLI power is justified by the usual adoption of standard detection techniques that are tailored to the AWGN channel, such as the matched filter/sampler receiver followed by symbol-by-symbol decision. Indeed, using such a conventional approach, the NLI can only be accounted for as an additive and independent noise source[10].

The main result of the GN-model is an analytical, and in some cases closed-form expressions, relating the main system parameters and the NLI power. This power is then typically used to estimate an *effective* received SNR, which under the assumption of additivity of the NLI, is given by

$$\mathrm{SNR} = \frac{P}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}} \tag{2.62}$$

where P is the signal power (one channel), P_{ASE} is the total ASE noise power accumulated along the fibre link.

The GN-model is based on two main hypothesis: i) nonlinearity impact on signal propagation is low (perturbative hypothesis); ii) the transmitted signal is replaced by a Gaussian process with the same average power. Under the first assumption, the NLI can be found performing the following steps: i) solving (2.34) without the nonlinear term, which yields the so-called zero-th order solution; ii) replacing the zero-th order solution to the nonlinear term; iii) solving again (2.34) to finally find the first-order solution. The first-order solution to (2.34) is a first approximation to the NLI which is the more accurate the lower is the amount of nonlinearity. This approach to solve Eq. (2.34) is shared among the almost entirety of the available channel models and it is referred to as first-order perturbation method [43, Ch. 1].

Once a first-order solution is obtained, the second assumption simplifies the problem of converting the analytical expression for the NLI field into an NLI power. Indeed, by assuming that the propagating field is a Gaussian process with a PSD $S(f)$, the PSD of the first-order NLI can be written as [40]

---

[10] Any noise process can be considered additive. However, when signal and noise are to some extent correlated, it is more suitable to see the noise as multiplicative.

$$G_{\mathrm{NLI}}(f) = \frac{16}{27}\gamma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(f_1)S(f_2)S(f_1+f_2-f)\theta(\Delta f_1,\Delta f_2)\,df_2\,df_1 \tag{2.63}$$

where $S(f)$ is the PSD of the transmitted signal, and

$$\theta(\Delta f_1,\Delta f_2) \triangleq \left|\frac{1-e^{-\alpha L_s}e^{j4\pi^2|\beta_2|L_s\Delta f_1\Delta f_2}}{\alpha-4\pi^2|\beta_2|\Delta f_1\Delta f_2}\right|^2 \frac{\sin^2(2N_s\pi^2\Delta f_1\Delta f_2|\beta_2|L_s)}{\sin^2(2\pi^2\Delta f_1\Delta f_2|\beta_2|L_s)} \tag{2.64}$$

is the FWM efficiency function already shown in Eq. (2.29) for a discrete spectrum. In Eq. (2.64), N_s is the number of fibre spans in the link, L_s is the span length, whereas Δf_1 and Δf_2 are defined as (f_1-f) and (f_2-f) respectively.

Eq. (2.63) is derived through three main steps: i) a discrete spectrum is first assumed and the discrete FWM products are calculated based as in [2]; ii) full uncorrelation between all frequency components of the transmitted spectrum (see Appendix [40, Appendix D]) is assumed, which allows for the power additivity of the FWM products; iii) the frequency spacing between the discrete components of the spectrum tends to zero. The assumption in step ii) is verified if the actual transmitted signal is replaced by a Gaussian process. However, this is not in general the case and the inaccuracy caused by this assumption is further discussed in the next section.

A general closed-form expression for Eq. (2.63) does not exist, as this is dependent on the specific spectral shape of $S(f)$. However, in all cases the NLI power can be written as [39]

$$P_{\mathrm{NLI}} = \eta P^3 \tag{2.65}$$

where η is a function of several system parameters, albeit independent of the transmitted power. For particular scenarios, approximated closed-form expressions of η have been proposed [39, 40, 44].

In particular, when the transmitted channel spectrum is exactly rectangular and the system under consideration is EDFA-amplified and dispersion-unmanaged an approximated closed-form expression for the η factor over one span of fibre is given by [39]

$$\eta_1 \approx \frac{8}{27}\frac{\gamma^2 L_{\mathrm{eff}}^2}{\pi|\beta_2|R_s^2 L_{\mathrm{eff}}} \operatorname{arcsinh}\left(\frac{\pi^2}{2}L_{\mathrm{eff}}R_s^2 N_{\mathrm{ch}}^{2\frac{R_s}{\Delta f_{\mathrm{ch}}}}\right) \tag{2.66}$$

where

$$L_{\mathrm{eff}} = \int_0^L \exp(-\alpha z)\,dz \tag{2.67}$$

is the effective length of a fibre span of length L, R_s is the symbol rate, N_{ch} is the number of transmitted WDM channels, and Δf_{ch} is the channel spacing. The dependence of η on the overall optical bandwidth $B = R_s N_{\mathrm{ch}}$ (under ideal Nyquist spacing conditions) is nonlinear through the arcsinh function, which can be approximated for large values of the argument to a logarithmic function. As a result, Eq. (2.66) predicts a monotonically increasing NLI power as the transmitted bandwidth is increased, with no saturation as B tend to infinity (although the slope of the increase does tend to zero). Furthermore, the dependency on the symbol rate of the NLI PSD is just apparent. In fact, multiplying η by P^3 to obtain the NLI power, a linear dependency R_s can be found, and therefore the NLI PSD (in this case equal to P_{NLI}/R_s) is independent on R_s.

Over multiple spans, the NLI accumulation is in general *noncoherent*, i.e. the NLI power does not scale linearly. An approximated scaling with the distance is given by [39]

$$\eta = \eta_1 N_s^{1+\varepsilon} \tag{2.68}$$

where

$$\varepsilon \approx \frac{3}{10} \ln \left(1 + \frac{6}{L_s} \frac{L_{\mathrm{eff}}}{\operatorname{arcsinh}(\frac{\pi^2}{2}|\beta_2| L_{\mathrm{eff}} B)} \right) \tag{2.69}$$

is referred to as coherence factor and varies between 0 and 1, albeit being typically a small number (≤ 0.2). Substituting Eq. (2.68) into Eq. (2.65) and then into Eq. (2.62) we find

$$\mathrm{SNR} = \frac{P}{N_s P_{\mathrm{ASE}} + \eta_1 N_s^{1+\varepsilon} P^3}. \tag{2.70}$$

Although, as discussed later on, Eq. (2.70) is just an approximation, it can be used to explain the typical performance behaviour of optical fibre transmission systems, where an optimum operating point is always found as a function of the transmitted power P. The analytical prediction of the SNR of the central channel given by Eq. (2.70) is illustrated in Fig. 2.12 as a function of the power per channel and for two different transmitted bandwidths and transmission distances.

It can be seen that all curves show the same qualitative behaviour. Indeed, for transmitted powers up to approximately -4 dBm, the SNR increases as 1 dB/dB, as expected for the linear case ($P_{\mathrm{NLI}} = 0$). This is due to the dominance of linear noise term (P_{ASE}) compared to the NLI term. As the NLI power becomes comparable to the total ASE noise the curves begin to saturate until a peak is achieved. The effect of the NLI becomes dominant beyond this point where, due to its scaling as P^3, the SNR decays as P^{-2} (see Eq. (2.70) when $N_s P_{\mathrm{ASE}}$ is considered negligible). At all transmission distances, the increase of the transmission bandwidth results in a decrease of the optimum SNR by about 1 dB and a shift of the optimum transmitted power by about -0.7 dB. The dependence on transmission distance is instead stronger, as the optimum SNR can be observed to decrease approximately linearly with N_s (e.g. $\approx$6 dB reduction going from a distance of 800 km to 3200 km). A detailed analysis of the scaling of the SNR based on the GN-model predictions are presented in the next

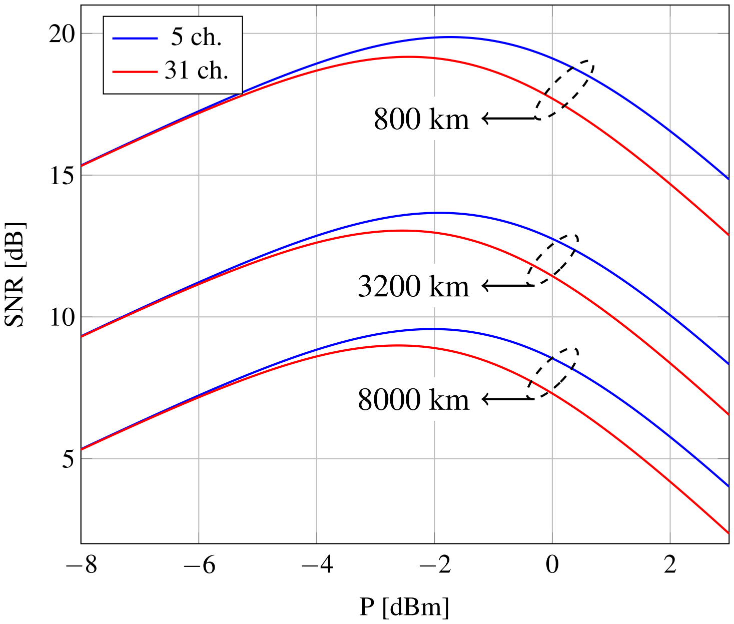


Figure 2.12: SNR of the central channel as a function of the transmitted power P for different transmitted optical bandwidths, and for different transmission distances. The SNR is obtained using Eq. (2.70).

chapter.

2.5.2 Recent developments on optical fibre channel modelling

Since its introduction in 2012 the GN-model has received a lot of attention from the research community. This research focus resulted in uncovering the main flaws of the model [45, 46] and in the introduction of new and more accurate models [47–50], which, at present, represent the state-of-art of analytical modelling of optical fibre transmission.

The main difference between such improved models and the GN-model consists in the dropping of the Gaussian assumption for the propagating signal. This assumption was justified on the observation that for typical dispersion-unmanaged systems the output constellation shows a distribution that closely approaches a circularly symmetric Gaussian distribution at all powers of interest for optical fibre communications [40]. This effect was observed both in experimental scenarios and simulated systems, even in the absence of ASE noise [40, 51], and hence the assumption of the NLI behaving as a Gaussian process. Furthermore, in a few works it was argued that, although the transmitted signal might have properties very dissimilar from a Gaussian process, the effect of dispersion would effectively transform the signal into a Gaussian process after a certain distance. Particularly, in [45], the *transient* during which the signal is still very dissimilar from a Gaussian process (first spans of a transmission link) was recognised to result in an underestimation of the NLI power. However, in the limit of infinite dispersion, it was argued that this signal would become a Gaussian process due to the central limit theorem, and as a result the error in the GN-model estimation would vanish.

In both [46, 50], it was observed this is not the case and, under a Gaussian assumption, a constant error in the NLI power estimation is accumulated over each fibre span after the initial ones. As a result, it was clear that this assumption, although bringing significant simplifications in the analytical expressions, would result in a fixed inaccuracy. Such inaccuracy was recognised to be strongly dependent on the properties of the transmitted signal and specifically on the modulation format adopted in transmission.

The first work recognising the effect of modulation format on the NLI was [47], where a first-order perturbation theory in the time domain to solve Eq. (2.34) was presented. The derivation of the first-order solution did not rely on any specific assumption about the signal and lead to analytical expressions for NLI in both its intra-channel and inter-channel components. One of the main shortcomings of this model (apart from the role of the modulation format on the generation of the NLI) was the fact that a good portion of the NLI could be identified as pure phase noise, in contrast with [40] where both ASE noise and NLI were assumed circularly Gaussian. However, probably due to the complexity of the analytical expressions, the significance of the model in [47] were never entirely understood.

The work in [47] was subsequently expanded in [50]. In this work, the expressions in [47] were further developed and an equivalent frequency-domain approach was introduced. Using the latter approach, the shortcomings of the GN-model could be better identified. Also, numerical results, in agreement with the new model, showed a gap between GN-model predictions and NLI generated by the modulated signals, such as PM-QPSK and PM-16QAM, can be found.

Independently from the previous two works, results in [48, 49] showed qualitatively similar results based on a variation of the regular perturbation approach used in both [47, 50] called logarithmic perturbation (LP) method. The derivation was performed in the frequency domain (hence the name *frequency-resolved* LP) an assuming XPM as the main component of the NLI (no SPM or FWM included in the description). Through this approach, this work showed that, to the first-order, optical fibre can be seen as a linear time-varying system, hence allowing a simplified analysis both in terms of NLI power [48] and in terms of achievable rates [49].

In response to these two works, the GN-model was refined to account for the specific properties of the transmitted signal [52, 53]. This new model, called enhanced GN (EGN)-model, introduced, in addition to the GN NLI power estimation, a correction term which was found to be modulation format-dependent. Furthermore, this new term includes contributions to the NLI which were discarded in the models in [47, 49, 50].

All the models discussed in this section do not in general provide closed-form approximations due to the higher complexity of the analytical relationships compared to GN model. However, in [54], an approximated closed-form formula was presented for the EGN-model. Similarly for the closed form expression in Eq. (2.66) the EGN formula in [54] is tailored for the ideal Nyquist shaping case (roll-off=0), with arbitrary spacing between the channels, in a dispersion-unmanaged, EDFA-amplified, fibre channel. According to this formula, the corrected NLI factor for a perfectly rectangular channel spectrum can be written as

$$\eta_{\text{EGN}} = \eta - \eta_{\text{corr}} \tag{2.71}$$

where η is the NLI coefficient according to the GN model and η_{corr} is the correction factor given by

$$\eta_{\text{corr}} \approx \frac{80}{81}\Phi\frac{\gamma^2 L_{\text{eff}}^2 N_s}{R_s \Delta f \pi \beta_2 L_s}\text{HN}([N_{\text{ch}}-1]/2). \tag{2.72}$$

Φ is a coefficient which is closely related to the modulation format fourth and sixth standardised moment (see [52]) and it can be found to be equal, for instance, to 1 for PM-QPSK, 17/25 for PM-16QAM and 13/21 for PM-64QAM. HN is the harmonic number function defined as

$$\text{HN}(N) = \sum_{n=1}^{N} \frac{1}{n}. \tag{2.73}$$

Based on this correction factor, each modulation format generates a different NLI power and therefore a different SNR. The impact of the correction factor introduced by the EGN-model on the SNR for different modulation formats is assessed in the following. The calculation of the SNR is performed for a transmission of 31x32 GBaud channels and a channel spacing of 33 GHz over a distance of 3200 km. The results are shown in Fig. 2.13.

As discussed earlier, the GN-model underestimates the SNR by an amount that varies based on the actual modulation format adopted. The largest underestimation is found for PM-QPSK, as in this case the GN-model underestimates the optimum SNR by 0.53 dB. The GN-model overestimation of η can be deduced by measuring the SNR gap for asymptotically large transmitted powers. It can be deduced that PM-QPSK η factor is overestimated by 1.47 dB in the nonlinear region. As the modulation order increases, the gap with the GN-model decreases to 0.3 dB and 0.85 dB compared to the optimum SNR and η of PM-64QAM, respectively. However, this gap does not close completely, as can be deduced by comparing the SNR performance of PM-16QAM and PM-64QAM, which show only a negligible difference. This suggests that a modulated signal preserves its non-Gaussian features, even for an arbitrary large constellation cardinality. The analytical results described in the last two sections are used in chapter 3 to validate the numerical results on the SNR performance of systems using EDC or NLI compensation at the receiver.

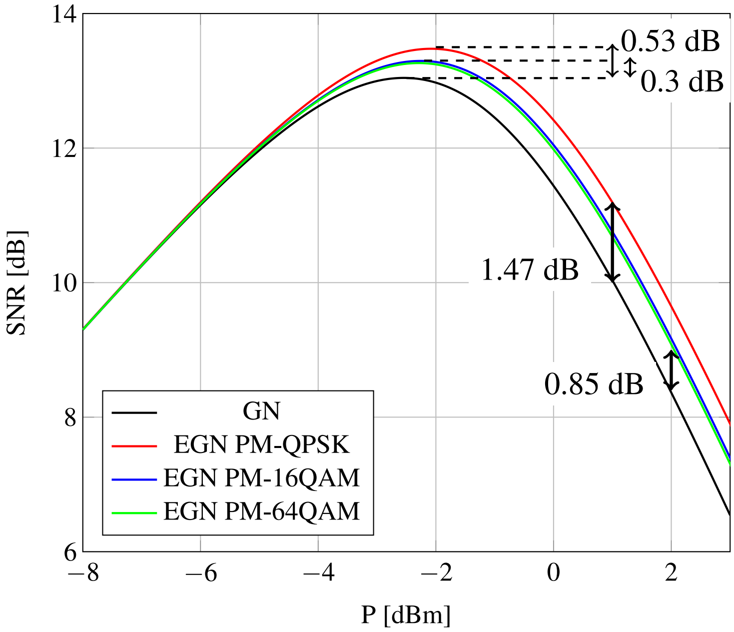


Figure 2.13: SNR of the central channel as a function of the transmitted power P for a 31x32 GBaud channels transmission over 3200 km and different modulation formats. The SNR is obtained using Eq. (2.70).

2.5.3 Signal-ASE beating

In the above sections different models of nonlinear fibre propagation were discussed in order to predict the NLI power and, consequently, the receiver SNR through Eq. (2.62).

One of the underlying assumptions of all the models presented is that the signal-signal NLI is generated only from interactions within the transmitted signal and itself. However, in EDFA-amplified systems (Fig. 2.6) the ASE noise is added at every span in a lumped fashion. This leads not only to an additive Gaussian noise accumulating linearly with distance, but also to a continuously generated nonlinear signal-ASE interaction. This phenomenon, also referred to as parametric noise amplification [12], is generally neglected as signal-signal terms usually dominate. However in some particular transmission scenarios (see chapter 3) this is no longer the case, and an accurate estimation of signal-ASE term is required.

In order to do that, we first need to understand how the signal-ASE NLI accumulates over a multi-span system. Fig. 2.14 schematically describes such an accumulation. Let us assume our system has N_s spans and N_s amplifiers distributed as shown in Fig. 2.14. The total accumulated signal-ASE term can be considered as the summation of multiple signal-ASE contributions originating by the interaction with the noise lots added by each EDFA in the link[11]. If we denote the signal-ASE NLI generated by the n-th EDFA as $P_{\text{SN}}(n) = \Gamma(N_s - n + 1)$, i.e. as a function of the number of spans over which the NLI accumulates, then the total signal-ASE NLI can be rewritten as

---

[11] Clearly, the noise lots are uncorrelated with each other and so can their relative signal-ASE NLIs.

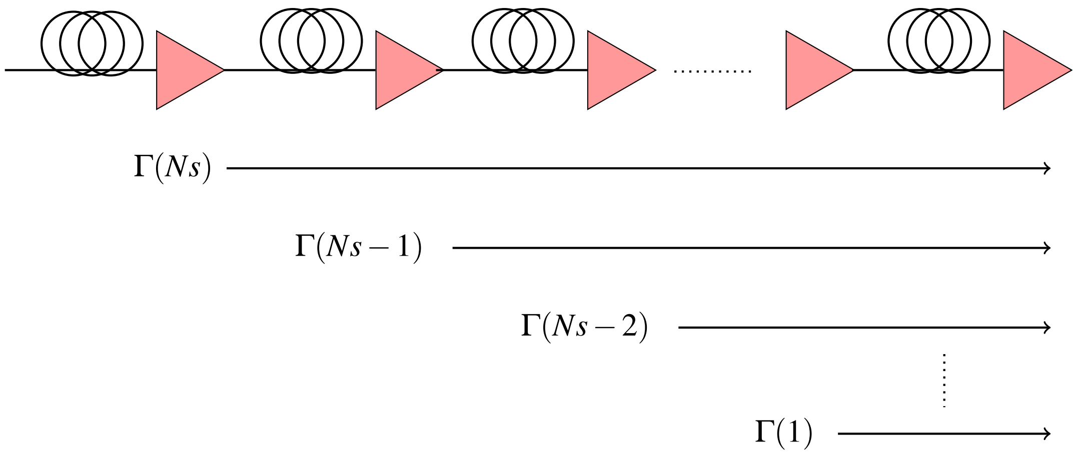


Figure 2.14: Schematic diagram of the accumulation of the signal-ASE NLI in a multi-span EDFA-amplified system.

$$P_{\mathrm{SNtot}} = \sum_{n=1}^{N_s} \Gamma(N_s - n + 1). \tag{2.74}$$

In order to find $\Gamma(N_s - n + 1)$, we can resort to the approach followed by the GN-model (or EGN) in [40]. In this case though, the signal PSD must be replaced with the combination of signal PSD and ASE noise PSD. A detailed mathematical derivation is presented in the Appendix, after which we find, if we use for simplicity the GN-model NLI factor,

$$P_{\mathrm{SNtot}} \approx 3\eta\zeta P_{\mathrm{ASE}} P^2 \tag{2.75}$$

where

$$\zeta \triangleq \sum_{n=1}^{N_s} n^{1+\varepsilon} \approx \frac{N_s^{2+\varepsilon}}{2+\varepsilon} + \frac{N_s^{1+\varepsilon}}{2} \tag{2.76}$$

and the approximation follows from the Faulhaber's formula [55, eq. (0.121))].

Eq. (2.75) shows two main properties of the signal-ASE NLI: it scales quadratically with transmitted power (as opposed to the signal-signal NLI which scales cubically with transmitted power); it accumulates super-quadratically ($N_s^{2+\varepsilon}$) with the distance (as opposed to the signal-signal NLI which accumulates only super-linearly).

Although often neglected in the design of present-day optical fibre communication systems, the role of the signal-ASE NLI is crucial as it represents one of the main performance limitations (probably the ultimate) on the performance of optical fibre transmission [38, 56]. In particular, the signal-ASE NLI defines the limits of receivers which are able to fully compensate for signal-signal nonlinear effects. The performance of such receivers will be analysed in chapter 3.

2.6 Numerical methods for fibre propagation

Throughout this chapter the differential equations governing the propagation of light-wave pulses in optical fibre were presented and discussed. It was also pointed out that such equations do not have in general an explicit analytical solution. We have also presented analytical models that attempt to provide an approximated solution under certain assumptions and specific regimes. Numerical methods remain the only option when the solution of Eq. (2.34) is required with arbitrary high accuracy and in all regimes where models do not hold. For instance, one of such regimes happen when high transmitted powers are used and, as a result, the first-order perturbation solution becomes increasingly inaccurate.

In this section, the main techniques for numerical evaluation of the performance of optical fibre communication systems are presented. Such numerical methods are heavily used throughout this thesis. In particular they are used in chapters 3, 4 and 5 to emulate fibre propagation, and numerically evaluate the system performance.

2.6.1 Split-step Fourier method

The most widespread method to numerically solve Eq. (2.34) is the SSFM algorithm [57], [1, Ch. 2]. Although other numerical methods are available such as time-domain finite difference methods [1, Ch. 2], [58], the SSFM has achieved popularity thanks to its fast implementation through the fast Fourier transform (FFT), which makes it the lowest complexity algorithm available to solve NLSE/Manakov equation [58].

The solution of (2.34) can be implicitly expressed using the exponential operator as [1, Ch. 2]

$$\mathbf{A}(z,t) = \mathbf{A}(0,t)\exp(\hat{D}z + \hat{N}z) \tag{2.77}$$

where $\mathbf{A}(0,t)$ is the initial condition imposed at section $z = 0$ and $\hat{D}$, $\hat{N}$ are the dispersion and nonlinear operator defined as

$$\hat{D} = j\frac{\beta_2}{2}z\frac{\partial}{\partial^2 t} \tag{2.78}$$

$$\hat{N} = -j\frac{8}{9}\gamma||\mathbf{A}(t,z)||^2 - \frac{\alpha}{2}. \tag{2.79}$$

The SSFM is based on breaking the integral solution into sections small enough such that each of the vector components of Eq. (2.77) can be approximated by

$$A(t,z+h) = \exp\left[(\hat{D}+\hat{N})h\right]A(t,z) \approx \exp(\hat{D}h)\exp(\hat{N}h)A(t,z) \tag{2.80}$$

i.e. the effect of the sum of the two operators can be expressed as the composition of each of the two acting separately [1, 57, 59]. This can be explained through the Baker-Campbell-Hausdorff formula on non-commutative operators [60], based on which we can expand the operator at the right-hand side of Eq. (2.80) as

$$\exp(\hat{D}h)\exp(\hat{N}h) = \exp(\hat{D}h + \hat{N}h + \frac{h^2}{2}([\hat{D},\hat{N}]) + \frac{h^3}{12}([\hat{D},[\hat{D},\hat{N}]) + \ldots \tag{2.81}$$

where

$$[\hat{D},\hat{N}] = (\hat{D}\hat{N} - \hat{N}\hat{D}) \tag{2.82}$$

is the *commutator* applied to the operators $\hat{D}$ and $\hat{N}$. We can observe that in the limit of the step-size h tending to zero, Eq. (2.80) converges to the exact solution at left-hand side of Eq. (2.80) with a local error (e.g. the error experienced in a single step) dominated by the term $\exp\left[\frac{h^2}{2}(\hat{D}\hat{N} - \hat{N}\hat{D})\right]$ for small values of h.

A way to reduce this local error is to resort to the alternative approximation

$$A(t,z+h) = \exp\left[(\hat{D}+\hat{N})h\right]A(t,z) \approx \exp\left(\hat{D}\frac{h}{2}\right)\exp\left(\hat{N}h\right)\exp\left(\hat{D}\frac{h}{2}\right)A(t,z) \tag{2.83}$$

i.e. further splitting the SSFM step in three sections: a first section applies the linear operator $\hat{D}$ over half the step-size $h/2$; a second section applies the nonlinear operator $\hat{N}$ over the entire step size h; finally a last section applies again the linear operator over the residual length $h/2$. In this case the $\hat{D}$ operator is symmetrically applied before and after the $\hat{N}$ operator, hence the name *symmetric* SSFM for this approach. Applying again the Baker-Campbell-Hausdorff formula for the two operators $\exp\left(\hat{D}\frac{h}{2}\right)$ and $\exp\left(\hat{N}h\right)\exp\left(\hat{D}\frac{h}{2}\right)$, we find

$$\exp\left(\hat{D}\frac{h}{2}\right)\exp(\hat{N}h)\exp\left(\hat{D}\frac{h}{2}\right) = \exp(h\hat{D} + h\hat{N} + \frac{h^3}{6}\left(\left[\frac{\hat{N}+\hat{D}}{2},\left[\hat{N},\frac{\hat{D}}{2}\right]\right]\right) + \ldots \tag{2.84}$$

In this case the dominant local error term is represented by $\frac{h^3}{6}([\frac{\hat{N}+\hat{D}}{2},[\hat{N},\frac{\hat{D}}{2}]])$ which tends to zero more rapidly than the local error in the bulk-step case.

Each of the exponential operators presented above represents the solution of Eq. (2.34) when $\hat{D}$ and $\hat{N}$ are in turn equal to zero. These solutions were discussed in section 2.1 and can be expressed as

$$\exp\left(\hat{D}(\omega)h\right) = \exp\left(j\frac{\beta_2}{2}\omega^2 h\right) \tag{2.85}$$

$$\exp(\hat{N}(z,t)h) = \exp\left(-j\frac{8}{9}\gamma\int_z^{z+h} ||\mathbf{A}(t,z')||^2 \exp(-\alpha z')dz' - \frac{\alpha}{2}h\right) \tag{2.86}$$

where it is explicitly indicated when the operator is performed in the time or in the frequency domain ω. Switching to the Fourier domain for the dispersion operator (hence the name split-step Fourier) allows to reduce the operator to a point-wise multiplication (in the frequency variable ω). Moreover, the Fourier transform can be efficiently implemented numerically using a fast Fourier transform (FFT). The operator $\exp(\hat{N}(t)h)$ can be also reduced to a point-wise multiplication in the time domain as the integral within the exponential Eq. (2.86) can be calculated as

$$\int_{z}^{z+h} \exp(-\alpha z')dz' = \frac{1-\exp(-\alpha h)}{\alpha} \triangleq h_{\text{eff}}. \tag{2.87}$$

As a result Eq. (2.86) can be written in a full closed-form expression as

$$\exp(\hat{N}(z,t)h) = \exp\left(-j\frac{8}{9}\gamma||\mathbf{A}(t,z)||^2 h_{\text{eff}} - \frac{\alpha}{2}h\right). \tag{2.88}$$

The numerical implementation of the SSFM can be summarised in Fig. 2.15 for both the bulk-step case (2.15a) and the symmetric split-step case (2.15b).

The sequence of samples of the optical field in the time domain $\mathbf{A}_t(z,n)$ is first transformed into the frequency domain using the FFT obtaining the sequence $\mathbf{A}_\omega(z,k)$. The spectrum is then point-wise multiplied by Eq. (2.85) and transformed back in the time domain. Finally, the resulting time-domain sequence is point-wise multiplied by Eq. (2.86) obtaining the sequence $\mathbf{A}_t(z+h,n)$.

In order to preserve an acceptable numerical accuracy, the solution of Eq. (2.34) for a fibre section of length L can be obtained by breaking the total length into small enough steps of length h and iteratively repeating the SSFM over each step. It is important to note that for multiple iterations of the the SSFM there is no difference between the bulk-step and the symmetric step implementation of the SSFM. This is due to the fact that consecutive dispersion blocks, resulting from cascading the blocks in Fig. 2.15b, can be merged in a single frequency domain block, effectively reobtaining the schematic in Fig. 2.15a. Thus, the global error, i.e. the error accumulated over multiple SSFM iterations, is the same for the bulk step and symmetric step case and it can be shown to decrease as $O(h^2)$ [59].

SSFM and one of its variants are used for the numerical simulations performed in

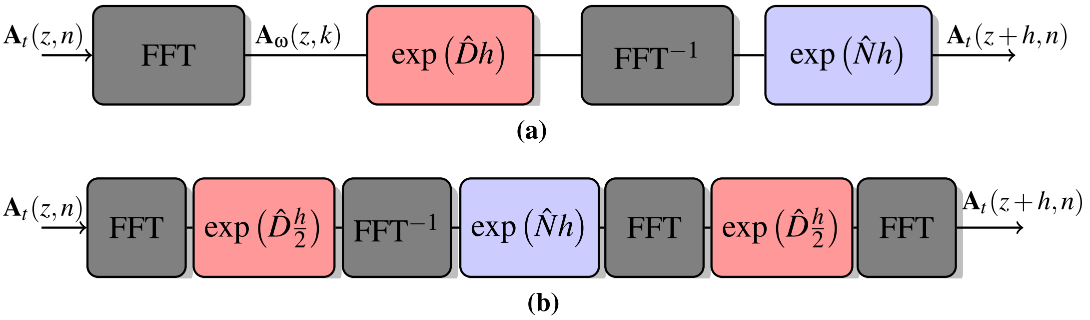


Figure 2.15: Schematic diagram of a SSFM section of length h for the (a) bulk step and (b) symmetric step implementation.

chapter 3, 4 and 5. Such a variant, which is more computationally effective, is discussed in the following section.

2.6.2 Adaptive SSFM methods

The SSFM method presented above does not allow to explicitly control the error arising due to the spatial discretisation. Furthermore, using uniform step sizes does not always represent the most efficient way to maximise the accuracy for a fixed number of computations. Indeed, for EDFA-amplified systems, most of nonlinear propagation effects takes place within a distance that can be quantified in the span effective length (see Eq. (2.67)). As a result, for a fixed error in the numerical solution, smaller steps are needed at the beginning of the span, while coarser steps could be used when the signal propagates beyond the effective length, because of a small (if not negligible) impact of nonlinearity. An effective variant of the SSFM should then be adaptive to the amount of nonlinearity generated (i.e. on transmitted power, bandwidth and fibre parameters) and at the same time should distribute the step sizes in order to match the attenuation profile of the fibre.

The first adaptive method was proposed in [61], where the step size Δz was determined by a maximum allowed nonlinear phase rotation Φ_{NLmax} as

$$\gamma \Delta z P_p < \Phi_{NLmax} \tag{2.89}$$

where P_p is the peak power of the signal at section z. Although such criterion qualitatively controls the accuracy of the calculation of the nonlinear step it does not explicitly control the amount of numerical artifact.

In [62] an alternative method was proposed to quantitatively control the accuracy of the SSFM. This method was based on controlling the power of the FWM spurious tones arising as artifacts of a coarse numerical integration of Eq. (2.34). By a calculation of the fictitious FWM products due to the SSFM spatial discretisation it was found that this term can be kept arbitrary small by choosing a step size distribution given by

$$\Delta z_n = -\frac{1}{2\alpha} \ln \left[\frac{1 - n\delta}{1 - (n-1)\delta} \right] \qquad n = 1, \ldots, K \tag{2.90}$$

where $\delta = (1 - e^{-2\alpha L})/K$ and K is the number of steps given by the criterion

$$K > \frac{3}{4} N_c^2 \gamma^2 L_e^2 P^2 10^{(x/10)} \tag{2.91}$$

where x represents the desired ratio (in dB) between the signal power and the power of the spurious FWM tone. The logarithmic step distribution, which also arises from the method in [61], was found to minimise the number of required steps to achieve signal-to-spurious tone ratio of at least x. The method in [62] was shown to decrease the number of required steps by more than an order of magnitude with respect to a uniform step size with comparable accuracy.

2.6.3 Monte-Carlo simulation of PMD

In this section we discuss the numerical methods utilised in chapter 3 to characterise the impact of PMD on the system performance. Numerical emulation of PMD is based on the discretisation of the Manakov-PMD equation presented in section 2.2. In order to do that, the different PMD sections of the fibre are emulated using a wave-plate approach [10]. Such an approach consists in cascading different sections (PMD sections) where the PMD effect can be assumed small enough over the entire transmitted signal bandwidth. In this case the PSP approximation (see section 2.2) can be used to accurately describe the effect of PMD over that section. To emulate the evolution of the PSPs over the fibre length, the PSPs of the subsequent PMD section are randomly rotated to obtain a full uncorrelation with respect to the previous ones. Slightly different methods can be followed in doing this [7, 10]. In our approach, the PSPs are scrambled over the Poincaré sphere using the approach in [63]. Such an approach guarantees that the PSPs of two adjacent PMD sections are independent and uniformly distributed over the Poincaré sphere.

The emulation of the interaction between PMD and fibre nonlinearity traditionally relies on the so-called *coarse step* method [7]. This method spatially discretise the Manakov-PMD equation over steps that are comparable to the fibre nonlinear length. The quick evolution of the SOPs along this length is instead accounted for by the Manakov equation. However, the coarse step method implies that both SSFM and wave-plate step are computed over the same section. Since the wave-plate sections need to be uniformly sized for the correct statistical accumulation of the DGDs, a SSFM fibre simulation including PMD is also required to keep a uniform step size. This can be prohibitively complex for SSFM simulations of wide-band systems, which would require a significantly small step along the whole propagation path. On the contrary, wide-band simulations significantly benefits from the log-step approach discussed in section 2.6.2.

In order to reconcile both needs, the method used in this thesis merges PMD sections and SSFM steps in a single steps distribution. The SSFM is performed over logarithmically-spaced steps. If each step is performed within the same PMD section (wave-plate) no action is taken with respect to the PMD. At the boundary between PMD sections, random PSP rotation and delay is applied. This implies that the SSFM step *containing* the interface between two adjacent PMD sections is further split into two smaller sections. Following this approach we can then guarantee that: i) an efficient simulation of fibre nonlinear propagation is performed using the log-step method instead that the uniform one; ii) PMD and its interaction with fibre nonlinearity is correctly emulated by keeping the wave-plate length fixed.

As for the statistical characterisation of PMD and its effect on fibre transmission, a Monte Carlo simulation of the PMD sections is typically performed [7]. The Monte Carlo method consists in generating a random cascade of PMD sections whose PSP is uniformly scattered over the Poincaré sphere and DGDs of each section are drawn from a Gaussian distribution with standard deviation equal to 10% of the mean [64].

More sophisticated and efficient techniques are available to capture peculiar features of the PMD stochastic process, such as the *importance sampling* method [65, Ch. 7]. However, despite its high computational complexity, the Monte Carlo approach allows to capture more comprehensively PMD stochastic behaviour, especially in its interaction with fibre nonlinearity. For this reason Monte-Carlo PMD method was chosen to obtain the results shown in section 3.7.

2.7 Summary

In this chapter, a theoretical background for the study of coherent optical fibre communication systems was presented. The topics discussed are key for understanding the results presented in the following chapters of this thesis. First, the equations governing the propagation of lightwave pulses in optical fibre transmission, such as the Manakov equation, were analysed, enabling an insight on the main transmission impairments. Then, the structure and the properties of the two optical fibre transmission systems studied in this thesis were described. The fundamentals of high SE, coherent optical fibre transmission, which is the focus of the work in this thesis, were discussed. Analytical models for optical fibre propagation were investigated with particular focus on the GN-model. Finally, numerical methods for solving the Manakov equation used for the numerical results presented in this thesis, were described.

In the next chapter, a study on the performance of receiver-side digital nonlinearity compensation in high-SE optical fibre transmission is presented. Such an investigation makes use of most of the theory presented here, in particular including SSFM for the numerical results and the GN-model to develop analytical predictions on the SNR performance. Furthermore, PMD theory and Monte-Carlo numerical methods are used to evaluate the impact on the performance of nonlinearity compensation.

References

[1] G. Agrawal, *Nonlinear Fiber Optics*, 5th ed. Academic Press, 2013.

[2] K. O. Hill, D. C. Johnson, B. S. Kawasaki, and R. I. MacDonald, "CW three-wave mixing in single-mode optical fibers," *J. Appl. Phys.*, vol. 49, no. 10, pp. 5098–5106, 1978.

[3] C. R. Menyuk, "Pulse propagation in an elliptically birefringent Kerr medium," *IEEE J. Quant. Electron*, vol. 25, no. 12, pp. 2674–2682, 1989.

[4] P. K. Wai, C. R. Menyuk, and H. H. Chen, "Stability of solitons in randomly varying birefringent fibers." *Opt. Letters*, vol. 16, no. 16, pp. 1231–1233, 1991.

[5] P. K. A. Wai and C. R. Menyuk, "Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 14, no. 2, pp. 148–157, 1996.

[6] J. P. Gordon and H. Kogelnik, "PMD fundamentals: polarization mode dispersion in optical fibers," *Proc. of the National Academy of Sciences*, vol. 97, no. 9, pp. 4541–4550, 2000.

[7] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1745, 1997.

[8] C. Poole and R. Wagner, "Phenomenological approach to polarisation dispersion in long single-mode fibres," *Electron. Lett.*, vol. 22, no. 19, pp. 19–20, 1986.

[9] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightw. Technol.*, vol. 9, no. 11, pp. 1439–1456, 1991.

[10] F. Curti, B. Daino, G. De Marchis, and F. Matera, "Statistical treatment of the evolution of the principal states of polarization in single-mode fibers," *J. Lightw. Technol.*, vol. 8, no. 8, pp. 1162–1166, 1990.

[11] C. Poole and D. Favin, "Polarization-mode dispersion measurements based on transmission spectra through a polarizer," *J. Lightw. Technol.*, vol. 12, no. 6, pp. 917–929, 1994.

[12] G. Agrawal, *Fiber-Optic Communication Systems*. Wiley, 2012.

[13] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, B. Goebel, and S. Member, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, 2010.

[14] E. Desurvire, *Erbium-doped fiber amplifiers: principles and applications*. Wiley-Interscience, 2002.

[15] K. Kikuchi and S. Tsukamoto, “Evaluation of sensitivity of the digital coherent receiver,” *J. Lightw. Technol.*, vol. 26, no. 13, pp. 1817–1822, 2008.

[16] A. Yariv, “Signal-to-noise considerations in fiber links with periodic or distributed optical amplification,” *Opt. Letters*, vol. 15, no. 19, pp. 1064–1066, 1990.

[17] J. D. Ania-Castañón, I. O. Nasieva, S. K. Turitsyn, N. Brochier, and E. Pincemin, “Optimal span length in high-speed transmission systems with hybrid Raman-erbium-doped fiber amplification,” *Opt. Letters*, vol. 30, no. 1, pp. 23–25, 2005.

[18] A. Carena, G. Bosco, V. Curri, P. Poggiolini, M. T. Taiba, and F. Forghieri, “Statistical characterization of PM-QPSK signals after propagation in uncompensated fiber links,” in *Proc. European Conference on Optical Communication (ECOC)*, Torino, Italy, 2010.

[19] G. Gavioli, E. Torrengo, G. Bosco, A. Carena, V. Curri, V. Miot, P. Poggiolini, F. Forghieri, S. J. Savory, L. Molle, and R. Freund, “NRZ-PM-QPSK 16×100 Gb/s transmission over installed fiber with different dispersion maps,” *IEEE Photon. Technol. Lett.*, vol. 22, no. 6, pp. 371–373, 2010.

[20] P. J. Winzer, “Advanced modulation formats for high-capacity optical transport Networks,” *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4711–4728, 2006.

[21] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Springer US, 1999.

[22] J. Barry, E. Lee, and D. Messerschmitt, *Digital Communication*. Springer US, 2004.

[23] T. Hirooka, P. Ruan, P. Guan, and M. Nakazawa, “Highly dispersion-tolerant 160 Gbaud optical Nyquist pulse TDM transmission over 525 km,” *Opt. Express*, vol. 20, no. 14, pp. 15 001–15 007, 2012.

[24] M. Nakazawa, T. Hirooka, P. Ruan, and P. Guan, “Ultrahigh-speed orthogonal TDM transmission with an optical Nyquist pulse train,” *Opt. Express*, vol. 20, no. 2, pp. 1129–1140, 2012.

[25] A. Vedadi, M. A. Shoaie, and C.-S. Brès, “Near-Nyquist optical pulse generation with fiber optical parametric amplification,” *Opt. Express*, vol. 20, no. 26, pp. B558–B565, 2012.

[26] R. Schmogrow, D. Hillerkuss, S. Wolf, B. Bäuerle, M. Winter, P. Kleinow, B. Nebendahl, T. Dippon, P. C. Schindler, C. Koos, W. Freude, and J. Leuthold, "512QAM Nyquist sinc-pulse transmission at 54 Gbit/s in an optical bandwidth of 3 GHz," *Opt. Express*, vol. 20, no. 6, pp. 6439–6447, 2012.

[27] R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Spectrally shaped DP-16QAM superchannel transmission with multi-channel digital back-propagation." *Sci. Rep.*, vol. 5, pp. 1–8, 2015.

[28] M. A. Soto, M. Alem, M. Amin Shoaie, A. Vedadi, C.-S. Brès, L. Thévenaz, and T. Schneider, "Optical sinc-shaped Nyquist pulses of exceptional quality," *Nature Communications*, vol. 4, pp. 2898 EP–, 2013.

[29] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, 2008.

[30] ——, "Digital coherent optical receivers: algorithms and subsystems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, 2010.

[31] M. G. Taylor, "Coherent detection method using DSP to demodulation of signal and for subsequent equalisation of propagation impairments," *IEEE Photon. Technol. Lett.*, vol. 16, no. 2, pp. 674–676, 2003.

[32] M. Paskov, D. Lavery, and S. J. Savory, "Blind equalization of receiver inphase/quadrature skew in the presence of Nyquist filtering," *IEEE Photon. Technol. Lett.*, vol. 25, no. 24, pp. 2446–2449, 2013.

[33] A. Leven, N. Kaneda, U. V. Koc, and Y. K. Chen, "Frequency estimation in intradyne reception," *IEEE Photon. Technol. Lett.*, vol. 19, no. 6, pp. 366–368, 2007.

[34] M. Morelli and U. Mengali, "Feedforward frequency estimation for PSK: A tutorial review," *European Transactions on Telecommunications*, vol. 9, no. 2, pp. 103–116, 1998.

[35] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Trans. Inf. Theory*, vol. 29, no. 4, pp. 543–551, 1983.

[36] I. Fatadin, D. Ives, and S. J. Savory, "Compensation of frequency offset for differentially encoded 16- and 64-QAM in the presence of laser phase noise," *IEEE Photon. Technol. Lett.*, vol. 22, no. 3, pp. 176–178, 2010.

[37] E. Agrell, A. Alvarado, and F. R. Kschischang, “Implications of information theory in optical fibre communications,” *Phil. Trans. Royal Soc. A*, vol. 374, no. 2062, 2015.

[38] M. Secondini and E. Forestieri, “Scope and limitations of the nonlinear Shannon limit,” *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Feb 2017.

[39] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, “The GN-model of fiber non-linear propagation and its applications,” *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, 2014.

[40] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, “Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links,” *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, 2012.

[41] P. Poggiolini, “The GN model of non-linear propagation in uncompensated coherent optical systems,” *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, 2012.

[42] G. Bosco, P. Poggiolini, A. Carena, V. Curri, and F. Forghieri, “Analytical results on channel capacity in uncompensated optical links with coherent detection,” *Opt. Express*, vol. 19, no. 26, pp. B438–B449, 2011.

[43] E. Forestieri, *Optical Communication Theory and Techniques*. Springer US, 2004.

[44] S. J. Savory, “Approximations for the Nonlinear self-channel interference of channels with rectangular spectra,” *IEEE Photon. Technol. Lett.*, vol. 25, no. 10, pp. 961–964, 2013.

[45] A. Carena, G. Bosco, V. Curri, P. Poggiolini, and F. Forghieri, “Impact of the transmitted signal initial dispersion transient on the accuracy of the GN-model of non-linear propagation,” in *Proc. European Conference on Optical Communication (ECOC)*, London, UK, 2013.

[46] P. Serena and A. Bononi, “On the accuracy of the Gaussian nonlinear model for dispersion-unmanaged coherent links,” in *Proc. European Conference on Optical Communication (ECOC)*, London, UK, 2013.

[47] A. Mecozzi and R. J. Essiambre, “Nonlinear Shannon limit in pseudolinear coherent systems,” *J. Lightw. Technol.*, vol. 30, no. 12, pp. 2011–2024, 2012.

[48] M. Secondini and E. Forestieri, “Analytical fiber-optic channel model in the presence of cross-phase modulation,” *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, 2012.

[49] M. Secondini, E. Forestieri, and G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, 2013.

[50] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “Properties of nonlinear noise in long , dispersion-uncompensated fiber links,” *Opt. Express*, vol. 21, no. 22, pp. 25 685–25 699, 2013.

[51] G. Bosco, R. Cigliutti, A. Nespola, and A. Carena, “Experimental investigation of nonlinear interference accumulation in uncompensated links,” *IEEE Photon. Technol. Lett.*, vol. 24, no. 14, pp. 1230–1232, 2012.

[52] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, “EGN model of non-linear fiber propagation.” *Opt. Express*, vol. 22, no. 13, pp. 16 335–62, 2014.

[53] ——, “On the accuracy of the GN-Model and on analytical correction terms to improve it,” *arXiv*, pp. 1–10, 2014. [Online]. Available: http://arxiv.org/abs/1401.6946

[54] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, “A simple and effective closed-form GN model correction formula accounting for signal non-Gaussian distribution,” *J. Lightw. Technol.*, vol. 33, no. 2, pp. 459–473, 2015.

[55] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, 2007.

[56] G. Gao, X. Chen, and W. Shieh, “Influence of PMD on fiber nonlinearity compensation using digital back propagation.” *Opt. Express*, vol. 20, no. 13, pp. 14 406–18, 2012.

[57] R. Hardin and F. Tappert, “Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equation ,” *SIAM review*, vol. 15, no. 1, pp. 423–428, 1973.

[58] T. R. Taha, “A numerical scheme for the nonlinear Schrödinger equation,” *Computers & Mathematics with Applications*, vol. 22, no. 9, pp. 77–84, 1991.

[59] O. V. Sinkin, R. Holzlöhner, J. Zweck, and C. R. Menyuk, “Optimization of the split-step Fourier method in modeling optical fiber communications systems,” *J. Lightw. Technol.*, vol. 21, no. 1, pp. 61–68, 2003.

[60] G. Weiss and A. Maradudin, “The Baker-Hausdorff formula and a problem in crystal physics,” *J. Math. Phys.*, vol. 3, no. 4, pp. 771–777, 1962.

[61] R. W. Tkach, A. R. Chraplyvy, F. Forghieri, A. H. Gnauck, and R. M. Derosier, “Four-photon mixing and high-speed WDM systems,” *J. Lightw. Technol.*, vol. 13, no. 5, pp. 841–849, 1995.

[62] G. Bosco, A. Carena, V. Curri, R. Gaudino, P. Poggiolini, and S. Benedetto, “Suppression of spurious tones induced by the split-step method in fiber systems simulation,” *IEEE Photon. Technol. Lett.*, vol. 12, no. 5, pp. 489–491, 2000.

[63] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, “Polarization drift channel model for coherent fibre-optic systems,” *Sci. Rep.*, vol. 6, p. 21217, 2016.

[64] C. H. Prola, J. A. Pereira da Silva, A. O. Dal Forno, R. Passy, J. P. Von der Weid, and N. Gisin, “PMD emulators and signal distortion in 2.48-Gb/s IM-DD lightwave systems,” *IEEE Photon. Technol. Lett.*, vol. 9, no. 6, pp. 842–844, 1997.

[65] A. Galtarossa and C. Menyuk, *Polarization Mode Dispersion*. Springer, 2005.

3

Digital backpropagation: theoretical gains and practical limitations

3.1 Inverting fibre impairments: digital back propagation

In chapter 2, it was shown how propagation of signals through an optical fibre can be comprehensively described by the Manakov equation (see Eq. (2.34)). It was also shown how, according to the Manakov equation, distortions arise during propagation, which limit transmission performance.

An intuitive approach to solve this problem consists in attempting to *undo* such distortions once the signal is detected at the receiver by applying the inverse operation applied by the fibre channel. This task is made possible by the fact that pulse distortion is described by a differential equation, which is a deterministic operator whose output can be (at least numerically) predicted and, hence, inverted. DSP techniques are clearly suited for this task, as any operation can be performed on the received signal, provided that the sampling rate of the conversion from the analogue to the digital domain is high enough. Indeed, this guarantees that there is a one-to-one relationship between the analogue and the digital domain and no information is lost.

The algorithm which inverts digitally the Manakov equation starting from the received signal is called *digital back propagation* (DBP) and it represents the most popular nonlinearity compensation scheme for optical fibre communication. The perfect cancellation of fibre nonlinear impairments would linearise the fibre channel, increasing transmission rates as the transmitted power is increased, similarly to the AWGN case. Although, as discussed during this chapter, the perfect linearisation of the channel is not achievable, the benefits shown by the reduction of fibre nonlinear effects through DBP are often considerable, making this technique extremely attractive.

Reversing Eq. (2.34) means being able to reconstruct the transmitted field $\mathbf{A}(0,t)$ using as an input the received field $\mathbf{A}(z,t)$ at a generic fibre section z. In the absence of any noise source added along the propagation path, this can be achieved by integrating Eq. (2.34) in the *backward* direction, i.e. using as initial condition the field $\mathbf{A}(z,t)$ and inverting the integration boundaries. This is equivalent to solve Eq. (2.34) with the same integration boundaries but reversing the sign of z, which effectively means solving Eq. (2.34) with reversed sign parameters

$$\alpha \rightarrow -\alpha \tag{3.1}$$

$$\beta_2 \rightarrow -\beta_2 \tag{3.2}$$

$$\gamma \rightarrow -\gamma. \tag{3.3}$$

The information on $\mathbf{A}(z,t)$ can be captured by a coherent receiver, and the backward integration of Eq. (2.34) can be performed numerically using a DSP algorithm. The reconstruction of the transmitted optical field from the channel-distorted version recorded at the receiver, corresponds to what, for linear channels, is referred to as a *zero-forcing* equalisation scheme [1, Ch. 7]. DBP can be therefore regarded as a generalisation of a zero-forcing equaliser for a nonlinear channel. As is well-known, one of the potential performance limitations of such schemes is represented by the effect of the noise which is added to the signal before the equalisation. Indeed, as will be discussed later in this chapter, noise added along the fibre link represents a major bottleneck of DBP performance.

As seen in section 2.6.1, the SSFM is generally adopted to integrate Eq. (2.34) in the forward direction. Thus, the DBP algorithm, which implements a reverse integration of Eq. (2.34), uses the schemes shown in Fig. 2.15, where, however, each of the exponential operators are changed in sign. If the number of iterations of the algorithm is as low as one per span then the symmetric-step variant of the SSFM is preferred for accuracy reasons (section 2.6.1). On the contrary, when the number of iterations performed is much larger than one per span the bulk-step SSFM is an obvious choice.

As a final remark, it is important to mention that in the DSP chain shown in 2.15 the choice of the rate at which the input optical field is sampled requires particular care. Indeed, in a linear system, no bandwidth expansion can be observed and the Nyquist sampling theorem guarantees that a sampling rate higher than twice the signal bandwidth will incur no aliasing through the DSP chain. This problem is further discussed in 3.6.

In this chapter, DBP performance is analysed, with the aim of highlighting potential benefits and practical limitations. The scenario under investigation is a multi-span EDFA-amplified, dispersion-unmanaged system like that discussed in Chapter 2. For the first time, DBP performance is numerically investigated for a large-bandwidth transmission ($\approx$1 THz).

First, an analytical investigation on the ideal benefits achievable through DBP is presented. Such an investigation made use of the GN-model discussed in the previous chapter, which allows to obtain predictions on the received SNR. Such predictions are then numerically verified through the SSFM.

Secondly, the limitations of the algorithm with respect to practical system constraints are then analysed. The impact of a limited DSP complexity available at the receiver is studied. In particular, DBP performance is studied as a function of the number of iterations performed in the algorithm and of the sampling rate.

Finally, the impact of PMD on the DBP performance is assessed, using the Monte-Carlo method explained in section 2.6.3.

3.2 Literature review

Before the advent of coherent optical detection, many techniques aiming to compensate for fibre propagation impairments were already under study. The first work using digital techniques to undo optical fibre linear and nonlinear impairments was reported in [2] in 2005, in the context of incoherent detection systems. In this paper, a transmitter-side digital *pre-distortion* technique, was adopted to compensate for both chromatic dispersion and intra-channel nonlinearity (SPM) for a return-to-zero differential phase-shift keying transmission. An extended work was later published as a journal paper in [3].

The experimental demonstration in [2] was shortly followed by the pre-distortion work in [4], which, for the first time, reportedly claimed to use SSFM to ”backpropagate“ the waveforms to be transmitted. In [5], pre-distortion was for the first time performed in a WDM scenario, digitally backpropagating an entire set of WDM channels, hence compensating for both intra-channel and inter-channel nonlinearities. In [6], a combination of pre-distortion and post-compensation of nonlinearity was demonstrated for OFDM transmission using a one-step nonlinear phase compensation scheme.

In all the works mentioned above, nonlinearity compensation was performed at the transmitter and for direct-detection systems. The first fully coherent implementation of DBP as a receiver-side technique for mitigating nonlinear fibre transmission impairments was studied in [7]. In this numerical study, a WDM transmission was coherently detected using a phase-locked comb of wavelenghts and DBP was performed using a parallel implementation of the SSF method applied on a channel-by-channel basis. In a transmission of 10x10 Gbit/s channels with 20 GHz spacing, a 2 dBQ gain at optimum launch power and a 60% distance increase was demonstrated in the short distance regime (500/800 km).

Later on, in [8], DBP capability of compensating for both intra-channel and inter-channel nonlinearities in WDM systems was studied for the joint backpropagation of the entire transmitted optical field. The performance of this implementation, referred to as *total-field* NLSE (T-NLSE) in [8], was compared to the performance of the parallel implementation introduced in [7] (referred to as coupled NLSE or C-NLSE in [8]). It was shown that for coarsely spaced WDM channels C-NLSE results in a small penalty but, in turn, it significantly reduces the hardware requirement with respect to receiver bandwidth and sampling rate, which could be set to as low as 2 Sa/sym per receiver. Moreover, it was shown that, if one gives up on compensating for the FWM (as in the C-NLSE implementation) the requirement on the DBP step size can be significantly relaxed and, as a result, fewer iterations are required in the algorithm to achieve a target gain.

A detailed description of the DBP algorithm was first given in [9], together with a characterisation of the DBP performance as a function of implementation parameters, such as number of steps and oversampling factor and system parameters such as channel spacing and percentage of in-line CD compensation. This work was later extended to the dual-polarisation case in [10], where a first study on the effect of PMD on DBP was also performed.

Many works have since focused on either reducing the complexity of the algorithm or studying its fundamental performance limits. As an example, in [11, 12], simplified versions of the SSF algorithm to compensate only for SPM and XPM effects that were based on an early work in [13], were presented. Also, additional techniques were explored in order to reduce the amount of required DBP iterations, while preserving an acceptable compensation performance. As an example, in [14–16], Gao et al. showed that using a low-pass filter before DBP significantly improves the system performance if a reduced number of steps is used. More recently, in [17, 18] Secondini et al., modified the conventional SSF method using a frequency resolved log-perturbation method to solve NLSE [19–21]. Due to the higher accuracy of this method, it was shown that the complexity of DBP can be globally reduced for a target performance compared to the conventional SSF implementation.

The fundamental performance limits of DBP have been extensively explored in several literature contributions [10, 22–31]. The two main performance bottlenecks were recognised to be: i) the presence of signal-ASE NLI in multi-span amplified systems [24, 25, 27, 30, 32] and ii) PMD [22, 25, 29]. However, additional numerical validation is needed for the analytical results obtained, for instance, in [25, 30–32].

In particular, a gap of numerical validation is apparent in those transmission regimes where simulations are particularly intense, such as wideband optical transmission or long-haul links affected by PMD. The results in this chapter contribute to fill this gap.

The impact of the signal-ASE NLI (see section 2.5.3) in the context of DBP performance was first explored in [24, 32] by Rafique et al. In these works, the residual signal-ASE NLI was estimated by numerically comparing the residual noise after DBP, which includes ASE and signal-ASE NLI, and the case of ideal linear transmission, where only ASE is present. The results showed that DBP is unable to compensate for the signal-ASE generated along the link and, because of that, performance increasingly degrades compared to the linear case as transmission distance increases. A more accurate analytical description of this phenomenon, based on the work of Chen and Shieh in [33], was given in [25, 26]. Here, expressions for the maximum SNR gains given by DBP in presence of signal-ASE beat noise were introduced for the first time. Finally, more sophisticated analytical models to account specifically for the effect of signal-ASE on DBP were developed in [27] using SSF approach and, more recently, in [30], the EGN-model [34].

In [10, 22], the detrimental effect of PMD on the performance of DBP was numerically studied for the first time. An analytical model to predict this effect was then introduced in [25]. Despite an heuristic approach was also discussed in [35], the model in [25], which was specifically derived for OFDM transmission systems, remains so far the only rigorous analytical study of the effect of PMD on DBP. The main result of the analytical tool was that the limitations induced by PMD on DBP were much more stringent than the ones due to the signal-ASE beat noise. As a result, the problem of PMD in the context of nonlinear compensation schemes has continued to attract the attention of following works such as [29, 35–38]. For instance, in [29, 36, 38], the effect of PMD on DBP gains was numerically studied as a function of the DBP bandwidth, PMD parameter and transmission distance for Nyquist spaced high-SE single-carrier systems. Particularly, in [38], it was shown that for ultra-large bandwidth systems ($\approx$1 THz), the DBP SNR gain rapidly saturates as a function of the nonlinearity compensation (NLC) bandwidth, as one can achieve a gain 0.8 dB below the maximum compensation gain by only using less than a fourth of the transmitted bandwidth. The results of the work on PMD in [29] and [38] are part of the original contribution of this thesis and are shown in section 3.7. Strategies to counteract the loss of effectiveness of DBP in the presence of PMD were proposed in [37, 39]. Specifically, in [37], it was shown that, in a single channel transmission, the SNR penalty due to PMD is reduced by 1 dB using the introduced modified DBP scheme as opposed to the conventional DBP algorithm. Although not shown in this thesis, the results in [37] were obtained as a joint work with Czegledi et al. and represent a continuation of the work on PMD presented in this thesis. For this reason, it is discussed as part of the future work.

Recent research on DBP has focused on understanding the real potential of DBP to enable high SE long-haul optical fibre transmission, both numerically [12, 29, 40] or experimentally [41–47].

The impact of pragmatic system constraints on the DSP, such as computational complexity or non-ideal knowledge of fibre parameters, was studied both for single channel [10, 11, 48] and multi-channel DBP [29]. In particular, in [29] the Q^2-factor penalty was analysed for the first time for multi-channel DBP, relating the backpropagated bandwidth and key parameters for the algorithm complexity, such as the number of DBP iterations and sampling rate. Also the penalty was characterised as a function of the DBP nonlinear coefficient, on which an ideal knowledge is not always available. The results of this inverstigation are presented in section 3.6.

Experimental works have shown the effectiveness of DBP in improving performance even in practical transmission scenarios. In [41] Cai et al. showed that in a transoceanic fully loaded C-band transmission scenario, despite the relatively small gain of up to 1 dB in Q^2 factor, single channel DBP enabled each transmitted channel to perform below the required FEC threshold (which in [41] was estimated in 4.9 dB in Q^2 factor for a 28% overhead LDPC FEC). This resulted in a reach extension of up to 3000 km for transmission distances in the range 6000 km–10000 km. DBP was also experimentally characterised in the context of super-channels transmission in [44, 45, 45, 47]. In these works a 100% reach increase was shown for DP-16QAM and DP-64QAM modulation formats and a total optical bandwidth of approximately 70 GHz. In [47], multi-channel DBP performance was instead experimentally characterised with respect to different 400G super-channel configurations. Finally, in [42], DBP was proven significantly beneficial also in the context of single-span transmission. Indeed, in this experiment, a 520 Gbits/s super-channel was transmitted over 250 km single SSMF span with EDFA-only amplification, with DBP increasing the transmission reach by approximately 8% for a fixed 15% FEC overhead.

3.3 Optical fibre communication system under study

In section 2.3, two different types of optical fibre systems were discussed. When analysing the performance of optical fibre transmission, it is critical to specify which of the possible configurations is analysed. As pointed out in, this is due to the fact that the physical properties of each of these systems are significantly different and, as a result, the performance of a specific transceiver scheme may substantially vary.

In this chapter, both the analytical and the numerical investigation presented will be focusing on a specific configuration for an optical fibre communication system. The system under consideration, schematically illustrated in Fig. 3.1, is a multi-span, EDFA-amplified, dispersion unmanaged, optical link. Such a transmission system represents

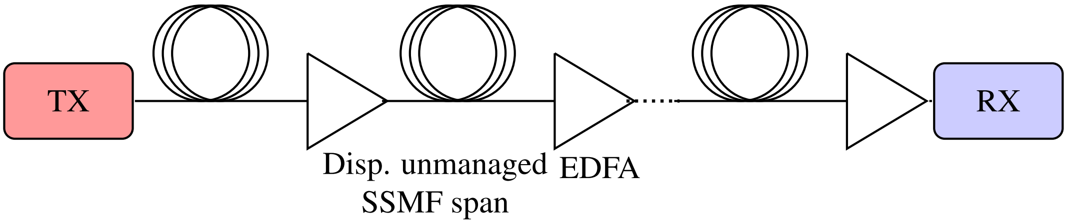


Figure 3.1: Schematic diagram of the optical fibre system under study in this chapter.

the standard for present-day long-haul fibre-optic links.

Although some system parameters were varied during the study, a *core* set of parameters can be identified as a reference. The fixed parameters are shown in table 3.1.

The transmitter consists of a PM-16QAM modulator at a 32 Gbaud rate. The spectrum of each transmitted channel was Nyquist shaped with a 1% roll-off and the spacing between the WDM channels was set to 33 GHz. Each WDM channel carries independent data and all of them are assumed to have the same transmitted power. No noise at the transmitter front-end was considered. The fibre link consists of multiple spans of SSMF with an ideally linear EDFA amplifier at the end of the span which exactly compensates for the span loss. No filtering effects are considered along the link due to possible add-and-drop multiplexers or other optical filters. The receiving front-end is assumed to be an ideal coherent receiver with no phase noise and no electrical noise. Fibre linear impairments such as the accumulated chromatic dispersion or the polarisation state rotation of the signal are ideally compensated [1] and a matched filter is used to select the bandwidth of the signal of interest.

Table 3.1: Parameters of the reference system used for the analytical and numerical study in this chapter.

Parameter Name	**Value**
Transmission Parameters	
Modulation Format	PM-16QAM
Symbol Rate	32 Gbaud
RRC Roll-Off	0.01
Channel Frequency Spacing	33 GHz
Fiber Channel Parameters	
Attenuation (α)	0.2 dB/km
Dispersion Parameter (D)	17 ps/nm/km
Nonlinearity Parameter (γ)	1.2 1/(W·km)
Fiber Span Length	80 km
EDFA Gain	16 dB
EDFA Noise Figure	4.5 dB

---

[1] The work *ideally* here refers to having at the receiver exact knowledge of the amount of in-line dispersion and state of polarisation of the received optical field.

3.4 Ideal DBP gain

One of the main performance parameters to characterise the performance of an equalisation scheme such as DBP is the SNR measured after DBP and subsequent matched filtering are applied. Although this does not necessarily represent the optimum receiver configuration (see chapter 4) it represents a good performance indicator for a matched filter based detection.

The GN-model (see chapter 2) provides a way to quantify the SNR at the matched filter output in presence of nonlinear distortion or when the nonlinear interference is partially or fully compensated for by the NLC scheme. As explained in chapter 2, according to any perturbative model, nonlinear distortions can be accounted for as additional noise. Thus, considering separately the different additive noise contributions, the SNR at the matched filter output after transmission over a fibre link is given by

$$\mathrm{SNR} = \frac{P}{N_s P_{\mathrm{ASE}} + \eta_{\mathrm{SS}} N_s^{(1+\varepsilon_{ss})} P^3 + \zeta \eta_{\mathrm{SN}} P_{\mathrm{ASE}} P^2} \tag{3.4}$$

where we indicate

P	Transmitted power per channel
N_s	Number of fibre spans
P_{ASE}	ASE noise power of a single EDFA and over the channel bandwidth
ε_{ss}	Coherence factor of the signal-signal term
η_{SS}	Signal-signal NLI factor
η_{SN}	Signal-ASE NLI factor
ζ	Signal-ASE NLI accumulation factor

Furthermore, the quantities at the denominator of the right-hand side of Eq. (3.4) can be separated as

$$P_{\mathrm{ASEtot}} \triangleq N_s P_{\mathrm{ASE}} \tag{3.5}$$

$$P_{\mathrm{SStot}} \triangleq \eta_{\mathrm{SS}} N_s^{(1+\varepsilon_{ss})} P^3 \tag{3.6}$$

$$P_{\mathrm{SNtot}} \triangleq \zeta \eta_{\mathrm{SN}} P_{\mathrm{ASE}} P^2 \tag{3.7}$$

where P_{ASEtot}, P_{SStot}, and P_{SNtot} are the total ASE noise power, total signal-signal NLI power and total signal-ASE NLI power, respectively.

The η_{SS} and η_{SN} are the so-called NLI factors for the signal-signal NLI and signal-noise NLI, respectively [28]. As discussed in section 2.5, these coefficients describe the proportionality of the signal-signal NLI and signal-noise NLI to P^3 and P^2, respectively, and they are functions of both transmission and fibre parameters (see section 2.5). Furthermore, their value depends on the specific model adopted. Although the domains of integration to derive η_{SS} and η_{SN} are in general different, it can be shown that (see Appendix A)

$$\eta_{\text{SN}} = 3\eta_{\text{SS}} \tag{3.8}$$

when the transmitted spectrum is perfectly flat, and the WDM bandwidth B is equal to the ASE bandwidth. This is typically the case as optical filters placed along the link are adjusted to the WDM bandwidth.

Thus, for simplicity we can refer to a single $\eta \triangleq \eta_{\text{SS}}$ and rewrite the SNR as

$$\text{SNR} \approx \frac{P}{N_s P_{\text{ASE}} + \eta N_s^{(1+\varepsilon_{ss})} P^3 + 3\zeta\eta P^2}. \tag{3.9}$$

Throughout this chapter the GN-model is used to make analytical predictions on the SNR with or without the use of DBP. The reason for using this model is in the availability of closed-form expressions for the η and ζ factors, through which a qualitative analysis can be more rapidly performed. However, the accuracy of such closed-form expressions is verified against numerical results based on SSFM simulations and shown in section 3.5.

The NLI term P_{SStot} represents the power of the additive term derived from a first-order perturbative solution of Eq. (2.34) (see Appendix). Since DBP calculates the solution of Eq. (2.34) in the reversed z direction, its effect on the SNR can be accounted for as a full or partial *reduction* of P_{SStot}. In contrast, P_{SNtot} cannot be cancelled because of its stochastic nature. However, the interaction between DBP and ASE noise is a relevant topic and requires an ad-hoc analysis. In the following sections we will closely analyse Eq. (3.9) and we will discuss the theoretical gains achievable via DBP whether it is applied on the entire transmitted signal bandwidth or just on a part of it.

3.4.1 Received SNR in absence of NLC

In absence of NLC, the receiver SNR after matched filtering and sampling is given by Eq. (3.9). The η and ζ factors can be expressed, according to the GN-model, in approximated closed-form given by Eq. (2.66) and Eq. (2.76), respectively.

By looking at Eq. (3.9) it can be observed that, while the total ASE noise power $N_s P_{\text{ASE}}$ has clearly no dependence on the transmitted power P, P_{SStot} and P_{SNtot} depend on P^3 and P^2, respectively. At low transmitted powers (linear region), P_{ASEtot} is the dominant term and the SNR increases at a rate of 1 dB/dB of transmitted power increase. For higher powers, P_{SStot} becomes significant with respect to P_{ASEtot} and the SNR increases with a rate lower than 1 dB/dB, until a maximum is achieved. In the assumption of $P_{\text{SStot}} \gg P_{\text{SNtot}}$, the optimum value of the SNR can be calculated approximating Eq. (3.9)

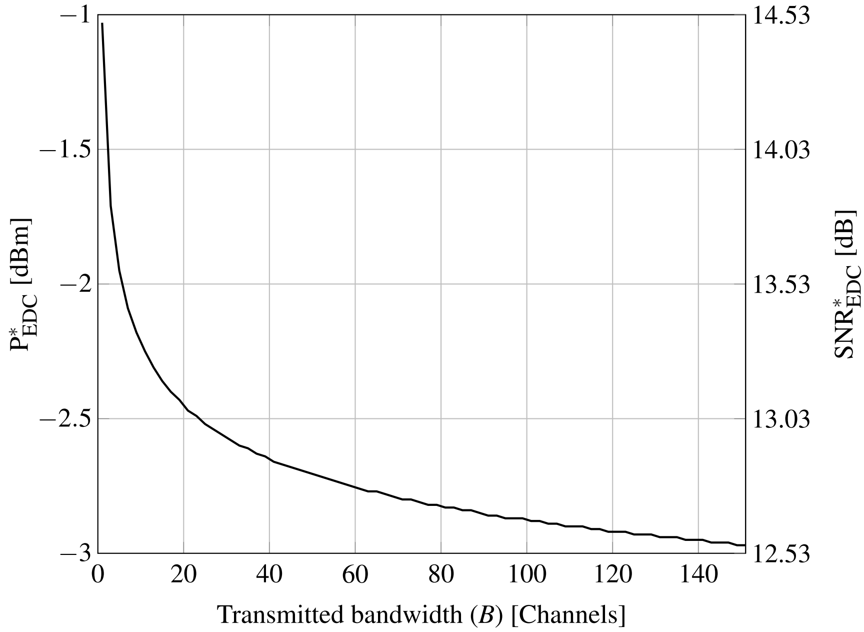


Figure 3.2: P^*_{EDC} and $\mathrm{SNR}^*_{\mathrm{EDC}}$ as a function of transmitted optical bandwidth B (expressed in number of 32 GBaud channels) for a 3200 (40x80) km transmission distance.

as

$$\mathrm{SNR} = \frac{P}{P_{\mathrm{ASEtot}} + P_{\mathrm{SStot}} + P_{\mathrm{SNtot}}} \approx \frac{P}{P_{\mathrm{ASEtot}} + P_{\mathrm{SStot}}}. \tag{3.10}$$

Such an assumption is based on the dominance of P term on P_{ASE} (for large enough values of P) and hence the dominance of P^3 with respect to the $P_{\mathrm{ASE}}P^2$ term [2]. Setting $\frac{\partial \mathrm{SNR}}{\partial P}$ to zero, we obtain the optimum value of the launch power P^*_{EDC} and the relative optimum SNR given by

$$P^*_{\mathrm{EDC}} = \sqrt[3]{\frac{P_{\mathrm{ASE}}}{2\eta N_s^{\varepsilon}}} \tag{3.11}$$

$$\mathrm{SNR}_{\mathrm{EDC}}(P^*_{\mathrm{EDC}}) \triangleq \mathrm{SNR}^*_{\mathrm{EDC}} = \frac{2}{3\sqrt[3]{2}} \frac{1}{P_{\mathrm{ASE}}^{2/3}\eta^{1/3}N_s^{(1+\varepsilon/3)}}. \tag{3.12}$$

It can be observed that the optimum launch power per channel $\mathrm{P}^*_{\mathrm{EDC}}$ presents a dependence on the ASE noise power ($\propto P_{\mathrm{ASE}}^{1/3}$), on the NLI coefficient ($\propto \eta^{-1/3}$), and very weakly on the transmission distance ($\propto N_s^{-(\varepsilon/3)}$). The optimum SNR instead is decreasing with respect to the ASE noise power ($\propto P_{\mathrm{ASE}}^{-1/3}$) and, almost linearly, with respect to the transmission distance ($\propto N_s^{-(1+\varepsilon/3)}$). Moreover it shows the same dependence than $\mathrm{P}^*_{\mathrm{EDC}}$ on the η factor.

In Fig. 3.2 $\mathrm{P}^*_{\mathrm{EDC}}$ and $\mathrm{SNR}^*_{\mathrm{EDC}}$ are plotted as a function of the transmitted bandwidth,

---

[2] For long enough distances Eq. (3.4) shows that P_{SNtot} can still be significant for P in the optimum launch power region, however such distances are much longer than any distance of interest for optical fibre communications.

using Eqs. (3.12), (3.11) and the GN-model based closed-form expressions in Eqs. (2.66) and (2.69) for the η and the ε coefficients, respectively. The system considered here is a 3200 (40x80) km transmission link with a number of channels spanning from 1 to 151 ($\approx$5 THz). It can be observed that the variation of both $\mathrm{P}^*_{\mathrm{EDC}}$ and $\mathrm{SNR}^*_{\mathrm{EDC}}$ is contained within only 2 dB out of which 1.5 dB are only caused by the first 21 transmitted channels.

3.4.2 Full-field nonlinearity compensation gain

When DBP is ideally applied on the entire received spectrum the signal-signal NLI is coherently cancelled and therefore P_{SStot} is forced to zero. The SNR after a full-field nonlinearity compensation is thus given by

$$\mathrm{SNR}_{\mathrm{DBP}} = \frac{P}{N_s P_{\mathrm{ASE}} + 3\zeta\eta P_{\mathrm{ASE}} P^2} \tag{3.13}$$

In this case, the optimisation of the launch power and the SNR gives

$$P^*_{\mathrm{DBP}} = \sqrt{\frac{N_s}{3\eta\zeta}} \tag{3.14}$$

$$\mathrm{SNR}^*_{\mathrm{DBP}} = \frac{1}{2\sqrt{3}} \frac{1}{P_{\mathrm{ASE}}} \sqrt{\frac{1}{3\eta\zeta N_s}}. \tag{3.15}$$

Truncating the ζ factor to the second term, as shown in 2.76, and substituting in Eqs. (3.14) and (3.15), we obtain

$$P^*_{\mathrm{DBP}} \approx \sqrt{\frac{2+\varepsilon}{3}} \frac{1}{\eta^{1/2} N_s^{(1+\varepsilon)/2}} \tag{3.16}$$

$$\mathrm{SNR}^*_{\mathrm{DBP}} \approx \frac{1}{2} \sqrt{\frac{2+\varepsilon}{3}} \frac{1}{\eta^{1/2} P_{\mathrm{ASE}} N_s^{(3/2+\varepsilon/2)}}. \tag{3.17}$$

It can be noted that, differently from Eq. (3.11), the optimum power in this case does not depend on the ASE noise power but only on the η factor and on the distance ($\propto N_s^{(1+\varepsilon)/2}$). However, the optimum SNR depends on the main system parameters including η factor, P_{ASE} and (strongly) on the transmission distance, as it decays as $\approx N_s^{-3/2}$.

To illustrate the scaling laws shown by the formulas in Eqs. (3.14), and (3.15), in Fig. 3.3 two contour plots show P^*_{DBP} and $\mathrm{SNR}^*_{\mathrm{DBP}}$ as a function of launch power after full-field DBP is applied.

For shorter distances ($\leq$ 4000 km), both P^*_{DBP} and $\mathrm{SNR}^*_{\mathrm{DBP}}$ show a weak dependence on the transmitted bandwidth when the number of transmitted channels is higher than

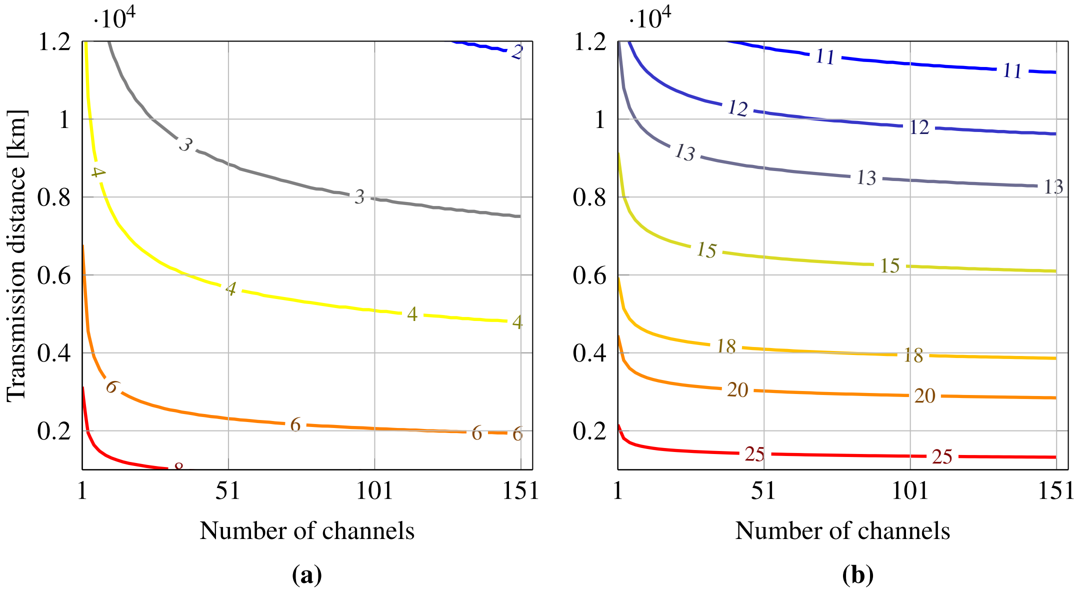


Figure 3.3: Contour plots of (a) P^*_{DBP} and, (b) $\text{SNR}^*_{\text{DBP}}$, as a function of the number of transmitted channels and transmission distance.

51 (B ≈1.65 THz). This is the region where P^*_{DBP} and $\text{SNR}^*_{\text{DBP}}$ assume the largest values. It can also be observed that for high values of P^*_{DBP} and $\text{SNR}^*_{\text{DBP}}$ and for very small transmitted bandwidths ($\leq$ 10 channels), further reducing the number of channels results in an abrupt increase in the transmission distance. For larger values of P^*_{DBP} and $\text{SNR}^*_{\text{DBP}}$, the trade-off between distance and number of channels transmitted is instead smoother. Finally, Fig. 3.3 shows that while P^*_{DBP} varies in a 6 dB range, the relative $\text{SNR}^*_{\text{DBP}}$ has a range of variation of $\approx$ 15 dB. This demonstates that the relationship between P^*_{DBP} and $\text{SNR}^*_{\text{DBP}}$ is not linear across different transmission distances and bandwidths.

Using the SNR expressions in Eqs. (3.12) and (3.17), the gain for an ideal full nonlinearity compensation can be written as

$$G_{\text{DBP}} \triangleq \frac{\text{SNR}^*_{\text{DBP}}}{\text{SNR}^*_{\text{EDC}}} = k\frac{1}{\eta^{1/6} P_{\text{ASE}}^{1/3} N_s^{1/2+\varepsilon/6}} \tag{3.18}$$

where k is a constant given by

$$k = \frac{3\sqrt[3]{2}}{4}\left(\frac{2+\varepsilon}{3}\right). \tag{3.19}$$

Eq. (3.18) shows that DBP gain is very weakly related to the nonlinearity factor η ($\propto \eta^{-1/6}$), to the ASE noise introduced by each EDFA amplifier ($\propto P_{\text{ASE}}^{-1/3}$) and to the transmission distance (approximately $\propto N_s^{-1/2}$). Furthermore, as shown in Eq. (2.66), the relationship between the NLI factor and the compensation bandwidth B_{DBP} is not linear, such that

$$G_{\text{DBP}} \propto \operatorname{arcsinh}\left(\frac{\pi^2}{2} L_{\text{eff}} B\right)^{-1/6}. \tag{3.20}$$

The scaling laws shown in Eqs. (3.18), (3.20) are of particular interest for the transceiver design of an optical fibre link as it indicates what is the maximum gain that a full NLC scheme can achieve compared to an EDC receiver. Fig. 3.4 illustrates such variation as a function of the number of transmitted channels and transmission distance. The contour lines in Fig. 3.4 show the weak variation of G_{DBP} on the number of transmitted channels, except in the case where the number of channels is low. For instance, for distances smaller than 4000 km the full-field DBP gain stays approximately constant from 11 channels up to 151 channels. A stronger variation can be appreciated with respect to the transmission distance, where a 5 dB/decade drop can be found in accordance with Eq. (3.18).

Such a drop in gain is to be attributed to the increasing amount of signal-ASE NLI as a function of N_s, (see Eq. (2.76)).

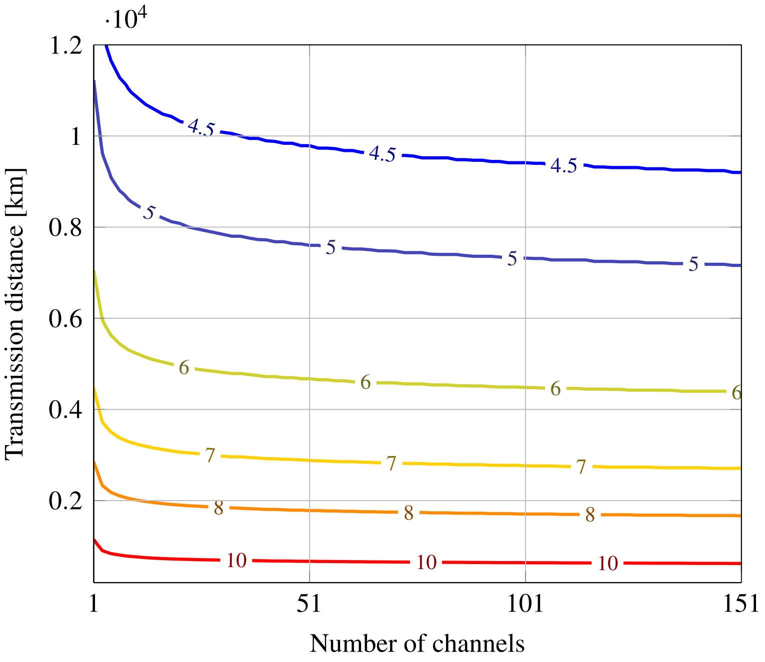


Figure 3.4: Contour plot of DBP SNR gain (G_{DBP}) in dB, as a function of the number of transmitted channels and transmission distance.

3.4.3 Intermediate bandwidth DBP

The case of full nonlinearity compensation often represents an ideal case for WDM systems, as jointly applying DBP over the entire transmitted bandwidth is an unrealistic task due to the limited receiver bandwidth and processing resources. Moreover, in networking scenarios, only a limited amount of channels will co-propagate along the same path between two network nodes. This makes DBP effective only over a limited portion of the WDM spectrum. It is, therefore, of great interest to understand what can be achieved by using DBP over a limited bandwidth.

To illustrate the dependence of DBP gain as a function of the compensation bandwidth B_{DBP} we again resort to a perturbational approach. To a first order approximation, DBP will cancel only the NLI generated within the spectral components that are captured by the receiver, while keeping the residual NLI unmodified [21]. This leads to a very simple estimation of the performance of partial bandwidth NLC using DBP, where the SNR after DBP can be expressed as

$$\text{SNR}_{\text{DBP}} = \frac{P}{P_{\text{ASE\_tot}} + [\eta(B) - \eta(B_{\text{DBP}})]\, P^3 N_{\text{s}}^{(1+\varepsilon)} + 3\eta\zeta P_{\text{ASE}} P^2}. \tag{3.21}$$

In the assumption that DBP is operated over a compensation bandwidth B_{DBP} much smaller that the full-field bandwidth B, Eq. (3.21) can be approximated as

$$\text{SNR}_{\text{DBP}} \approx \frac{P}{P_{\text{ASE\_tot}} + [\eta(B) - \eta(B_{\text{DBP}})]\, P^3 N_{\text{s}}^{(1+\varepsilon)}}. \tag{3.22}$$

In order to fully understand the efficacy of NLC schemes it is important to relate the reduction of the signal-signal NLI to the increment in the optimum launch power and in the optimum SNR. We can recognise two regions of operation: the case $B_{\text{DBP}} << B$, where Eq. (3.22) holds; $B_{\text{DBP}} \approx B$ where Eq. (3.22) does not hold and another approximation has to be used in order to find simple closed-form scaling rules. In the first case, the form of Eq. (3.22) is analogous to Eq. (3.9) where η is now replaced by the residual nonlinearity factor after DBP $\eta - \eta(B_{\text{DBP}})$. Thus, the analysis performed in section 3.4.1 also holds here but using the residual nonlinearity factor. From Eq. (3.12) it can be seen that for $B_{\text{DBP}} \ll B$

$$G_{\text{DBP}} \propto \sqrt[3]{\frac{\eta(B)}{\eta(B) - \eta(B_{\text{DBP}})}} \tag{3.23}$$

where

$$\rho \triangleq \frac{\eta(B)}{\eta(B) - \eta(B_{\text{DBP}})} \tag{3.24}$$

is a coefficient indicating to what extent nonlinearity is suppressed by DBP. The same scaling can be observed for the increase in optimum launch power ΔP^* from Eq. (3.11). The relationship in Eq. (3.23) indicates for that if DBP reduces P_{SStot} by 3 dB (half of the original NLI power) the expected optimum SNR gain will be 1 dB.

In the second case, for the NLC bandwidth B_{DBP} varying in a neighbourhood of the full-field bandwidth B, P_{SNtot} becomes dominant compared to P_{SStot} and Eq. (3.22) can be approximated by Eq. (3.13). However, the accuracy of this approximation for a given B_{DBP} close to B depends on the amount of P_{SNtot} which is in turn dependent on the transmission distance and ASE noise power, as shown by Eq. (3.7).

The dependence of G_{DBP} on the NLI suppression is illustrated in Fig. 3.5. The plot shows G_{DBP} vs. ρ for three transmission distances. The asymptotic behaviour of G_{DBP} is also highlighted for the two DBP operation regions: $B_{\text{DBP}} \ll B$ (small NLC compensation bandwidth) and $B_{\text{DBP}} \approx B$ (full-field DBP region). It can be observed that, for small NLC bandwidths, G_{DBP} increases by 0.33 dB/dB of NLI suppression, as predicted by Eq. (3.23). Such asymptotic behaviour is independent on both the transmission distance and the P_{SNtot} as in such region it is negligible compared to P_{SStot}. For large NLI suppression factors the gain saturates to a value which is dependent on the amount of signal-ASE noise and, thus, on the transmission distance.

Making the relationship with the transmitted WDM bandwidth B explicit, as in Eq. (3.20), we obtain

$$G_{\text{DBP}} \propto \sqrt[3]{\frac{\operatorname{arcsinh}(aB^2)}{\operatorname{arcsinh}(aB^2) - \operatorname{arcsinh}(aB_{\text{DBP}}^2)}}, \tag{3.25}$$

where $a = \frac{\pi^2}{2} L'_{\text{eff}}$.

The dependency of the DBP gain on the compensation bandwidth described by Eq. (3.25) is illustrated in Fig. 3.6. In Fig. 3.6a G_{DBP} is plotted as a function of the normalised NLC bandwidth defined as the fraction of the transmitted bandwidth B_{DBP}/B over which DBP is applied. These graphs are shown for a 151 channel transmission

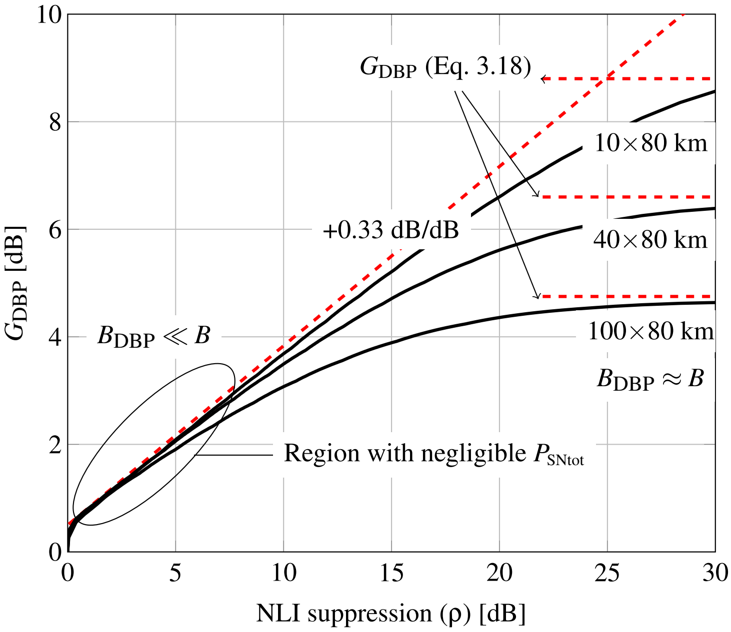
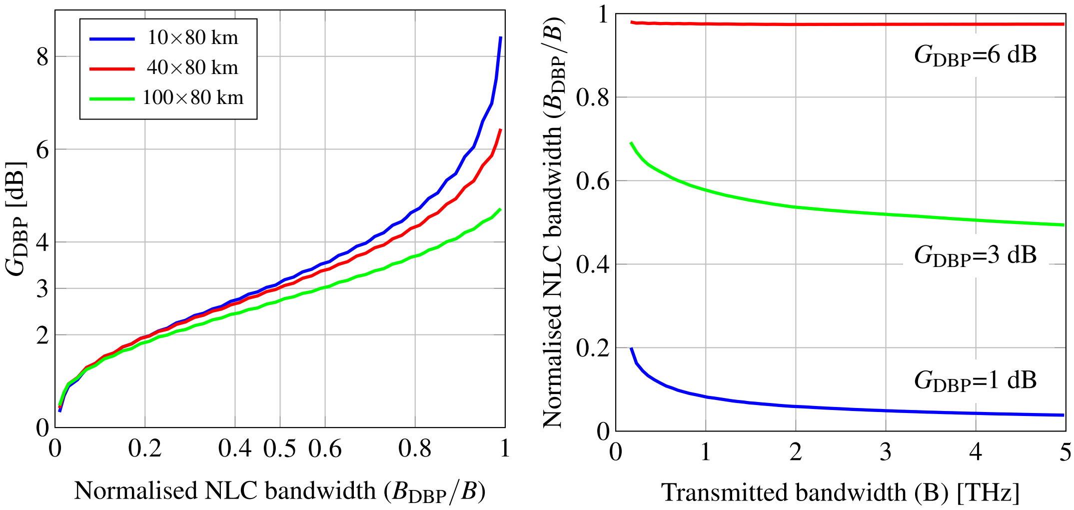


Figure 3.5: DBP SNR gain as a function of the NLI suppression factor (ρ) for different transmission distances.

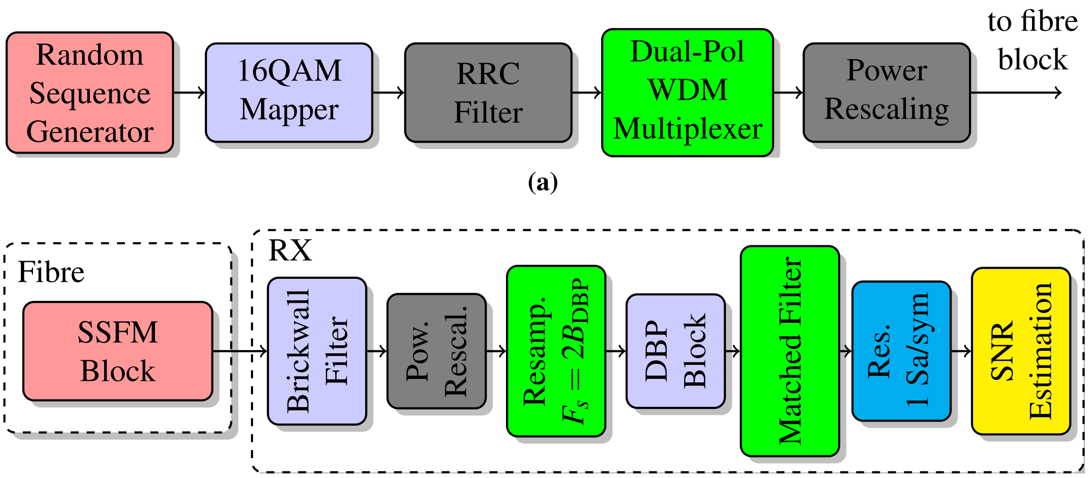


(a) **(b)**

Figure 3.6: DBP SNR gain as a function of (a) the NLC bandwidth for different transmission distances, and (b) required NLC bandwidth as a function of the overall transmitted bandwidth B for different DBP SNR gains.

corresponding to B≈ 5 THz and for the three transmission distances 10x80 km, 40x80 km and 100x80 km also shown in Fig. 3.5. It can be observed that in order to obtain a 3 dB gain in SNR the amount of compensation bandwidth B_{DBP} has to be at least 50% of the transmitted one or higher, depending on the transmission distance. As B_{DBP} approaches the transmitted bandwidth B, the gain slope increases and achieves its peak in the surrounding of $B_{\mathrm{DBP}} = B$. The faster gain increase for compensation bandwidths close to the full-field bandwidth can be explained due to the full cancellation of the signal-signal NLI and the residual signal-ASE term which is typically much smaller. This is confirmed by the behaviour of the three curves for the different transmission distances in Fig. 3.6: at shorter distances, where the signal-ASE NLI is smaller, the gain increase close the full-field compensation bandwidth appears more abrupt.

Because of the nonlinear relationship between B_{DBP}, B and G_{DBP}, shown in Eq. (3.25), the trend illustrated in Fig. 3.6a cannot be generalized to any transmitted bandwidth B. In order to better understand the impact of B, in Fig. 3.6b the relationship between B and B_{DBP} is plotted for fixed values of G_{DBP}. As it can be seen, the percentage of the transmitted bandwidth that we require to backpropagate to achieve gains for instance of 1 or 3 dB is significantly larger when such bandwidth is smaller. However, for higher values of G_{DBP} (e.g. see G_{DBP}=6 dB in Fig. 3.6b) the fraction of required B_{DBP} stays approximately constant regardless of the transmitted bandwidth B.



(a)

(b)

Figure 3.7: Schematic diagram of the system adopted for the numerical simulation of DBP performance: (a) transmitter, and (b) fibre block followed by receiver structure.

3.5 Numerical results on DBP performance

In order to validate the previously introduced analytical expressions and evaluate their accuracy, a comparison with numerical results obtained using SSFM simulations is presented in this section.

3.5.1 Numerical setup

Numerical simulations were performed based on the schematic diagram presented in Fig. 3.7. $2N_{\text{WDM}}$ different sequences of integer numbers (between 0 and 15), where N_{WDM} is the number of WDM channels transmitted were first generated. As a result a different independent data sequence was produced for each of the two orthogonal polarisations on each WDM channel. The integer sequences were then mapped onto a 16QAM constellation which were then shaped using an ideal frequency-domain RRC filter. A PM-WDM multiplexer then encoded each analogue signal onto a polarisation channel at a specified WDM wavelength.

The so-obtained WDM signal is first rescaled to match the specific launch power requirement and then fed into a SSF block emulating the optical fibre propagation. Specifically, such block numerically implements the Manakov equation in Eq. (2.34). The step distribution was selected adaptively to the transmitted power and following the log-step approach described in section 2.6.2. At the end of each fibre span an ideally linear EDFA is added, whose parameters are shown in table 3.1.

The first block of the receiver part is an ideal electro-optical front end which selects the signal at the fibre output over the specified bandwidth B_{DBP} using a brick-wall filter. The filtered signal is then passed to a resampling block which adjusts sampling rate to $F_s = 2B_{\text{DBP}}$, i.e. two times the Nyquist rate. Such operation makes sure the sampling

Table 3.2: Parameters used in numerical simulations for the results presented in this chapter.

Parameter Name	Value
Transmitted symbols	2^{18}
SSF sampling rate	$2B$
SSF step size	Adaptive log-step
DBP sampling rate	$2B_{\mathrm{DBP}}$
DBP step size	Adaptive log-step

rate is always adapted to the compensation bandwidth. Thus, as discussed in section 3.6.2, the complexity of the DBP algorithm is reduced while maintaining its highest accuracy. Before going into the DBP block the signal is properly rescaled to match the exact signal power (per channel) exhibited at the fibre output. In order to ideally operate the DBP algorithm, the inverse of the Manakov equation is implemented using the same step size distribution used in the forward propagation emulation. The signal is then passed to a matched filter to select the central channel and further down-sampled at 1 sample/sym.

The SNR in Eq. (3.9) is numerically calculated using a fully data-aided approach. Such approach consists in isolating each symbol's noisy cloud based on the knowledge of transmitted sequence. Firstly, for each noisy cloud, the expected value s_i for $i = 1,2,...,M$ is first estimated, where M is the constellation cardinality. Secondly, the variance of the residual noise σ_i^2 is calculated (additive noise assumption). The average SNR is then calculated as

$$\mathrm{SNR} = \frac{\sum_{i=1}^{M} s_i^2}{\sum_{i=1}^{M} \sigma_i^2}. \tag{3.26}$$

The accuracy of such estimation, which depends on the number of transmitted symbols, was found to be very high in the SNR region of interest. The parameters of the numerical simulations performed for the results in this chapter are shown in table 3.2.

3.5.2 EDC SNR performance

In order to test the accuracy of Eq. (3.9) in Fig. 3.8 numerical results on the SNR are shown for a transmission over a 40x80 km link and for 5x32 Gbaud and 31x32 Gbaud channels. The discrepancy between the values of the optimum SNR calculated from the closed-form expression and SSFM is $\leq$ 0.5dB and always results in an underestimation of the SNR for the closed-form formula. As already discussed in 2.5, this behaviour is due to the GN approximation of the signal as a Gaussian stochastic process which disregards the properties of the transmitted modulation format (16QAM in this case). The difference between the analytically predicted SNR and the numerical results increases to up to 0.6 dB at optimum launch power and 1 dB in the nonlinear regime. Such a gap represents the ratio between the NLI coefficient predicted by the GN and the actual one.

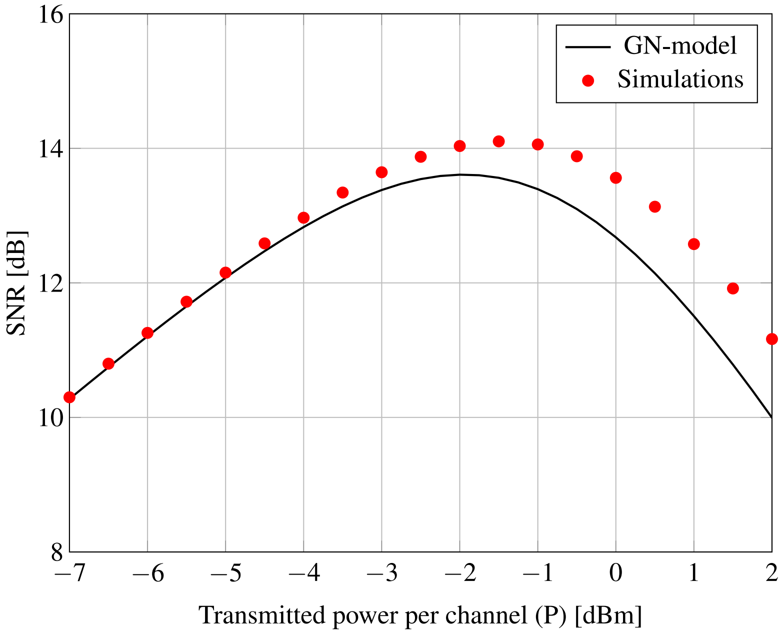


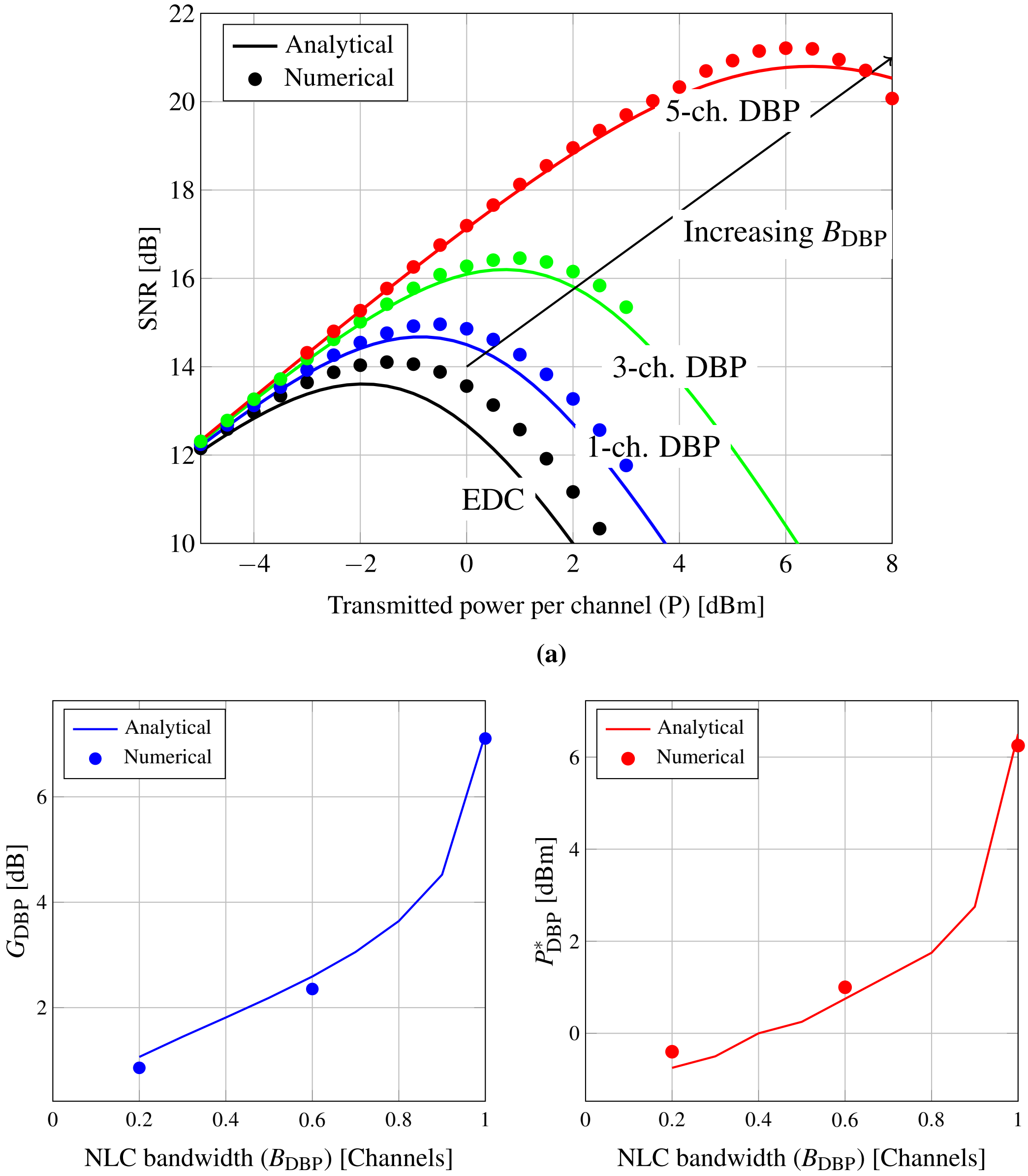
Figure 3.8: EDC receiver SNR as a function of the transmitted power per channel P in the absence of NLC and for a 5×32 Gbaud transmission over 40x80 (3200) km distance.

A more accurate analytical prediction of the η can be given by the EGN-model and the model discussed in 2.5. For both of them, unfortunately, closed-forms are not available and only numerical integration is possible. The scaling rules of the optimum transmitted power and the optimum SNR shown in Eq. (3.10) can also be checked from Fig. 3.8. In this case, as can be seen from Eqs. (3.11) and (3.12), the accuracy of the GN-model does not impact the accuracy of the scaling rule, provided that the NLI coefficient η remains the same. As for the bandwidth variation, it can be observed that the optimum launch power decreases by 1 dB going from 5 channels to 33 channels. Following Eq. (3.13) this corresponds to an increase of 2 dB in the η factor. The same trend is confirmed for a different transmission distance.

3.5.3 DBP SNR performance

In Figs. 3.9 and 3.10 the SNR vs. transmitted power per channel P is shown, when DBP is operated over different compensation bandwidths. Two different scenarios are shown: Fig. 3.9 shows the case of 5 WDM transmitted channels whereas in Fig. 3.10 the case of 31 channels is shown. In both cases the DBP compensation bandwidth B_{DBP} is varied from 1-channel DBP up to the full-field bandwidth B. The EDC performance is also shown as a reference.

Fig. 3.9 shows the 5 transmitted channels case. Solid lines represent the analytical expressions in Eqs. (3.9) and (3.21), whereas the circle markers represent the numerical results. In 3.9a the SNR vs. launch power per channel is shown as the NLC bandwidth



(a)

(b)

(c)

Figure 3.9: DBP SNR performance for a transmission of 5x32 GBaud PM-16QAM channels over 3200 km: (a) SNR as a function of P for EDC and DBP over different NLC bandwidths; (b) DBP SNR gain as a function of NLC bandwidth; (c) optimum transmitted power (P^*_{DBP}) as a function of NLC bandwidth.

is varied between 1 channel and 5 channels by steps of two channels at a time. The agreement between analytics and numerics is within 0.5 dB for the optimum SNR for all curves shown. As already discussed for Fig. 3.8, the EDC SNR performance is underestimated by the GN-model as we enter the nonlinear regime. Although the error is lower, this is also the case for the SNR when DBP is applied. It can also be noticed that, for the full-field DBP case, again the SNR is underestimated, indicating a modulation format dependency of the η_{SN} factor. However, the optimum SNR discrepancy as the full-field bandwidth is approached can be attributed to two reasons: there is a difference, albeit small, between the η_{SN} and $3\eta_{\mathrm{SS}}$ factors (see appendix); both η_{SS} and

η_{SN} are different from the η predicted by the GN-model, as they both are modulation format-dependent.

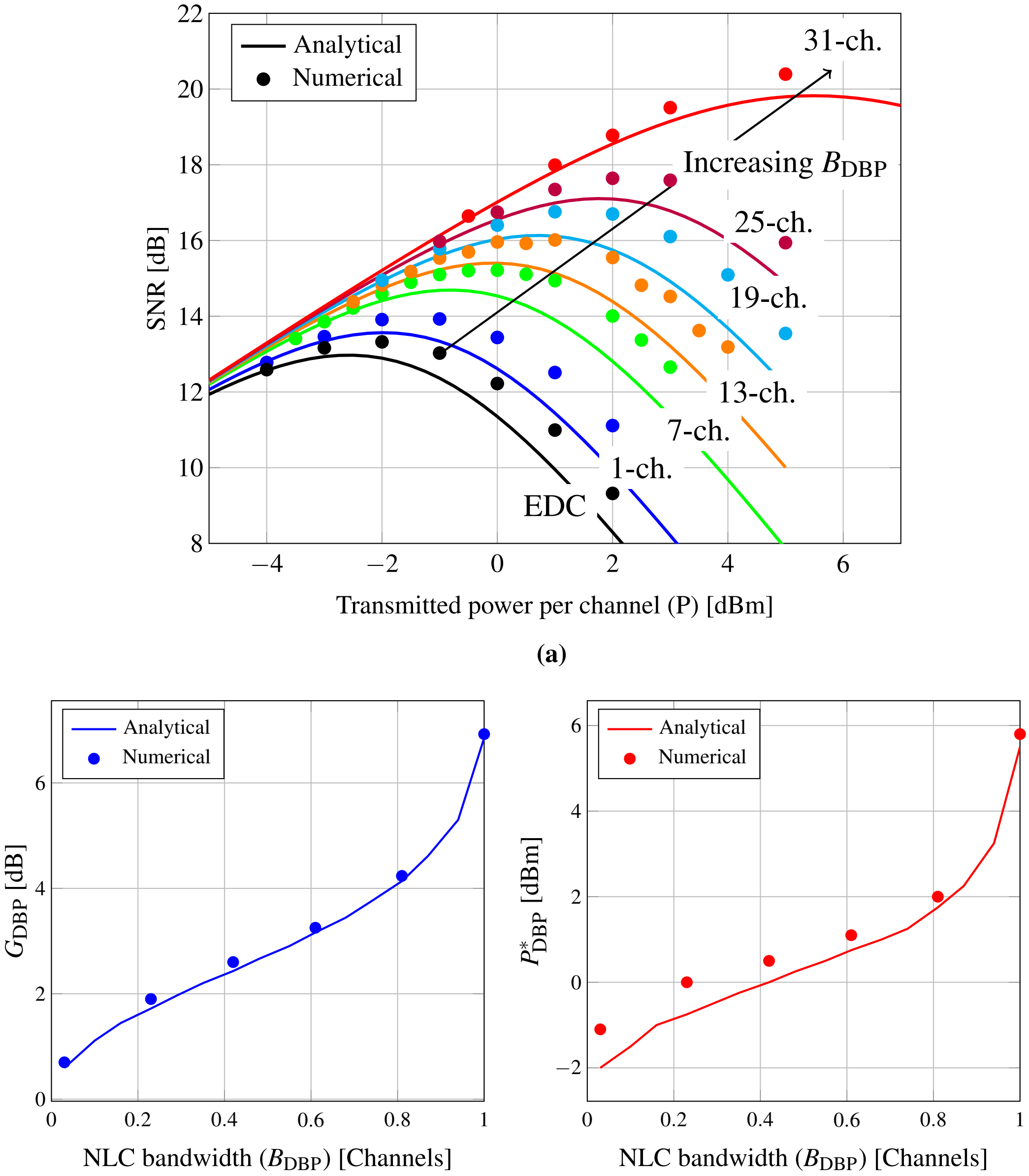
In Figs. 3.9b and 3.9c the DBP gain and the optimum launch power are shown as a function of the NLC bandwidth, respectively. Again, a good agreement between the analytical expressions and the numerical results can be observed. It can also be seen that the DBP gain is overestimated by the GN-model, whereas the optimum launch power is underestimated, as it can be expected from an overestimation of η factor.

The accuracy of the analytical model was also tested in the context of a large bandwidth transmission scenario. Numerical results on the DBP performance were obtained for 31x32 GBaud channels with an overall transmitted optical bandwidth of approximately 1 THz and are shown in Fig.3.10.

In Fig. 3.10a the SNR vs. launch power is shown for the EDC case and for different NLC bandwidths spanning from 1 channel up to 31 channels (full-field bandwidth). Again the analytical curves are in good agreement with the numerical results with the model underestimating the SNR in the nonlinear regime, as already discussed for the 5-channel scenario above. The only inaccuracy in the SNR prediction can be still attributed to the modulation format-dependent missing term (see discussion in chapter 2), which appears to be independent on the transmitted optical bandwidth. This inaccuracy is not reflected in the G_{DBP} vs. B_{DBP} plot in Fig. 3.10b, where it can be seen that the numerical results are in perfect agreement with the analytical results. This is due to the fact that the underestimation of the optimum SNR given by the GN-model is roughly equal for all the NLC bandwidths, except for the full-field bandwidth case. Higher inaccuracy can be observed for the optimum launch power shown in Fig. 3.10c as a function of the B_{DBP}. The optimum power is underestimated by up to 1 dB for 1-channel DBP. As B_{DBP} increases the agreement between numerics and analytical predictions improves.

Overall, the good predictions provided by the closed-form expressions shown in Fig. 3.10 confirm the validity of the model in the large bandwidth transmission scenario, where numerical validation is computationally intense.

Finally, the scaling law of DBP SNR performance as a function of the transmission distance was numerically studied. Fig. 3.11 shows the DBP gain G_{DBP} as a function of the transmitted distance for the 5-channel transmission scenario already studied in Fig. 3.9. The numerical results (circle markers) are compared with the closed-form expression in Eqs. (3.12) and (3.18) (solid lines) for three NLC bandwidths. The results in Fig. 3.11 show that a very good agreement is found between the scaling laws described in Eqs. (3.12) and (3.18) for the transmission distance and the numerically evaluated SNR at all NLC bandwidths presented. Particularly, the gain for the partial NLC bandwidths cases (i.e. 1-channel and 3-channel DBP shown by the blue and the green curve, respectively) is substantially flat across all distances. The only variation at



(a)

(b)

(c)

Figure 3.10: DBP SNR performance for a transmission of 31x32 GBaud PM-16QAM channels over 3200 km: (a) SNR as a function of P for EDC and DBP over different NLC bandwidths; (b) DBP SNR gain as a function of NLC bandwidth; (c) optimum transmitted power (P^*_{DBP}) as a function of NLC bandwidth.

short distances is given by the slight variation in the ε coefficient that is dependent on the NLC bandwidth. Such discrepancy disappears for large transmission distances. In the full-field case (red curve), the DBP gain decreases approximately as the square root of the transmission distance. Numerical results clearly confirm such a trend, although an underestimation of the gain can be noticed in general, due to the η factor overestimation of the GN-model previously discussed.

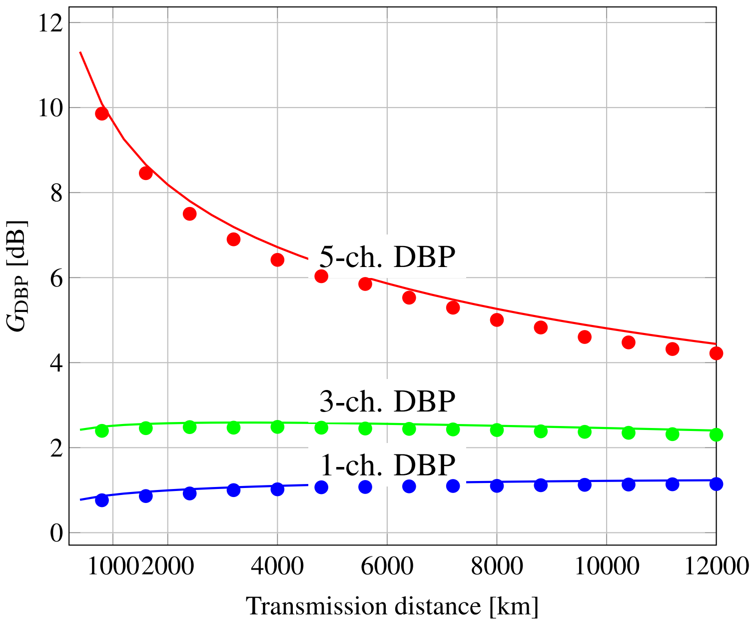


Figure 3.11: DBP SNR gain as a function of the transmission distance for a 5-channel system and for three different NLC bandwidths.

3.6 Practical limitations on DBP performance

The performance of DBP investigated in the previous sections of this chapter assumes an *ideal* implementation of the algorithm. We refer to DBP implementation as *ideal* when the algorithm is performed with a number of iterations (steps) and at sampling rate high enough to guarantee the required accuracy and, thus, to perform an ideal signal-signal NLI cancellation within the NLC bandwidth.

This implementation often implies a, currently unfeasible, computational demand on the DSP at the receiver. The impossibility of providing such resources has therefore driven research to investigate the trade-off in performance vs. computational complexity. Also, the analysis of the limitations of DBP performance due to pragmatic constraints provides an insight on why, in many experimental results, due to the reduced DSP complexity, DBP has shown significantly lower gains than the ones promised by the theory [9, 10, 49].

In this section we will discuss the performance/DSP complexity trade-off for multichannel DBP. Particularly the analysis will focus on the two main DSP parameters impacting the complexity of the algorithm, such as number of steps and sampling rate. Numerical results will be shown for a 5 channel transmission. However, it is argued that qualitatively similar results can be obtained as the number of transmitted channels is increased.

3.6.1 Number of iterations in DBP algorithm

As discussed in section 3.1, the conventional DSP implementation of the DBP algorithm involves a number of iterations. Each iteration includes two FFTs, and two point-wise complex multiplications (see Fig. 2.15). As each iteration is equally complex, the complexity of the DBP algorithm scales linearly with the number of steps required. On the other hand, in section 2.6.1 it was shown how the accuracy of the numerical integration of Eq. (2.34) scales quadratically on the step size, and as a result on number of steps used for the integration of a finite section of fibre. These two opposite requirements highlight a stringent trade-off between the performance and maximum number of iterations allowed for a fixed implementation complexity.

The impact of employing a limited number of iterations was first studied in [8] for the single-channel DBP case. However, as the DBP is applied to larger bandwidths, in order to correctly capture faster dispersion phenomena, a finer spatial discretisation is needed [11]. Particularly, it was suggested that the accuracy scales quadratically with the bandwidth [40].

In order to characterise the impact of the spatial resolution and the number of iterations required for a given NLC bandwidth, numerical results are shown in Fig. 3.12 for a 5-channel transmission over 3200 (40x80) km. The plot shows the DBP gain as a function of the number of steps utilised per fibre span for three different NLC bandwidths (1, 3 and 5 channels). As expected, larger NLC bandwidths require a larger number of iterations to achieve the ideal gain predicted from the theory for that given DBP bandwidth (see section 2.6.1). The minimum required number of steps to achieve

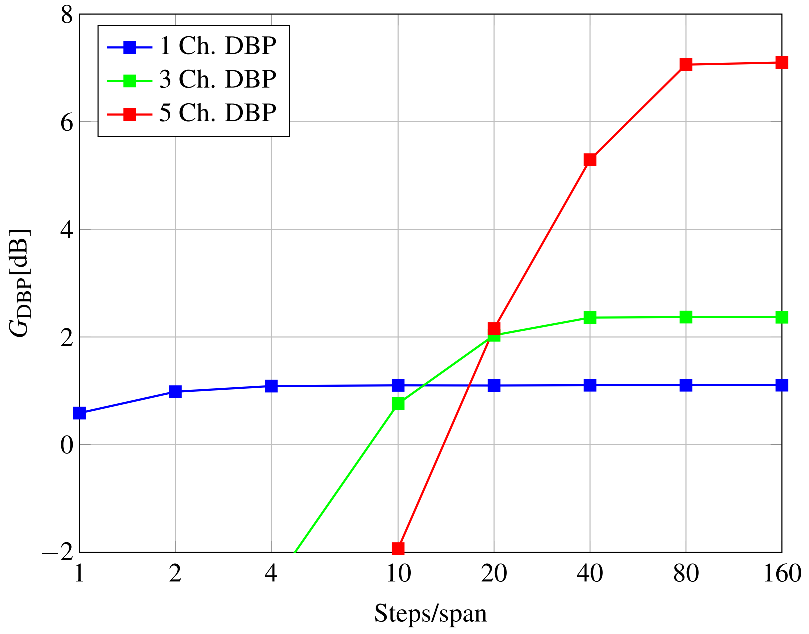
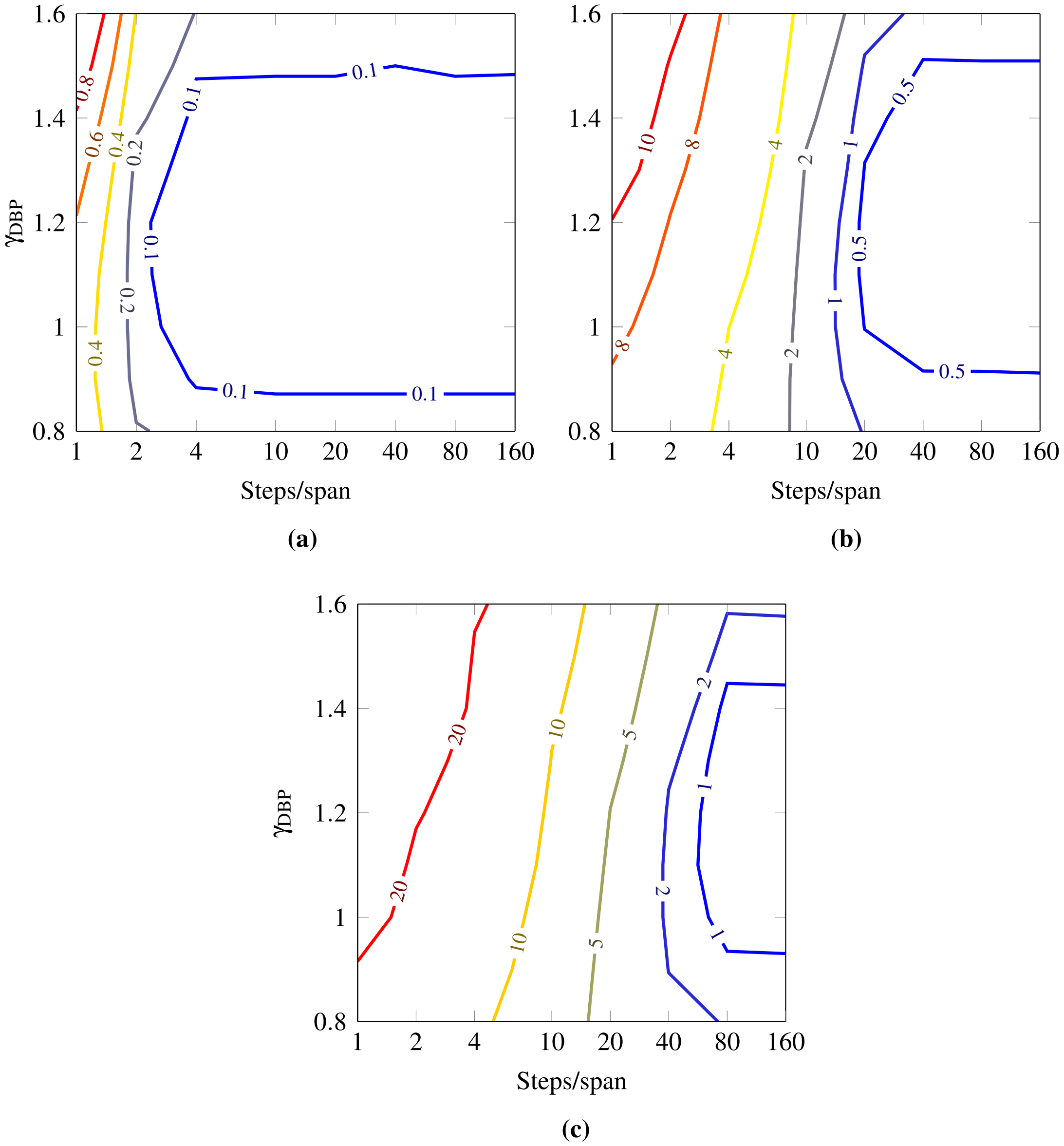


Figure 3.12: DBP SNR gain as a function of the number of steps per span used in the algorithm. The results are for a 5-channel system after transmission over 3200 km.

the ideal gain is 4, 40 and 80 for a single-channel, 3-channel, and 5-channel DBP, respectively. It can also be noticed that for a limited complexity (number of available iterations) increasing the DBP bandwidth is not always beneficial and instead it can be detrimental for the SNR performance. For instance, in this scenario, using full-field DBP starts to be beneficial (compared to lower NLC bandwidths) only when using more than 20 steps/span. Also, surprisingly, a penalty (negative gains in Fig. 3.12) can be incurred compared to an EDC receiver if a 3-channel or 5-channel NLC bandwidth is used with less than 4 or 20 steps/span, respectively. On the contrary single-channel DBP is more robust to a reduction in the number of steps/span, as going down to 1 single step per span reduces the SNR gain only by 0.5 dB. This can be attributed to the relatively small NLC bandwidth.

Usually, when DBP is used in a real transmission scenario, an additional issue is represented by the imperfect knowledge of the fibre parameters that are required to ideally *reverse* the fibre propagation. As discussed in section 3.1, these parameters are the fibre attenuation (α), the GVD coefficient (β_2) and the nonlinearity coefficient (γ). The knowledge of α is generally very accurate as it can be obtained through attenuation measurements and, for EDFA-amplified systems, the power profile is well-known to be exponentially decaying. However, in real systems, both the β_2 and γ parameters used in the DBP algorithm need to be properly tuned. We exclude the impact of the β_2 variation from the study as this parameter can be accurately tuned operating EDC in the linear regime. The impact of the variation of the γ parameter used in the DBP algorithm is presented in Fig. 3.13. Here, for each of the three NLC bandwidths mentioned above, a contour plot of the SNR performance degradation is shown as a function of both γ used in the DBP algorithm (referred to as γ_{DBP}) and the number of steps/span. The fibre nonlinear parameter γ used for these numerical results is the same of the reference system shown in table 3.1, i.e. $1.2\,W^{-1}\mathrm{km}^{-1}$. When single-channel DBP is used 3.13a the SNR penalty, compared to the ideal value, is limited to a maximum of 1 dB even for values of γ_{DBP} considerably off from the fibre γ parameter (e.g. $\gamma_{\mathrm{DBP}} = 1.6W^{-1}\mathrm{km}^{-}1$). For larger NLC bandwidths such as 3-channel DBP 3.13b and 5-channel DBP 3.13c, it can be observed how most of the SNR penalty is due to the reduction of the number of steps per span. However, when a suboptimal number of steps is utilised the performance can be improved using a γ_{DBP} which does not match the γ parameter of the fibre. Particularly, using a $\gamma_{\mathrm{DBP}} \leq \gamma$ results in a better performance in the region where the number of steps is low. This suggests that an *undercompensation* of the nonlinearity per step is beneficial when the step size is too coarse to accurately account for the interaction between dispersion and nonlinearity.



(a)

(b)

(c)

Figure 3.13: Contour plots of the SNR degradation after DBP as a function of number of algorithm iterations and the nonlinearity parameter γ_{DBP} for (a) single-channel DBP, (b) 3-channel DBP, and (c) 5-channel DBP.

3.6.2 DBP sampling rate

Another important DBP implementation parameter is the sampling rate at which the algorithm is operated. The complexity of the algorithm scales superlinearly with the number of samples per second N_s (complexity is dominated by the fast Fourier transform which scales as $O(N_s \log_2(N_s))$ such that its reduction is advisable. A lower bound on the sampling rate is given by the need to correctly reproduce the signal waveform without aliasing. While this is sufficient if we want to preserve the information of the analog signal in the digital domain when performing linear DSP, it may be insufficient if a nonlinear processing is performed. To explain this, let $x(t)$ be the analog signal to be backpropagated and $X_n \triangleq x(nT_s)$ its sampled version at time instants $t = nT_s$, where $F_s = \frac{1}{T_s}$ represents the sampling rate. We call BP$\{\cdot\}$ the operator performing backpropagation in the analog domain and $\tilde{\mathrm{BP}}\{\cdot\}$ the equivalent operator performed on sequences of samples. These 2 operators can be described as a repetition of a two-stage operation: (i) multiplication in the frequency domain by the inverse of dispersion frequency response; (ii) multiplication in the time domain by an instantaneous nonlinear phase shift. An ideal inversion of the NLSE performed in the analog domain would generate the sequence of samples Y_n:

$$Y_n = \int_{-\infty}^{\infty} \delta(t - nT_s)\big\{BP\{x(t)\}\big\}\, dt. \tag{3.27}$$

Since DBP is performed digitally, we have the sequence

$$\tilde{Y}_n = \tilde{\mathrm{BP}}\big\{X_n\big\} \tag{3.28}$$

In general $Y_n \neq \tilde{Y}_n$ because BP$\{\cdot\}$ is not linear and swapping the sampling operator with BP$\{\cdot\}$ results in a different output. Specifically, in both BP$\{\cdot\}$ and $\tilde{\mathrm{BP}}\{\cdot\}$, the multiplication of the signal by an instantaneous phase shift generates new frequency components that in Eq. (3.28) can introduce aliasing if the sampling rate is too low. Therefore Eq. (3.27) and Eq. (3.28) can be considered equivalent when F_s is sufficiently large. By setting F_s to a sufficiently higher value than the Nyquist frequency the digital spectrum will have a guard band between the replicas of the original spectrum (red spectra), thus allowing new DBP-generated out-of-band frequency components to be correctly represented.

The SNR gain versus the oversampling ratio (relative to the Nyquist rate) is plotted in Fig. 3.14. It is shown that for each backpropagated bandwidth there is a threshold sampling rate needed to obtain the optimal gain that corresponds to oversampling the backpropagated bandwidth by a factor of approximately 1.3. In our case, in order to observe the full advantage of the full-field backpropagation we need to use a sampling rate greater than 200 GSamples/s. Backpropagation of the entire spectrum of 5 channels sampled at the Nyquist rate (165 GSamples/s) results in a performance worse than backpropagating just 3 channels at the same sampling rate. Therefore, as shown in [29], backpropagating an oversampled smaller bandwidth can result in better performance compared to a larger backpropagated bandwidth sampled at Nyquist rate. These results provide a criterion for the choice of the bandwidth (or the number of channels) to be backpropagated when the sampling rate is a system constraint.

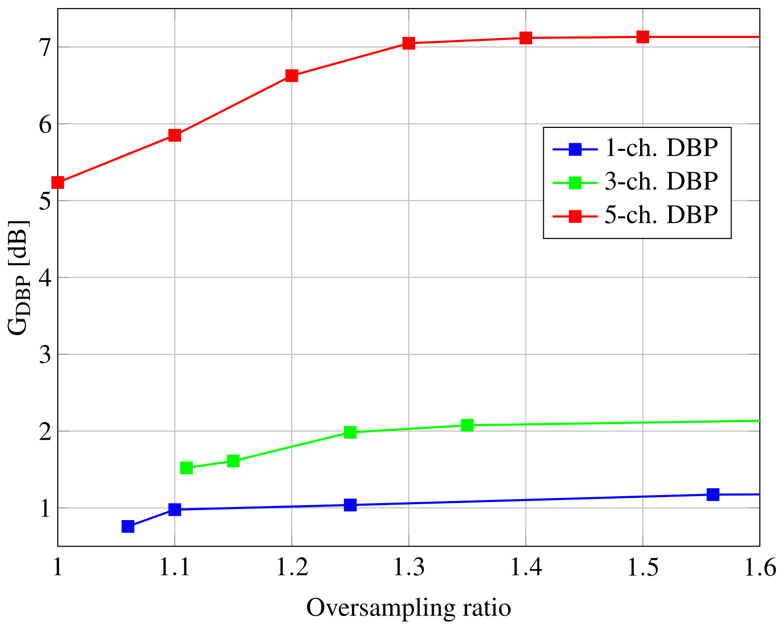


Figure 3.14: DBP SNR gain as a function of the sampling rate for different NLC bandwidths. The results are for a 5-channel system after transmission over 3200 km.

3.7 Impact of PMD

In addition to the effects of the choice of DSP parameters on the system performance as discussed above, another system parameter which has a significant effect on the performance of DBP is PMD [25, 29, 36].

As discussed in Sec 2.2 PMD is a linear phenomenon due to the inherent fibre birefringence. However, the interaction between the SOPs evolution and the fibre nonlinearity can also have a noticeable impact on the total NLI generated during optical fibre transmission. More importantly, it can have an even more significant impact on the performance of NLC schemes such as DBP.

This detrimental effect is due to the lack of knowledge of the polarisation evolution of the optical signal propagating through an optical fibre link and, as such, DBP can only perform an imperfect nonlinear cancellation. This concept is illustrated in Fig. 3.15. The signal launched into the optical fibre is here represented in the frequency domain. As seen in section 2.2, due to PMD the different frequency components of the transmitted signal encounter a different polarisation state evolution. The FWM product generated at a generic fibre section by any triple of frequency components f_1, f_2 and f_3 will be strongly dependent in both phase and magnitude on the specific polarisation state of each of these frequencies, as schematically illustrated in Fig. 3.15.

On the other hand, DBP uses as an input a *snapshot* of the (frequency dependent) polarisation states captured at the receiver section of the fibre. As illustrated in Fig. 3.16, the conventional DBP approach will then statically backpropagate such a snapshot, i.e. the polarisation states of the signal are not (to the first order) altered by the DBP

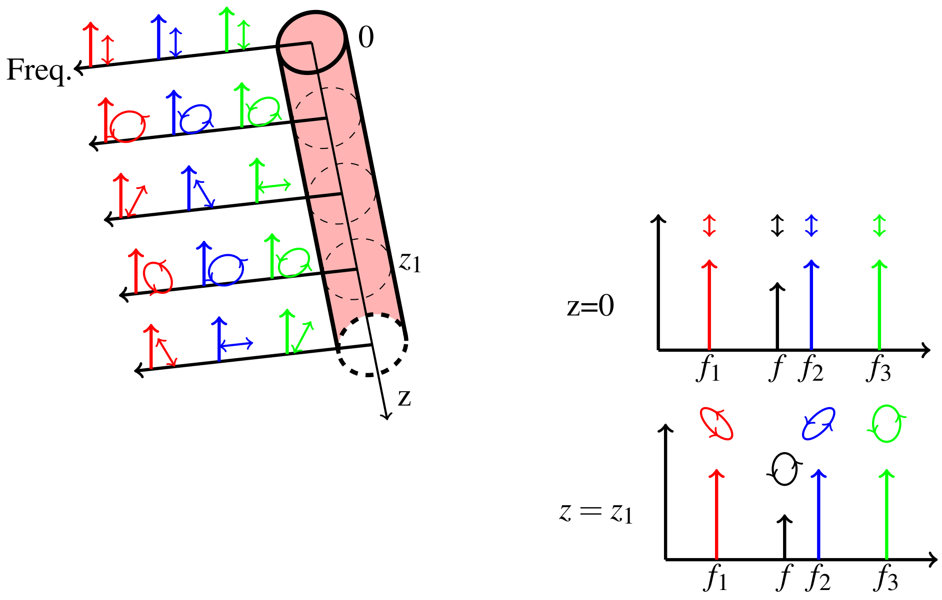


Figure 3.15: Schematic diagram of the SOPs evolution in a birefringent optical fibre and corresponding FWM generation.

algorithm. As a result, the FWM products generated in the backward direction by DBP do not match the ones in the forward propagation, effectively leaving the signal after DBP with a residual uncompensated NLI.

Various works have previously analysed the performance of DBP in the presence of PMD both numerically and (in a limited amount) analytically. In the following the results of our numerical study on the impact of PMD on DBP will be presented.

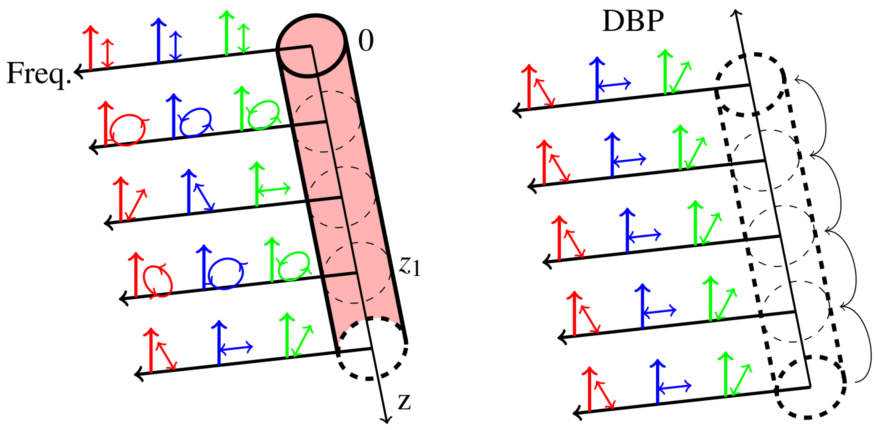


Figure 3.16: Illustration of the mismatch between forward and virtual backward (DBP) signal SOPs evolution.

3.7.1 Numerical evaluation of DBP performance in the presence of PMD

The parameters involved in the full characterisation of the performance of DBP in the presence of PMD are: the fibre PMD parameter, the average accumulated DGD, the transmitted bandwidth and the NLC bandwidth.

A comprehensive numerical study involving all these four parameters is prohibitive, as accurate Monte-Carlo simulations of PMD are in general computationally intense (see 2.6.3), particularly if DBP is also taken into account. In the following, results are shown aiming to capturing significant sections of the full parameters space.

As described in section 2.6.3, numerical results shown in this section are obtained through Monte-Carlo simulations of multiple PMD realisations using an hybrid method merging the log-step SSFM and the wave-plate approach. For each realisation the output SNR is measured and stored for subsequent statistical characterisation of the PMD effects on the performance of the system. DBP was implemented following the conventional algorithm described in 3.1, i.e. without any attempt of reversing the PMD evolution.

To characterise the impact of the fibre PMD parameter, in Fig. 3.17, the average DBP SNR gain is shown as a function of such parameter. The system considered in this case is a 5-channel WDM system over a transmission distance of 3200 km, using single-channel (blue curve), 3-channel (green curve) and 5-channel (red curve) DBP. The ideal DBP gains are shown by the black dashed lines as a reference. Ten

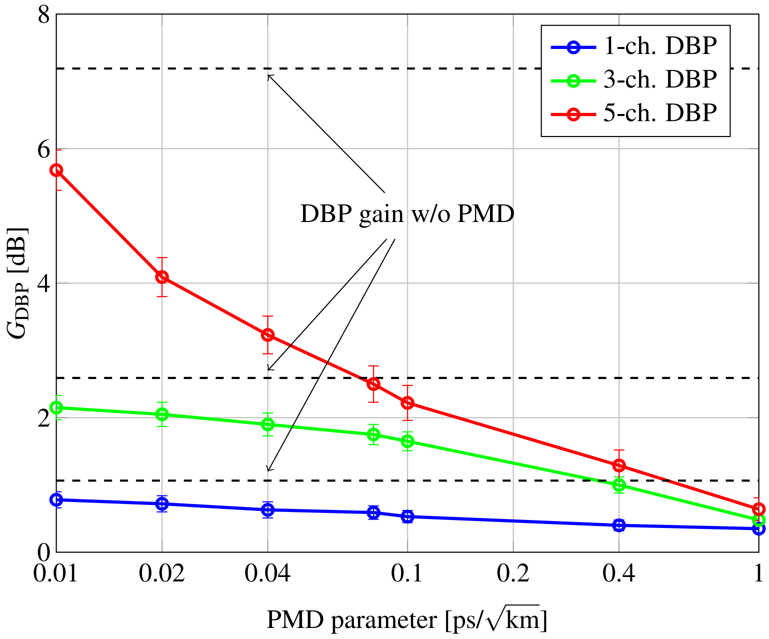
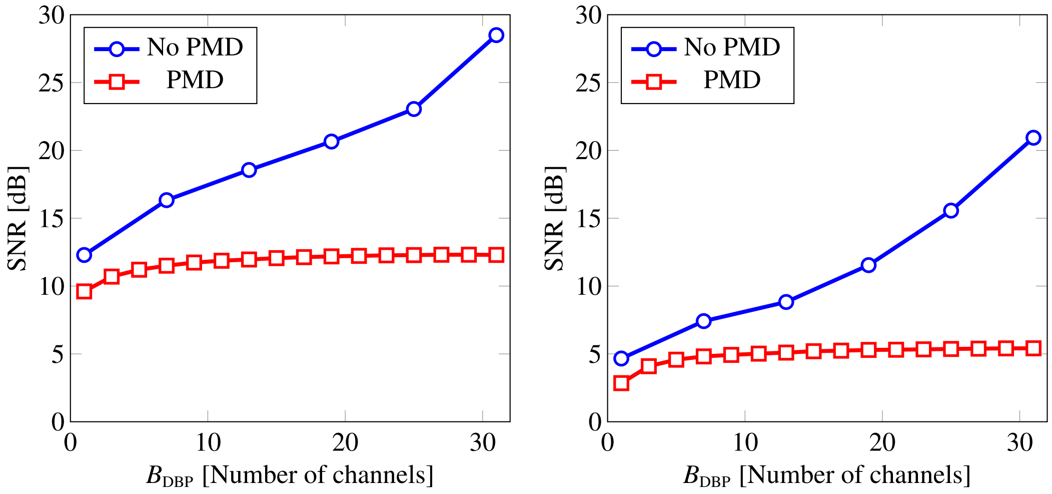


Figure 3.17: DBP SNR gain as a function of the fibre PMD parameter and different NLC bandwidths, for a 5-channel system and transmission over 3200 km.

independent PMD realisations were simulated and the error bars are shown in the plot to specify the standard deviation of the calculated SNR values.

The figure shows that the PMD parameter has a strong impact on the performance of DBP, especially as the NLC bandwidth is increased. Indeed, for large values of this parameter, (≥ 0.1 ps/$\sqrt{\text{km}}$) the advantage of using full-field DBP (red curve) compared to a 3-channel DBP almost entirely vanishes due to the effect of PMD. For modern fibres, more realistic values of the PMD parameter are in the range 0.02–0.1 ps/$\sqrt{\text{km}}$. However, in this range, still a significant reduction of the full-field DBP gain compared to the ideal gain can be observed. Conversely, single-channel DBP is less affected by the PMD parameter across the all range of investigated values, with a penalty of only 0.3 dB going from a PMD parameter of 0.01 ps/$\sqrt{\text{km}}$ to one of 1 ps/$\sqrt{\text{km}}$. In all NLC bandwidth cases, however, it can be noticed that the PMD phenomenon introduces an abrupt loss in SNR even at very low values of the PMD parameter. According to these results, PMD reduces the effectiveness of DBP particularly when it attempts to compensate NLI generated by frequencies far away from the channel of interest.

In order to confirm this behaviour and have a better understanding of the impact of PMD on the performance of wide-band DBP, additional results were obtained for a 31-channel WDM system and for a fixed PMD parameter of 0.1 ps/$\sqrt{\text{km}}$. In this case the transmitted power per channel was fixed to 5 dBm but the transmission distance was varied to study the role of the accumulated DGD. Fig. 3.18 shows (red curves) the average SNR after 50 PMD realisations as a function of the NLC bandwidth B_{DBP}, for 800 km distance (Fig. 3.18a) corresponding to an average accumulated DGD $\bar{\tau}$=2.83 ps, and 3200 km distance (Fig. 3.18b) corresponding to $\bar{\tau}$=5.66 ps. The blue curves



(a) **(b)**

Figure 3.18: Average SNR performance of multi-channel DBP as a function of B_{DBP} with and without PMD and for P=5 dBm. In (a) for an average DGD $\bar{\tau}$=2.83 ps (800 km distance), and in (b) for an average DGD $\bar{\tau}$=5.66 ps (3200 km distance).

represent instead the DBP gain in the absence of PMD. The results show that the presence of PMD induces a strong saturation of the multi-channel DBP performance when used in a wide-band (1 THz) transmission scenario. Such saturation behaviour appears to be independent on the transmission distance as, for instance, the penalty incurred by full-field DBP (B_{DBP}=31 channels) is equal to 15 dB in both cases.

A closer comparison between the SNR curves for the two transmission distances is shown in Fig. 3.19. Here the DBP SNR gain G_{DBP} compared to the EDC SNR (at a fixed transmitted power of 5 dBm) is shown as a function of B_{DBP} for the two average DGD cases of $\bar{\tau} = 2.83$ ps (800 km) and $\bar{\tau} = 5.66$ ps (3200 km). The plots show that, depending of the $\bar{\tau}$ value, G_{DBP} will behave differently. Specifically, a lower DGD value will result in a slower saturation of G_{DBP}. For $\bar{\tau} = 2.83$ ps, to achieve a gain 0.8 dB smaller from maximum gain the required NLC bandwidth is B_{DBP}=5 channels, as opposed to B_{DBP}=7 channels for $\bar{\tau} = 5.66$ ps. Also, a slight decrease in the maximum G_{DBP} (DBP gain for full-field DBP) can be noticed for larger values of the average DGD (0.15 dB lower gain for $\bar{\tau} = 5.66$ ps compared to $\bar{\tau} = 2.83$ ps). The results in Fig. 3.19 suggest that when the fibre PMD parameter is fixed, the average DGD has only a weak impact on the performance of DBP.

Although the average SNR measured over multiple realisations is an effective measure to summarise of the impact of PMD, it does not convey the whole picture. In most cases, a system designer is rather interested in the worst-case scenario performance or in the system performance that can be guaranteed with a certain probability of outage. It is therefore useful to complement the average SNR picture with an histogram of the SNR values obtained for the different PMD (or fibre) realisations. In Fig. 3.20, two

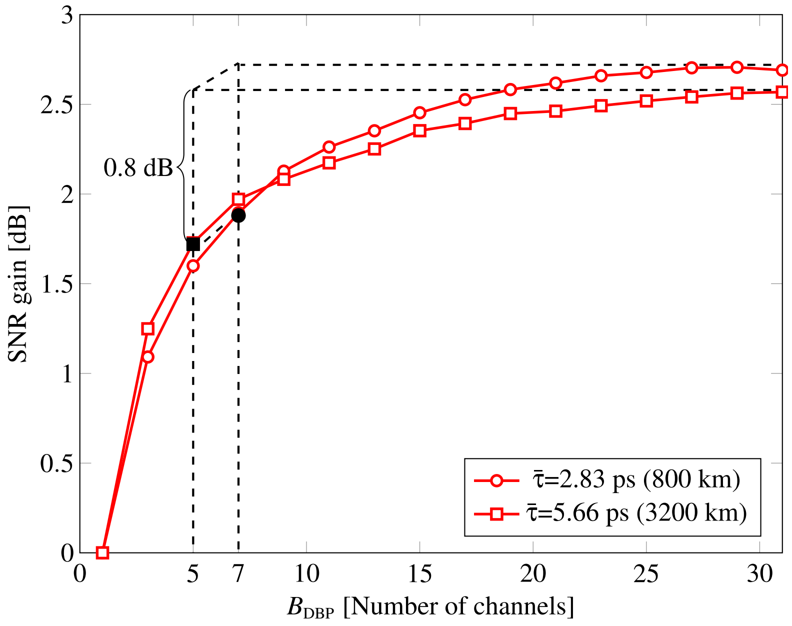


Figure 3.19: DBP SNR gain relative to B_{DBP}=1 as a function of B_{DBP} in the presence of PMD for P=5 dBm.

of such histograms are shown to show the SNR statistical behaviour for different fibre realisations and for two NLC bandwidths: single-channel DBP (blue) and full-field DBP (red). Interestingly, it is shown that the two histograms overlap. This does not mean that full-field DBP can potentially perform worse than the single-channel case as the fibre realisations for which the overlapping happens are different. However, statistically, operating DBP over a large NLC bandwidth could result, for a *bad* fibre, in a lower SNR than the single channel case operated with a *good* fibre. Also, the SNR values after full-field DBP is applied, are more statistically spread (larger standard deviation) compared to the single channel case. The SNR range of variation for the full-field DBP case spans from as little as 11 dB for a bad realisation to as much as 31 dB for a very good one.

This behaviour can be explained due to PMD inducing a larger decorrelation of the polarisation states of frequency components spaced far away from each other. This, in turn, translates into an ineffective mitigation of the NLI generated by these components of the signal spectrum.

In conclusion, it was shown that PMD is a major source of impairment of the performance of DBP, especially for wide-band scenarios where the NLC bandwidth needs to be extended to achieve the gain given by the full nonlinearity compensation. Although additional numerical results would help to gain more insight on the impact of PMD on DBP performance, due to the already mentioned computational complexity of accurate Monte-Carlo simulations, a properly validated analytical model would represent a much more powerful tool. Recent works [37, 39] have instead tackled the

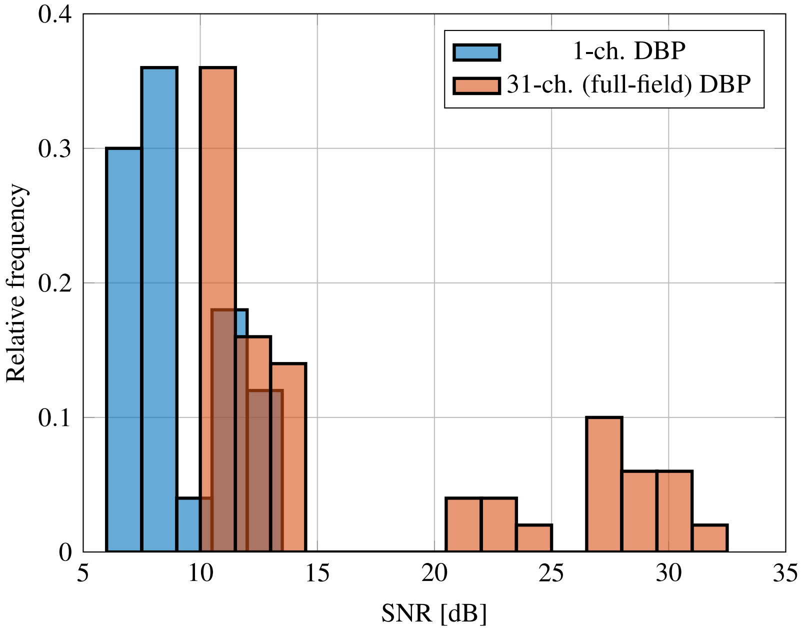


Figure 3.20: Histograms of SNR values obtained for B_{DBP}=1 and B_{DBP}=31 channels at 800 km transmission distance ($\bar{\tau}$=2.83 ps).

problem of mitigating the penalties due to PMD, through modified DBP schemes with some knowledge on the fibre PMD. This is discussed in more detail in chapter 6.

3.8 Conclusions

In this chapter, a comprehensive study of the performance of the DBP algorithm for optical fibre transmission systems using coherent detection was presented, with particular focus on its application over multiple channels to compensate for both intra-channel and inter-channel nonlinear impairments. Theoretical gains compared to an EDC receiver were analysed using an analytical closed-form expressions derived from the GN-model. This analysis was validated by means of numerical results obtained from simulations of optical fibre transmission based on the SSFM.

DBP showed substantial beneficial effects on the receiver SNR due to its ability to undo the signal-signal NLI, which is a major source of SNR deterioration in the region of interest for high-SE optical fibre systems. For instance, up to 10 dB SNR gain can be achieved for a 1000 km transmission of 5 channels, where full-field DBP was applied. DBP gain was found to be a decreasing function of both the overall transmitted bandwidth and transmission distance, due to the growing impact of signal-ASE NLI, which cannot be compensated. In particular, when full-field DBP is used, DBP gain depends weakly on the transmitted bandwidth ($\propto \operatorname{arcsinh}(B)^{-1/6}$), and a stronger dependence on the transmission distance given by $N_s^{1/2}$. In more pragmatic scenarios, where DBP is applied only over a portion of the transmitted bandwidth, the SNR gain was found to increase the quickest for the first compensated channels, where the gain slope is 1 dB every 3 dB of NLI suppression. Saturation of the SNR gain was observed as the signal-ASE NLI becomes dominant compared to the residual signal-signal NLI. Since the signal-ASE increases with transmission distance, the DBP gain was shown to saturate earlier for longer distances than for shorter ones. The SNR showed instead an abrupt increment as the NLC bandwidth approached the full-field bandwidth.

Ideal DBP implementation often represents an unrealistic task. Thus, limitations of the algorithm performance due to more realistic system scenarios, such as receiver DSP with limited computational complexity and PMD, were also analysed. As for the computational complexity, number of iterations (or steps) and sampling rate of operation play an important role in the performance of DBP, particularly as the NLC bandwidth was increased. For example, in a 5-channel transmission scenario, full-field DBP was shown to bring no improvement compared to lower NLC bandwidths (single channel or 3 channels), if operated with less than 20 steps/span. Unexpectedly, when less than 10 steps/span were used, NLC bandwidths in excess of a single channel are not only ineffective but also detrimental to DBP performance. This is due to the higher required spatial accuracy in the numerical integration performed by the SSFM, as the signal bandwidth increases.

Finally, PMD was identified as a major source of impairment of DBP performance even with infinite complexity available at the receiver. The results showed that typical fibre PMD prevents the compensation of inter-channel effects due to the lack of knowledge of the polarisation states evolution of the transmitted signal through the optical fibre. For a typical fibre PMD parameter of 0.1 ps/$\sqrt{\text{km}}$ and a 5-channel transmission, the difference between the DBP gain for a 3-channel DBP and 5-channel (full-field) DBP is smaller than 0.5 dB, whereas the DBP gain increased by more than 1 dB going from single-channel DBP and 3-channel DBP. This indicates a saturation of DBP gain as NLC bandwidth is increased. Such saturation was confirmed in a wideband transmission scenario of 1 THz optical bandwidth. In this case, DBP gain was observed to reach a value 0.8 dB below the *full-field* bandwidth maximum for an NLC bandwidth between 16% (5 channels) and 23% (7 channels) of the total bandwidth (31 channels), depending on the amount of accumulated DGD.

References

[1] J. Barry, E. Lee, and D. Messerschmitt, *Digital Communication.* Springer US, 2004.

[2] D. Mcghan, C. Laperle, A. Savchenko, C. Li, G. Mak, and M. O. Sullivan, "5120 km RZ-DPSK transmission over G652 fiber at 10 Gb/s with no optical dispersion compensation," in *Proc. Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA, 2005.

[3] K. Roberts, C. Li, L. Strawczynski, M. O'Sullivan, and I. Hardcastle, "Electronic precompensation of optical nonlinearity," *IEEE Photon. Technol. Lett.*, vol. 18, no. 2, pp. 403–405, 2006.

[4] R.-J. Essiambre and P. J. Winzer, "Fibre nonlinearities in electronically predistorted transmission ," in *Proc. European Conference on Optical Communication (ECOC)*, Glasgow, Scotland, UK, 2005.

[5] E. Yamazaki, F. Inuzuka, K. Yonenaga, A. Takada, and M. Koga, "Compensation of interchannel crosstalk induced by optical fiber nonlinearity in carrier phase-locked WDM system," *IEEE Photon. Technol. Lett.*, vol. 19, no. 1, pp. 9–11, 2007.

[6] A. J. Lowery, "Fiber nonlinearity pre- and post-compensation for long-haul optical links using OFDM." *Opt. Express*, vol. 15, no. 20, pp. 12 965–12 970, 2007.

[7] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, "Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing." *Opt. Express*, vol. 16, no. 2, pp. 880–888, 2008.

[8] E. Mateo, L. Zhu, and G. Li, "Impact of XPM and FWM on the digital implementation of impairment compensation for WDM transmission using backward propagation." *Opt. Express*, vol. 16, no. 20, pp. 16 124–16 137, 2008.

[9] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, 2008.

[10] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *J. Lightw. Technol.*, vol. 28, no. 6, pp. 939–951, 2010.

[11] E. F. Mateo, F. Yaman, and G. Li, "Efficient compensation of inter-channel nonlinear effects via digital backward propagation in WDM optical transmission." *Opt. Express*, vol. 18, no. 14, pp. 15 144–15 154, 2010.

[12] E. F. Mateo, X. Zhou, and G. Li, “Improved digital backward propagation for the compensation of inter-channel nonlinear effects in polarization-multiplexed WDM systems.” *Opt. Express*, vol. 19, no. 2, pp. 570–583, 2011.

[13] J. Leibrich and W. Rosenkranz, “Efficient numerical simulation of multichannel WDM transmission systems limited by XPM,” *IEEE Photon. Technol. Lett.*, vol. 15, no. 3, pp. 395–397, 2003.

[14] L. B. Du and A. J. Lowery, “Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems.” *Opt. Express*, vol. 18, no. 16, pp. 17 075–17 088, 2010.

[15] Y. Gao, J. H. Ke, K. P. Zhong, J. C. Cartledge, and S. S. H. Yam, “Intra-channel nonlinear compensation for 112 Gb/s dual polarization 16QAM systems,” *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3902–3910, 2012.

[16] Y. Gao, J. H. Ke, J. C. Cartledge, K. P. Zhong, and S. S.-H. Yam, “Implication of parameter values on low-pass filter assisted digital back propagation for DP 16-QAM,” *IEEE Photon. Technol. Lett.*, vol. 25, no. 10, pp. 917–920, 2013.

[17] M. Secondini, S. Rommel, F. Fresi, E. Forestieri, G. Meloni, and L. Poti, “Coherent 100G nonlinear compensation with single-step digital backpropagation,” in *International Conference on Optical Network Design and Modeling (ONDM)*, Florence, Italy, 2015.

[18] M. Secondini, S. Rommel, G. Meloni, F. Fresi, E. Forestieri, and L. Potì, “Single-step digital backpropagation for nonlinearity mitigation,” *Photonic. Netw. Commun.*, vol. 31, no. 3, pp. 1–10, 2015.

[19] E. Forestieri and M. Secondini, *Optical Communication Theory and Techniques*, 2004.

[20] M. Secondini and E. Forestieri, “Analytical Fiber-Optic Channel Model in the Presence of Cross-Phase Modulation,” *IEEE Photon. Technol. Lett.*, vol. 24, no. 22, pp. 2016–2019, 2012.

[21] M. Secondini, E. Forestieri, and G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, 2013.

[22] F. Yaman and G. Li, “Nonlinear impairment compensation for polarization-division multiplexed WDM transmission using digital backward propagation,” *IEEE Photon. Technol. Lett.*, vol. 2, no. 5, pp. 816–832, 2010.

[23] D. Rafique, J. Zhao, and A. D. Ellis, “Performance improvement by fibre non-linearity compensation in 112 Gb/s PM M-ary QAM,” in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, USA, 2011.

[24] ——, “Fundamental limitations of digital back-propagation in coherent transmission systems,” in *International Conference on Transparent Optical Networks*, Stockholm, Sweden, 2011.

[25] G. Gao, X. Chen, and W. Shieh, “Influence of PMD on fiber nonlinearity compensation using digital back propagation.” *Opt. Express*, vol. 20, no. 13, pp. 14 406–18, 2012.

[26] T. Tanimura, M. Nölle, J. K. Fischer, and C. Schubert, “Analytical results on back propagation nonlinear compensator with coherent detection.” *Opt. Express*, vol. 20, no. 27, pp. 28 779–85, 2012.

[27] L. Beygi, N. V. Irukulapati, E. Agrell, P. Johannisson, M. Karlsson, H. Wymeersch, P. Serena, and A. Bononi, “On nonlinearly-induced noise in single-channel optical links with digital backpropagation.” *Opt. Express*, vol. 21, no. 22, pp. 26 376–86, 2013.

[28] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, “The GN-model of fiber non-linear propagation and its applications,” *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, 2014.

[29] G. Liga, T. Xu, A. Alvarado, R. I. Killey, and P. Bayvel, “On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission,” *Opt. Express*, vol. 22, no. 24, pp. 30 053–30 062, 2014.

[30] P. Serena, “Nonlinear signal-noise interaction in optical links with nonlinear equalization,” *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1476–1483, 2016.

[31] R. Dar and P. J. Winzer, “On the Limits of Digital Back-Propagation in Fully Loaded WDM Systems,” *IEEE Photon. Technol. Lett.*, vol. 28, no. 11, pp. 1253–1256, 2016.

[32] D. Rafique, J. Zhao, and A. D. Ellis, “Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM M-ary QAM transmission.” *Opt. Express*, vol. 19, no. 6, pp. 5219–5224, 2011.

[33] X. Chen and W. Shieh, “Closed-form expressions for nonlinear transmission performance of densely spaced coherent optical OFDM systems.” *Opt. Express*, vol. 18, no. 18, pp. 19 039–54, 2010.

[34] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation." *Opt. Express*, vol. 22, no. 13, pp. 16 335–62, 2014.

[35] A. D. Ellis, S. T. Le, M. A. Z. Al-Khateeb, S. K. Turitsyn, G. Liga, D. Lavery, T. Xu, and P. Bayvel, "The impact of phase conjugation on the nonlinear-Shannon limit: The difference between optical and electrical phase conjugation," in *Summer Topicals Meeting Series (SUM)*, Nassau, Bahamas, 2015.

[36] K. Goroshko, H. Louchet, and A. Richter, "Fundamental limitations of digital back propagation due to polarization mode dispersion," in *Asia Communications and Photonics Conference (ACP)*, Hong Kong, 2015.

[37] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, "Polarization-mode dispersion aware digital backpropagation," in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[38] G. Liga, C. Czegledi, T. Xu, E. Agrell, R. I. Killey, and P. Bayvel, "Ultra-wideband nonlinearity compensation performance in the presence of PMD," in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[39] K. Goroshko, H. Louchet, and A. Richter, "Overcoming performance limitations of digital back propagation due to polarization mode dispersion," in *International Conference on Transparent Optical Networks*, Trento, Italy, 2016.

[40] E. F. Mateo, X. Zhou, and G. Li, "Selective post-compensation of nonlinear impairments in polarization-division multiplexed WDM systems with different channel granularities," *IEEE Journal of Quantum Electronics*, vol. 47, no. 1, pp. 109–116, 2011.

[41] J.-X. Cai, H. G. Batshon, H. Zhang, M. Mazurczyk, O. Sinkin, and D. G. Foursa, "200 Gb/s and dual wavelength 400 Gb/s transmission over transpacific distance at 6.0 b/s/Hz spectral efficiency," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 832–839, 2014.

[42] L. Galdino, G. Liga, D. Lavery, R. Maher, T. Xu, M. Sato, R. I. Killey, S. J. Savory, B. C. Thomsen, and P. Bayvel, "Unrepeatered transmission over 253.4 km ultra low loss fibre achieving 6.95 (b/s)/Hz SE using EDFA-only pre-amplifier," in *Proc. European Conference on Optical Communication (ECOC)*, Cannes, France, 2014.

[43] R. Maher, L. Galdino, M. Sato, T. Xu, K. Shi, S. Kilmurray, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Linear and nonlinear impairment mitigation in a Nyquist spaced DP-16QAM WDM transmission system with full-field DBP," in *Proc. European Conference on Optical Communication (ECOC)*, Cannes, France, 2014.

[44] R. Maher, T. Xu, L. Galdino, M. Sato, A. Alvarado, K. Shi, S. J. Savory, B. C. Thomsen, R. I. Killey, and P. Bayvel, "Spectrally shaped DP-16QAM super-channel transmission with multi-channel digital back-propagation." *Scientific reports*, vol. 5, pp. 1–8, 2015.

[45] R. Maher, D. Millar, A. Alvarado, K. Parsons, R. Killey, and P. Bayvel, "Reach enhancement of 100 % for a DP-64QAM super-channel using MC-DBP," in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, USA, 2015.

[46] R. Maher, D. Lavery, A. Alvarado, M. Paskov, and P. Bayvel, "Multi-channel DBP for reach enhancement of high capacity M-QAM super-channels," in *Signal Processing in Photonic Communications (SPPComm)*, Boston, MA, USA, 2015.

[47] F. P. Guiomar, S. B. Amado, R. M. Ferreira, J. D. Reis, S. M. Rossi, A. Chiuchiarelli, J. R. F. De Oliveira, A. L. Teixeira, and A. N. Pinto, "Multicarrier digital backpropagation for 400G optical superchannels," *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1896–1907, 2016.

[48] R. Asif, C.-Y. Lin, M. Holtmannspoetter, and B. Schmauss, "Optimized digital backward propagation for phase modulated signals in mixed-optical fiber transmission link." *Opt. Express*, vol. 18, no. 22, pp. 22 796–22 807, 2010.

[49] N. K. Fontaine, G. Raybon, B. Guan, A. Adamiecki, P. J. Winzer, R. Ryf, A. Konczykowska, F. Jorge, J.-Y. Dupuy, L. L. Buhl, S. Chandrasekhar, R. Delbue, P. Pupalaikis, and A. Sureka, "228-GHz coherent receiver using digital optical bandwidth interleaving and reception of 214-GBd (856-Gb/s) PDM-QPSK," in *Proc. European Conference on Optical Communication (ECOC)*, Amsterdam, The Netherlands, 2012.

4

Optimum detection for the nonlinear optical fibre channel

The properties of the optical fibre channel described in chapter 2 are clearly very different from the classical additive white Gaussian noise (AWGN) channel ones well studied in communication theory. The main difference lies in the fact that the channel properties change considerably based on the signal that is transmitted through it. Also, effects such as chromatic dispersion (CD) and its interaction with nonlinear propagation effects introduce memory in channel which in most cases cannot be entirely removed by equalisation schemes applied at the receiver, such as electronic dispersion compensation (EDC).

The properties of the channel are crucial to devise a proper signal detection scheme at its output. By *detection* we can broadly define the strategy put in place to *guess* which signal was transmitted in a set of M signals, when an observation is given at the output of a noisy channel. Although the optical fibre channel shows features that are quite divergent from the AWGN channel, the most typical detection scheme used in current optical fibre systems is still the conventional matched filter (MF)/sampler approach followed by symbol-by-symbol decision. This approach, which is well-known to be the optimum for the AWGN case [1], is not in general the best for channels with memory, and particularly for nonlinear channels.

In the context of coherent detection systems, only a relatively small number of works have so far investigated the problem of optimising the detection process for the optical fibre channel. The explanation for this can be found in three main arguments: i) historically, coherent optical systems have always been operated in a linear (or pseudo-linear) region of transmitted powers which results in the MF approach being the optimal one; ii) fibre transmission impairments such as CD or Kerr nonlinearity are in general considered to be detrimental and therefore the most widespread approach consists in attempting to undo them using *zero-forcing* equalisation scheme such as digital backpropagation (DBP). By doing that, the assumption is to recover a perfectly linear channel whose detection is thus optimal when using the MF strategy; iii) a comprehensive and satisfactory knowledge of the statistical properties of the optical fibre channel, required for the design of a tailored detection scheme, is still missing.

As a matter of fact, sophisticated detection schemes were already adopted in pre-coherent era, when techniques such as maximum-likelihood sequence detection (MLSD) were used to mitigate the effect of CD in intensity-modulation direct-detection (IMDD) systems. For many years MLSD has represented the reference for such systems, as the compensation of CD through DSP was not a viable option due to the loss of phase information caused by squared-law detectors [2–4]. When coherent detection was introduced in the mid-2000s, the MF and symbol-by-symbol approach was never really questioned until [5], where MLSD was utilised for the first time to mitigate nonlinear fibre impairments. In this work, approximated statistical models were proposed to account for the channel memory with or without prior application of DBP. Based on this model branch metrics for the Viterbi algorithm [6] were derived.

An alternative detection method called *stochastic backpropagation* [7, 8] proposed to *reconstruct* the channel statistical law based on the received channel samples. In order to do that, the received noisy samples were backpropagated and the optimal detection rule was derived through the use of message-passing algorithms.

The previous methods both tackled the problem of detecting in a multi-span system environment where signal-ASE interaction plays a significant role and needs to be factored in for the design of the optimal decision strategy.

In [9], an MLSD strategy was instead applied to the unrepeatered (single-span) fibre channel where the signal-ASE term is absent or at least negligible. As discussed in section 2.3.1, unrepeatered systems are used for a wide variety of applications but their performance is strongly affected by fibre nonlinear distortions. Thus, increasing transmission rates and extending their reach represents a crucial and challenging research topic. More importantly, devising optimal detection strategies for more complex systems, such as multi-span fibre systems, requires a clear statement of the problem, which is easier to do for a simplified version thereof. In this sense, single-span systems represent the main building blocks of multi-span systems, and solving the detection problem in such a scenario represents the first necessary step to tackle the more sophisticated detection problem regarding multi-span systems.

The main aim of the study in [9] was showing that for a single-span system, properly designed MLSD strategies can be claimed to be optimum for that specific channel. Furthermore, it was shown that performance similar to the AWGN channel can be arbitrarily approached if the complexity constraint is removed. This chapter is based on the work performed on this topic.

4.1 Properties of the nonlinear single-span optical fibre channel

The interaction of CD and nonlinearity during the propagation of lightwave pulses in an optical fibre generates a nonlinear distortion with memory [10]. This means that the received waveform associated to a symbol transmitted at a given time slot is affected by the transmitted symbols surrounding that time slot [1]. Such property is mainly to be attributed to the CD phenomenon. Indeed, in section 2.1.2 CD was observed to cause a spread of the transmitted pulse over multiple time slots. This pulse spreading destroys the pulse orthogonality when multiple of them are transmitted next to each other. Let us assume that the transmitted signal is given by

$$s(t) = \sum_{k=-\infty}^{\infty} x_k p(t - kT_s) \tag{4.1}$$

where x_k is the transmitted symbol at time-slot k, $p(t)$ is the modulation pulse and T_s is the symbol period. Then the (noiseless) signal affected by CD can be written as

$$y(t) = \sum_{k=-\infty}^{\infty} x_k q(t - kT_s) \tag{4.2}$$

where

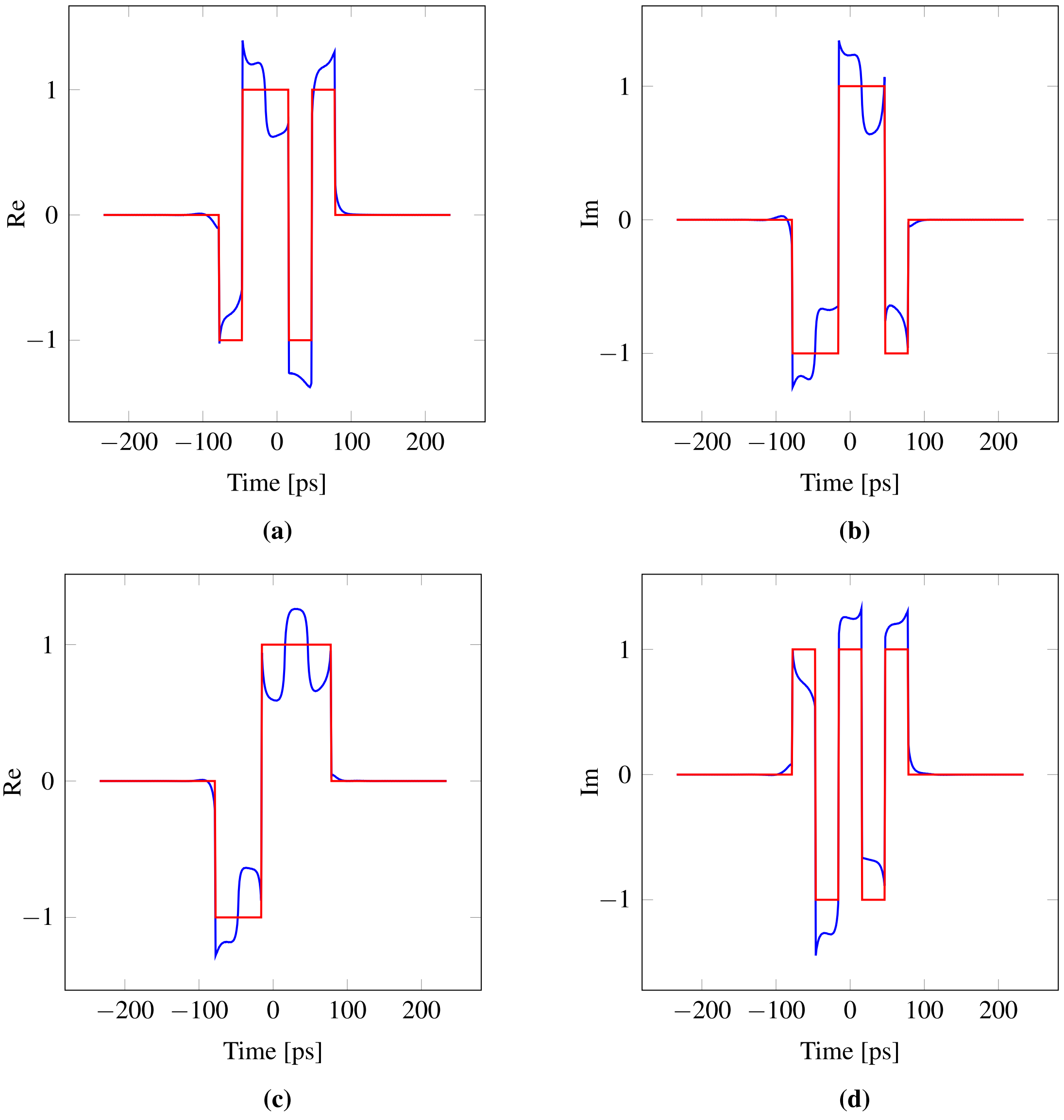
$$q(t) = \int_{-\infty}^{\infty} p(\tau)h(t-\tau)d\tau \tag{4.3}$$

and $h(t)$ is the impulse response associated with CD. The effect of the pulse $q(t)$ on other pulses on time slots $k \neq 0$ is typically referred to as (linear) inter-symbol interference (ISI). However, because of the interaction between CD and nonlinear propagation effects, the received signal $y(t)$ cannot be in general written in the form of Eq. (4.2), i.e. the ISI acting on one pulse cannot be expressed as superposition of pulses located at different time-slots. An illustration of this is given in Fig. 4.1 that shows the impact of the interplay between CD and nonlinear effects on the transmitted waveforms. Two QPSK 5-symbol sequences modulated with rectangular pulses at a symbol rate of 32

---

[1] Symbols transmitted at future time slots can also have an impact on the output at the present time slot. This is due to the delayed time frames between input and output typically adopted when analysing NLSE or Manakov equation.

GBaud are shown at an instantaneous power of 10 dBm and after propagation over a 300 km SSMF span and full CD compensation. The transmitted sequences shown are the waveforms associated with the QPSK symbols $(-1-j, 1-j, 1+j, -1+j, 1-j)$ and $(-1+j, -1-j, 1+j, 1-j, 1+j)$ for Figs. 4.1a, 4.1b, and 4.1c, 4.1d, respectively. It can be observed that: i) each received pulse is temporally confined but a unique received pulse shape cannot be identified; ii) a comparison of the middle pulses between the two sequences (sharing the same middle symbol) shows that the received pulse shape depends not only on the transmitted pulse but also on the surrounding ones; iii) the imaginary part of the received pulse shape depends also on what is transmitted on the



(a) **(b)**

(c) **(d)**

Figure 4.1: Transmitted (red) and received (blue) waveforms after a 300 km fibre span, for two 5-symbol QPSK sequences. In (a) and (b), the waveforms corresponding to the symbol sequence 13203 (in decimal notation and using Gray mapping) are shown in their real and imaginary part, respectively. In (c) and (d), the waveforms corresponding to the symbol sequence 01232 are shown in their real and imaginary part, respectively.

real part (and vice versa), due to the nonlinear effects (see e.g. Figs. 4.1c, 4.1d). This interaction between pulses is referred to as *nonlinear ISI* [5, 11, 12] and it represents one of the main properties of the optical fibre channel.

Since the memory is introduced by the fibre CD, it is limited (see discussion in [13]) and it varies depending on the channel bandwidth and transmission distance (see section 2.1.2). The channel memory also depends on the specific transmission scenario as different configurations of an optical fibre transmission system can show a significantly different behaviour.

In this chapter, as a preliminary study, the focus was put on the single-span fibre channel illustrated in Fig. 4.2a. It can be noticed that the booster amplifier is here discarded compared to the conventional single-span channel shown in Fig. 2.5. This simplification is based on the realistic assumption that the OSNR at the fibre input is very high. On the other hand, this allows to simplify the detection problem as the signal-ASE NLI is entirely removed and the only source of impairment is given by the signal-signal NLI. As the signal-ASE NLI statistical distribution is not known (at least not analytically), devising an optimal receiver represents a much harder problem.

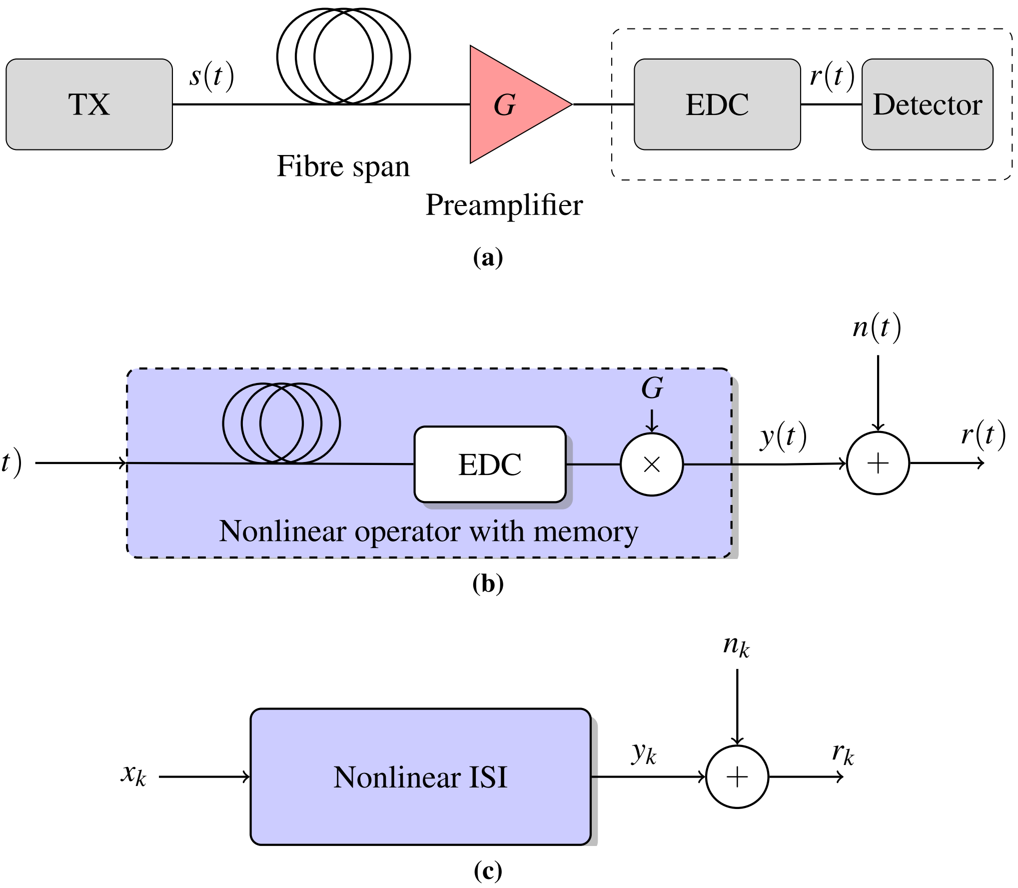


Figure 4.2: Schematic diagrams of a single-span optical fibre channel where (a) represents the physical channel, (b) represents the equivalent continuous-time model, and (c) is the equivalent discrete-time model.

As it can be seen from Fig. 4.2a, at the receiver the first block is assumed to be an EDC filter followed by a detector. The EDC filter cannot entirely remove the channel memory because of the interplay between CD and nonlinear effects. However, it can significantly reduce it (the memory of zero-th order solution is undone). Furthermore, since the EDC block is a linear all-pass filter, it has no impact on the properties of the Gaussian noise added by the EDFA amplifier. Because of this reason, the EDC block has no effect on the optimality of the detection process at the channel output, as a sufficient statistic can be equally collected at the block output. As a result, it can be considered as a fixed block in the receiver and hence absorbed as part of the channel itself. The physical channel in Fig. 4.2a can be equivalently represented by the block diagram in Fig. 4.2b. The channel is here modelled as the cascade of the optical fibre nonlinear operator and the EDC block. As mentioned before, the linearity of the EDC block allows switching its position before the AWGN addition without changing the properties of the channel. A scaling block G implements the amplification gain before the AWGN process $n(t)$ is added.

A possible quantification of the memory of the continuous-time channel in 4.2b can be given through the calculation of the auto-correlation function on the noiseless output signal $y(t)$ [14]. However, from the receiver standpoint, the relevant quantity is the memory shown by the *channel observations*. Such observations are a discrete set r_k of samples that the receiver extracts from the received waveform $r(t)$ in order to estimate the transmitted data x_k. Thus, a discrete-time channel model can be then derived and is illustrated in Fig. 4.2c. Although the optimal way to extract such set of samples will be discussed in the next section, a possible option is represented by the combination of a MF and sampler (1 Sa/sym). Following this approach, Fig. 4.3 shows the memory effect on the received samples r_k. The three scatter plots in Fig. 4.3 are obtained by transmitting all possible sequences of 3 (Fig. 4.3a), 5 (Fig. 4.3b) and 7 (Fig. 4.3c) QPSK symbols through a 350 km fibre span at a transmitted power of

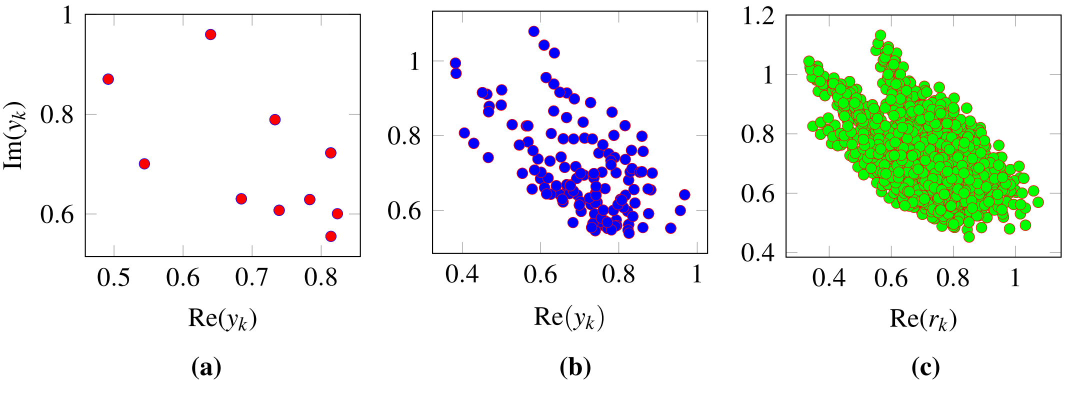


Figure 4.3: Nonlinear ISI as observed at the output of a MF. The ISI is observed on a fixed central symbol and varying all possible sequences of (a) 3, (b) 5, and (c) 7 symbols.

15 dBm. The points in the diagram show the nonlinear ISI on the (noiseless) middle sample y_k as the surrounding symbols are varied. The effect of the nonlinear ISI consists in a movement of the constellation point onto different locations, depending on the surrounding symbols. Furthermore, it can be seen that adding more surrounding symbols increases the number of ISI *locations* due to the additional memory effect coming from the outermost symbols. Such symbols affect the ISI on the middle symbol more weakly the further away in time from this symbol they are located. This can be understood by looking at the *filling* effect in Figs. 4.3a, 4.3b, and 4.3c as the number of the symbols in the sequence is increased. This vanishing effect is a sign of the finite memory of the channel which is quantified in Fig. 4.4.

Fig. 4.4 shows the normalised variance of the NL ISI illustrated in Fig. 4.3 as a function of the transmitted power and the number of symbols surrounding the central one. The normalised NLI variance increases in all cases with a rate of approximately 2 dB/dB as predicted by the models discussed in chapter 2. A gap in the NLI variance between the 3-symbol transmission and the 5-symbol (or 7-symbol) transmission of 0.45 dB can be observed. However, no significant difference in the variance can be observed between the 5-symbol and 7-symbol case. This indicates a small influence in the NLI variance of the symbols located three time-slots away from the symbol of interest. In this sense, the memory of the channel can be quantified in 5 symbols. However, it is worthwhile mentioning that such measure of the channel memory, although providing an intuitive understanding of the channel properties, is not necessarily an indication on how much memory a well-performing receiver should account for. Rather, the performance of receivers operating on channels in which the noise is AWGN, such as the one in

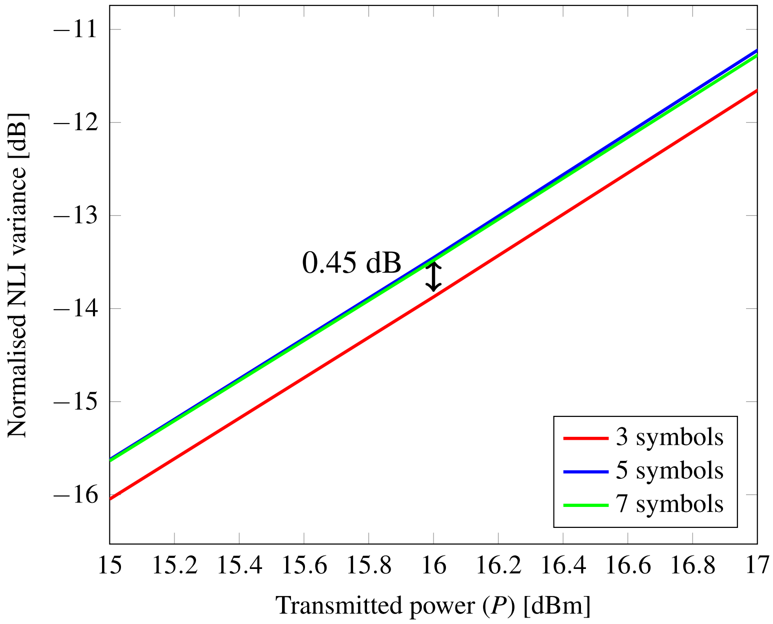


Figure 4.4: Normalised NLI variance over the central symbol as a function of power and varying number of transmitted symbols.

Fig. 4.2b, is determined by the minimum Euclidean distance (in the asympotically high SNR regime) between the interference points shown in Fig.4.3. This quantity is not necessarily related to the NLI variance in Fig. 4.4.

4.2 Improving detection for the optical fibre channel

A precise definition of the concept of *memory* for discrete-time channels can be found in communication and information theory. Let $(X_1, X_2, \ldots, X_n)$ and $(R_1, R_2, \ldots, R_n)$ be transmitted and received sequences in a channel, and $(x_1, x_2, \ldots, x_n)$, $(r_1, r_2, \ldots, r_n)$ their respective realisations. A discrete-time channel is said to have *memory* if, the received sample at present time r_n is statistically dependent on the present but also on a certain number of past inputs $(x_{n-1}, x_{n-2}, \ldots, x_{n-2p})$ and outputs $(r_{n-1}, r_{n-2}, \ldots, r_{n-2q})$ [15]. This means that the full description of the channel is given by the conditional pdf[2]

$$p(r_n|x_n, x_{n-1}, \ldots, x_{n-2p}, r_n, r_{n-1}, \ldots, r_{n-2q}) \qquad \forall n \geq 2p \geq 2, \tag{4.4}$$

and the *channel memory* can be defined as

$$2m \triangleq \max(2p, 2q). \tag{4.5}$$

As a result, a channel can be considered *memoryless* if and only if

$$p(r_n, r_{n-1}, \ldots, r_1|x_n, x_{n-1}, \ldots, x_1) = \prod_{k=1}^{n} p(r_k|x_k). \tag{4.6}$$

Under the above definition, the family of channels with memory includes several kind of specific channels [3]. Particularly, for the channel discussed in the previous section and illustrated in Fig.4.2c the present output r_n statistically depends only on the present and past *inputs*, but not on the past *outputs*. Therefore such channels are fully described by

$$p(r_n|x_n, x_{n-1}, \ldots, x_{n-2p}) \qquad \forall n \geq 2p \geq 2, \tag{4.7}$$

and are referred to as ISI channels.

The question addressed in this section is: *what is the optimal detection strategy for an ISI channel?* We refer to optimum detection strategy as the rule that minimises the probability of wrongly estimating a generic parameter of interest θ, which for a communication channel is either a transmitted symbol or a sequence of transmitted symbols. Such strategy can be devised following two steps [17, Ch. 7], [1, Ch. 4]:

---

[2] For simplicity of notation the indication of the random variable is dropped in the pdf subscript.

[3] In turn, channels with finite memory can be considered as a subset of a larger family of channels called *finite-state channels* [16].

i. extracting a *sufficient statistic* $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ on the parameter θ from the observation of the noisy received waveform $r(t)$ (see Fig. 4.2b)

ii. implementing the rule that maximises the a-posteriori probability

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|\mathbf{r}) \tag{4.8}$$

hence called maximum a-posteriori probability (MAP) rule.

A sufficient statistic on the parameter θ and for the observation of $r(t)$ can be defined as a set of samples $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ such that for any statistics $\mathbf{r}' = (r'_1, r'_2, \ldots, r'_n)$ extracted from $r(t)$ it is verified [18], [19], [1, Ch. 4]

$$p(\mathbf{r}'|\mathbf{r}, \theta) = p(\mathbf{r}'|\mathbf{r}) \tag{4.9}$$

i.e. when knowing the sufficient statistics $\mathbf{r}$, any other statistics $\mathbf{r}'$ is independent on the parameter to estimate. This is equivalent to have

$$p(\mathbf{r}'|\theta) = f(\mathbf{r}')q(\mathbf{r}|\theta) \tag{4.10}$$

which is known as Fisher-Neyman factorisation [18] criterion and indicates that $\mathbf{r}$ contains all the information that is needed to perform an estimation on θ.

Because of the memory in Eq. (4.7), it is clear that a sufficient statistic to estimate, e.g., the symbol x_k at time-instant k has to be collected over different time-slots depending on the memory of the channel. Also in order for Eq. (4.9) to be verified, a necessary condition is that the samples y_k of the noiseless waveform $y(t)$ (see Fig.4.2b) need to be a complete representation of $y(t)$ (over the observation time), i.e. y_k and $y(t)$ need to be in a one-to-one relationship. For instance this can be obtained by projecting $r(t)$ onto an orthonormal basis of signals [1]. For our specific channel, such a method is also sufficient to obtain a sufficient statistic $\mathbf{r}$ because $n(t)$ is assumed AWGN [1, 17].

Let us now assume that $\mathbf{r} = (r_1^1, r_1^2, \ldots, r_1^N, \ldots, r_n^1, \ldots, r_k^N)$[4] are samples obtained by the projection of $r(t)$ over a complete (for $y(t)$) orthonormal set of N signals and over an observation time long enough to include the channel memory. Such samples can be considered sufficient for the estimation of a single x_k but also for a *sequence* of transmitted symbols $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Indeed, for n large enough, the observations $\mathbf{r}$ can be considered (approximately) statistically independent from transmitted symbols outside of the sequence $\mathbf{x}$ and, as a result, detection can be performed optimally by applying the rule

---

[4] Superscripts indicate the signal index within the orthonormal basis whereas subscripts indicate the discrete-time index.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{x}|\mathbf{r}). \tag{4.11}$$

In most cases, sequences of transmitted symbols are all equally likely and the rule in Eq. (4.11) can be rewritten as [20, Ch. 5]

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, p(\mathbf{r}|\mathbf{x}) \tag{4.12}$$

which is referred to as MLSD and will be used in the next sections.

It is worth mentioning that the MLSD strategy is optimal in the minimum error probability on the estimated parameter which is in this case the sequence $\mathbf{x}$. Optimum strategies in the minimum symbol error probability sense are also possible based on the criteria discussed earlier in this section. However, as is known (e.g. [21]), optimal symbol-by-symbol receiver for ISI channels are considerably more complex than MLSD receivers and their performance is in general very similar. For this reason, in the following only sequence detection will be discussed.

4.3 Bank-of-correlators receiver

The most natural way to apply the criteria discussed in the above section to a single-span fibre channel is detecting different sequences by applying the minimum Euclidean distance criterion on the waveform channel in 4.2b. Indeed, for times $T = nT_s$ that are much longer than channel memory, the waveform $r(t)$ can be written as

$$r(t) = \sum_{k=-\infty}^{\infty} s(t-kT,\mathbf{x}) + n(t) \tag{4.13}$$

where $s(t-kT,\mathbf{x})$ is a set of waveforms with support in $kT \le t \le (k+1)T$, associated with the sequence of transmitted symbols $\mathbf{x} \in \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{M^{2m+1}}\}$, and $n(t)$ is a white complex Gaussian process with power spectral density equal to $2N_o$. Eq. (4.13) assumes that each waveform $s(t-kT,\mathbf{x})$ is determined only by the transmitted sequence of symbols $\mathbf{x}$ within a time T. Thus any *inter-sequence* interaction is assumed to have a negligible effect on the shape of the entire sequence waveform. For long enough sequences (large T), this approximation can be considered true. Detecting by sequences allows then to avoid the nonlinear ISI observed in Fig.4.3, which is due to the fact that the actual waveform associated with a given transmitted symbol varies depending on the channel state.

The minimum Euclidean distance criterion over all possible waveforms $s(t-kT,\mathbf{x})$ is well-known to represent the optimal detection strategy for the channel in 4.2b, when $n(t)$ is AWGN [1]. Such a strategy effectively consists in maximising the log-likelihood [22, Sec. 2.6]

$$\Lambda_k = \frac{2}{N_o} \mathrm{Re} \int r(t) s^*(t, \mathbf{x}_k) dt - \frac{1}{N_o} \int |s(t, \mathbf{x}_k)|^2 dt. \tag{4.14}$$

A receiver implementing (in continuous-time) such a detection strategy is referred to as a *bank-of-correlators* (BC) receiver and is illustrated in Fig. 4.5. The reason for this name is due to the set of parallel correlators calculating the scalar products (in the complex signal space)

$$l_k = \int_{-\infty}^{\infty} \mathrm{Re}\left[r(t) s^*(t - T, \mathbf{x}_k)\right] dt \tag{4.15}$$

between the received signal $r(t)$ and the waveforms $s(t-T, \mathbf{x}_k)$ for $k = 1, ..., M^{2m+1}$ where $\mathbf{x}_k$ represent a specific element in the ensemble of all possible transmitted sequences of length n. The signal correlations are then offset by a quantity that is proportional to the signal energy to account for the case where different signals $s(t-T, \mathbf{x}_k)$ have unequal energies.

The number of required correlators is equal to the number of possible waveforms which is in turn equal to the number of possible transmitted sequences M^{2m+1}. Although the correlation calculations can be performed in parallel, still $M^{2m+1}-1$ comparisons are required in order to find the most likely symbol sequence. This imposes a stringent constraint on the length of the sequences if a manageable receiver complexity is to be

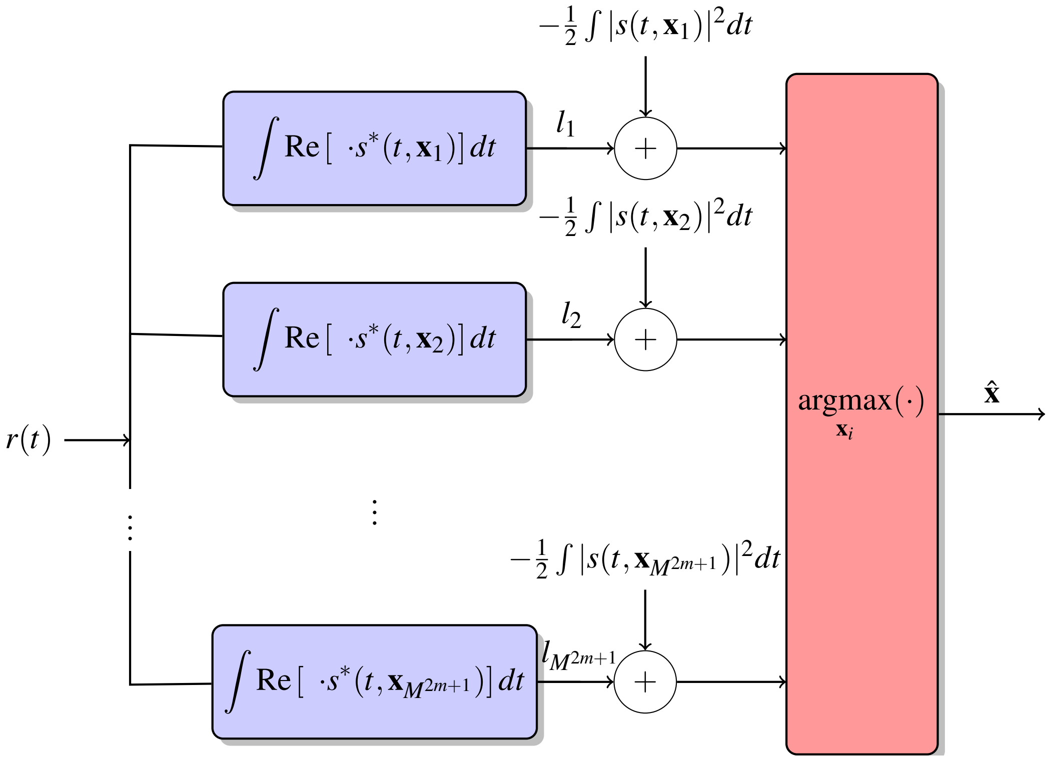


Figure 4.5: Schematic diagram of a BC receiver.

preserved.

The performance of the BC receiver were numerically analysed in Fig. 4.6 where the BER after detection is shown for different correlator lengths $2m+1$. The system parameters used for this numerical study are shown in table 4.1 As a reference, the BER of a QPSK transmission in an AWGN channel and the BER of a MF/symbol-by-symbol receiver are also shown.

The length of the correlators is varied between 3 and 7 symbols which yields a number of correlators varying between 4^3 and 4^7. It can be observed that at a distance of 300 km, the BER of a BC receiver is substantially improved compared to the MF case. This improvement is increased as the sequence length is extended up to 7-symbol sequences. This is due to the progressive reduction of the nonlinear ISI discussed before. On the other hand, as the power increases the beneficial effects of the BC receiver are less evident which can be attributed to the increasing inter-sequence interactions that are not taken into account by the receiver. However, it is interesting to note that applying ML detection over small sequences extends the region of powers where the performance of the AWGN channel can be closely approached. This behaviour can be clearly extended by increasing the correlators' length which jointly allows to account for a longer memory and to detect over longer sequences. However, the exponential scaling of the complexity with m makes the implementation of receiver rapidly infeasible.

The BC receiver structure can be considerably simplified if an orthonormal basis is used for the signal space generated by all possible signals $s(t,\mathbf{x}_k)$, with $k = 1,2,...,M^{2m+1}$. Based on the $2BT$-theorem (see e.g. [22, Ch. 2]), the number of dimensions of this vector space increases only linearly with the number of symbol periods spanned by the $s(t,\mathbf{x}_i)$, which instead are in a number growing exponentially with the number of symbols. A significant number of correlators can then be saved. However, the main contribution for the complexity of the structure in 4.5 comes from the argmax block which needs to perform $M^{2m+1}-1$ comparisons every M detected

Parameter Name	**Value**
Transmission Parameters	
Modulation Format	QPSK
Symbol Rate	32 GBaud
Number of WDM channels	1
Fibre Channel Parameters	
Attenuation Coefficient (α)	0.2 dB/km
GVD (β_2)	-21.66 ps^2/km
Nonlinearity Parameter (γ)	1.2 1/(W·km)
Span Length	350 km
EDFA NF	3 dB

Table 4.1: System parameters used the numerical study performed in this chapter.

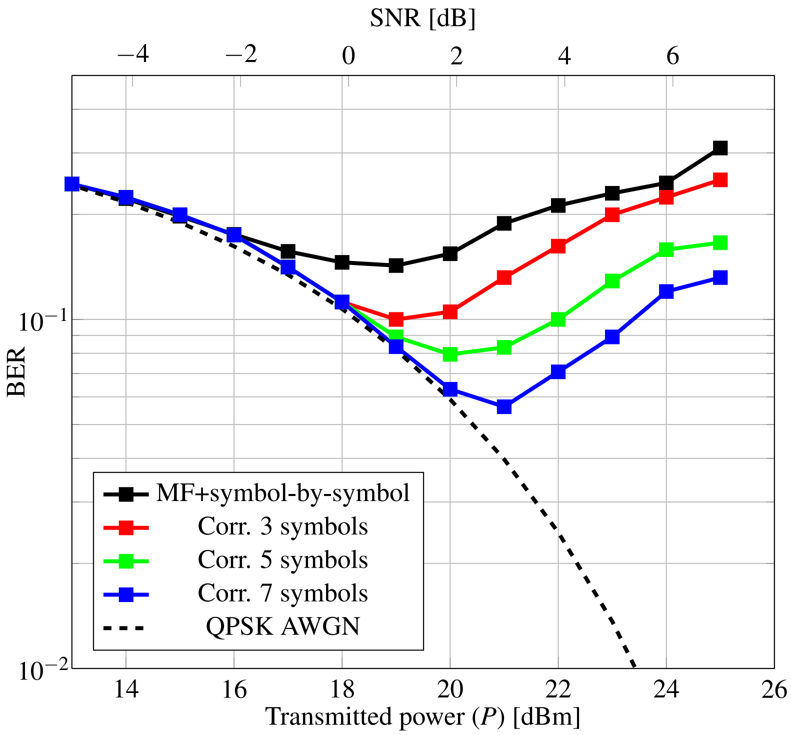


Figure 4.6: BER as a function of the transmitted power (P) for BC receivers with different correlator lengths and transmission distance L=350 km.

symbols, regardless of the number of vector components l_k at its input.

In the next section, the well-known Viterbi algorithm will be employed to solve this problem and arbitrarily increase the sequence length without incurring an exponentially scaling detection complexity.

4.4 Optimum MLSD for the nonlinear fibre channel

In order to avoid interaction between sequences, detection should be performed on sequences significantly longer than the channel memory. This can be done without incurring in an exponentially growing detection complexity by resorting to the well-known Viterbi algorithm [6, 23]. The Viterbi algorithm, allows to find the most likely sequence without having to compare all of them which would represent an infeasible task due to their exponential growth. The complexity increases only linearly with the length of the detected sequence and remains constant on per-detected-symbol basis.

Let us assume $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is the sequence of transmitted symbols and $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ is the sequence of channel observations on the waveform $r(t)$. A necessary condition to use the Viterbi algorithm is the factorisation of the sequence likelihood as

$$p(\mathbf{r}|\mathbf{x}) = \prod_{k=1}^{n} p(r_k|x_k, \sigma_k) \tag{4.16}$$

where

$$\sigma_k = (x_{k-m}, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{k+m}) \tag{4.17}$$

is the channel state and $2m$ is the channel memory. Such factorisation is possible for the discrete-time channel in Fig. 4.2b only when the r_k are extracted in some specific ways from the waveform $r(t)$. For instance, when the r_k are obtained as the output of a MF followed by a sampler at 1 Sa/sym (see Fig. 4.7a,b), the noise samples n_k are Gaussian and uncorrelated [17, Ch. 5], and as a result they are also statistically independent. Thus, Eq. (4.16) is verified as long as the $2m$ is equal or larger than the channel memory.

As discussed in section 4.2, in order to optimise the detection strategy r_k also needs to cover a sufficient statistic. For the channel considered in this chapter, r_k is a sufficient statistic if and only if the sequence of samples y_k is a complete representation of the waveform $y(t)$[5]. For a MF/sampler receiver this is not in general the case as we have seen that a transmitted sequence is not in general representable as in Eq. (4.1). This can also be understood by considering that the received signal bandwidth is in general expanded compared to the transmitted one because of the strong nonlinear effects. Thus, the MF cannot be used as basis function for the representation of the received signals. A simple way to obtain a sufficient statistic is instead based on the sampling theorem [22, Ch. 2]: sampling the received signal at a rate $F_s > 2B$ guarantees a complete representation of the waveform $y(t)$ of bandwidth B over a sequence period T.

In order to understand the impact of using a sufficient statistic on the receiver BER performance, three different receiver schemes shown in 4.7 are compared in the following: a typical MF/symbol-by-symbol receiver (4.7a); a MF/Viterbi processor (4.7b); a receiver filtering $r(t)$ over the bandwidth B of the received signal $y(t)$ (rectangular low-pass filter (RLPF)) with a subsequent Viterbi processor (4.7c). The Viterbi processor implements MLSD over arbitrarily long sequences without incurring the unbearable complexity of the BC receiver approach. Both receiver schemes in Figs. 4.7a, 4.7b represent suboptimal approaches as, albeit in different ways, they do not produce estimates based on sufficient statistics. In the third option the receiver instead attempts to estimate sequences of symbols based on a sufficient statistic and therefore it can be claimed optimal as long as the receiver memory matches that of the channel.

In Fig. 4.8, the BER as a function of the transmitted power is shown for the three receiver schemes in Fig. 4.7. The MF/MLSD receiver (4.7b) is implemented using 6-symbol states, whereas the RLPF/MLSD receiver (4.7c) uses either 2-, 4-, or 6-symbol states. The SNR axis is shown as a reference for the performance of the AWGN QPSK transmission.

All MLSD schemes show a significant improvement with respect to the MF/symbol-

---

[5] For instance in channels where $n(t)$ is not AWGN this condition might not be sufficient.

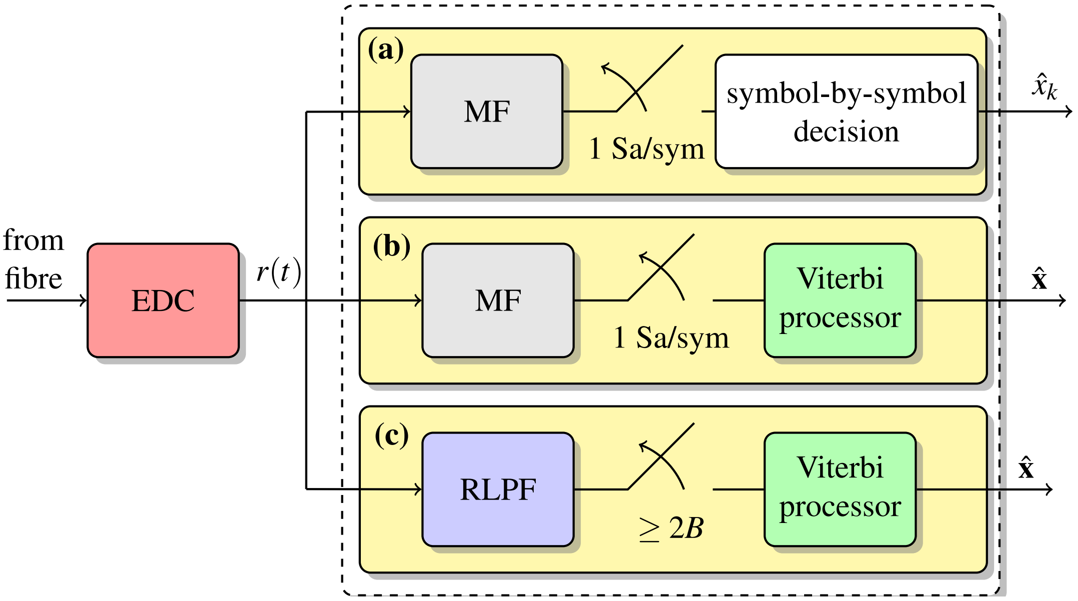


Figure 4.7: Schematic diagram of the MLSD receiver schemes analysed in this chapter.

by-symbol scheme (black line), demonstrating the effectiveness of sequence detection (as already shown for the correlators receiver). However it is interesting to note that, despite its 6-symbol states, the MF/MLSD receiver performs worse than all other RLPF schemes. This reflects the loss of information due to the suboptimal statistics used in the receiver in 4.7b. RLPF/MLSD receivers show a significantly better performance even for the reduced-state cases (2- and 4-symbol states), which can be attributed to the fact that they all operate on sufficient statistics. Particularly, the 6-symbol state RLPF receiver can closely approach the AWGN performance up to 22 dBm of transmitted power and shows a monotonically decreasing BER for the investigated range of transmitted powers. The reduced-state RLPF receivers instead show minimum BER at around 25 dBm. This behaviour is due to their reduced-memory states which do not allow the factorisation in Eq. (4.16). The BER floor observed for the 6-symbol state RLPF receiver also suggests that such state length is not sufficient to cover the actual channel memory. The BER floor could be also explained by assuming a minimum Euclidean distance between sequences which decreases with the transmitted power. Since the minimum Euclidean distance between received sequence waveforms dominates the sequence error probability and, as a result, the BER [6], [22, Ch. 7], this could result in a non-decreasing BER. However, non-increasing (as a function of transmitted power) minimum Euclidean distances between sequence waveforms were not observed in a preliminary investigation performed using the Viterbi trellis. However, further investigation on this topic is desirable.

In conclusion, in this section we have numerically tested the detection principles discussed in section 4.2 in a single-span optical fibre channel modelled as in Fig. 4.2. The results show remarkable BER gains when detection is performed by taking into

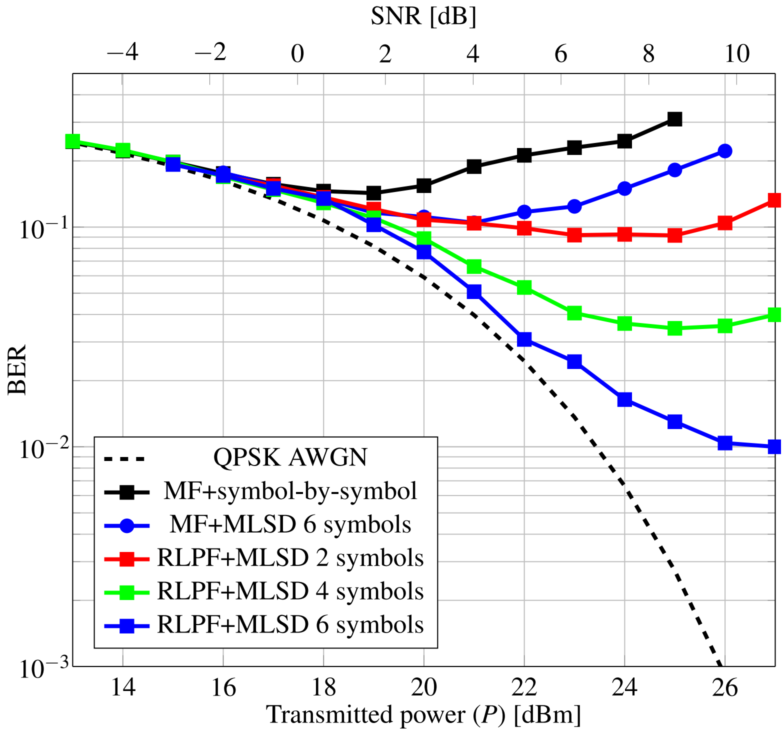


Figure 4.8: BER as a function of P for different detection strategies and transmission distance L=350 km.

account the channel memory and using a sufficient statistic on the data to be estimated.

4.5 Summary

In this chapter, receiver schemes, alternative to the zero-forcing nonlinear equalisation approach such as DBP, were investigated. In particular, the structure of the optimal receiver was studied when the channel is characterised by nonlinear ISI which is a peculiar characteristic of the unrepeatered fibre channel, due to the need to operate in the deep nonlinear regime of transmitted powers.

In such systems, due to the low receiver OSNR, zero-forcing equalisers incur a strong noise enhancement. When signal-ASE interactions can be neglected, it was shown that optimum receivers need to be based on MLSD schemes in order to properly account for the channel memory.

Different MLSD strategies were discussed and numerically implemented. The BER results show significant gains obtained by such improved receiver strategies compared to the conventional MF/symbol-by-symbol receiver. In particular, a receiver based on a bank-of-correlators was shown to attain a BER 2.5 times lower that a conventional MF/symbol-by-symbol receiver, when 4^7 correlators are employed. Close-to-optimum MLSD schemes, such as the RLPF/Viterbi detection with 6-symbol states, closely approach the AWGN performance up to very high transmitted powers (22 dBm) with a

BER saturating beyond that power. The minimum BER of the RLPF/Viterbi receiver was shown in this case to be more than one order of magnitude lower than the minimum BER achieved by a MF/symbol-by-symbol receiver. For the cases where the state length was reduced (4 or 2 symbols) the BER reached a minimum value between 4.3 and 1.5 times lower than that of a MF/symbol-by-symbol receiver.

Finally, the importance played by the collection of a sufficient statistic for the detection of the transmitted symbols was highlighted. Indeed, a MF/sampling receiver does not collect a sufficient statistic even when MLSD detection is performed. Using suboptimal statistics yields a markedly deteriorated performance compared to a MLSD receiver with the same number of states but using instead a sufficient statistic.

Despite their poor complexity scaling, MLSD receivers were shown to perform extremely well in single-span transmission scenarios, even with a relatively low number of states. Therefore, they represent a valid alternative to conventional receiver schemes (MF/symbol-by-symbol detection) or nonlinear equalisers for systems operating at very low OSNR.

MLSD techniques can also potentially be extended to multi-channel scenarios, with or without the knowledge of the other transmitted channels, and to multi-span systems. The major challenge in this case is still represented by the complexity, as the channel memory is significantly increased. Thus, ways to substantially reduce the number of states without compromising the performance are of great research interest. Finally, in systems where the signal-ASE noise cannot be neglected, such as a multi-span system, optimum detection strategies still need to be devised due to the incomplete knowledge of the signal-ASE noise statistical properties.

References

[1] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. John Wiley & Sons, 1965.

[2] T. Foggi, E. Forestieri, G. Colavolpe, and G. Prati, "Maximum-likelihood sequence detection with closed-form metrics in ook optical systems impaired by gvd and pmd," *J. Lightw. Technol.*, vol. 24, no. 8, pp. 3073–3087, 2006.

[3] P. Poggiolini and G. Bosco, "Long-haul WDM IMDD transmission at 10.7 Gbit/s in a dispersion-managed multispan system using MLSE receivers," *J. Lightw. Technol.*, vol. 26, no. 17, pp. 3041–3047, 2008.

[4] J. Zhao, M. E. McCarthy, A. D. Ellis, and P. Gunning, "Chromatic dispersion compensation using full-field maximum-likelihood sequence estimation," *J. Lightw. Technol.*, vol. 28, no. 7, pp. 1023–1031, 2010.

[5] D. Marsella, M. Secondini, and E. Forestieri, "Maximum likelihood sequence detection for mitigating nonlinear effects," *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, 2014.

[6] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, 1972.

[7] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, "Stochastic Digital Backpropagation," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3956–3968, 2014.

[8] N. V. Irukulapati, D. Marsella, P. Johannisson, E. Agrell, M. Secondini, and H. Wymeersch, "Stochastic digital backpropagation with residual memory compensation," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 566–572, 2016.

[9] G. Liga, A. Alvarado, E. Agrell, M. Secondini, R. I. Killey, and P. Bayvel, "Optimum detection in presence of nonlinear distortions with memory," in *Proc. European Conference on Optical Communication (ECOC)*, 2015.

[10] G. Agrawal, *Nonlinear Fiber Optics*, 5th ed., ser. Academic Press. Academic Press, 2013.

[11] Y. Cai, "MAP detection for linear and nonlinear ISI mitigation in long-haul coherent detection systems," in *2010 IEEE Photonics Society Summer Topical Meeting Series, (PHOSST)*, Playa del Carmen, Mexico, 2010.

[12] G. Bosco, P. Poggiolini, and M. Visintin, “Performance analysis of MLSE receivers based on the square-root metric,” vol. 26, no. 14, pp. 2098–2109, 2008.

[13] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, “Capacity of a nonlinear optical channel with finite memory,” *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, 2014.

[14] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “Properties of nonlinear noise in long , dispersion-uncompensated fiber links,” *Opt. Express*, vol. 21, no. 22, pp. 25 685–25 699, 2013.

[15] S. Yang, “The Capacity of Communication Channels with Memory,” Ph.D. dissertation, 2004.

[16] H. Pfister, “On the capacity of finite state channels and the analysis of convolutional accumulate-m codes,” Ph.D. dissertation, 2003.

[17] J. Barry, E. Lee, and D. Messerschmitt, *Digital Communication.* Springer US, 2004.

[18] R. A. Fisher, “On the mathematical foundations of theoretical statistics,” vol. A, no. 222, pp. 309—-368, 1922.

[19] H. L. V. Trees, *Classical Detection and Estimation Theory Part 1*, 2001, vol. 6.

[20] R. Gallager, *Stochastic Processes: Theory for Applications.* Cambridge University Press, 2013.

[21] “Optimal sequence detection and optimal symbol-by-symbol detection: similar algorithms,” *IEEE Trans. Commun.*, vol. 30, no. 1, pp. 152–157, 1982.

[22] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications.* Springer US, 1999.

[23] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 260–269, 1967.

5

Achievable information rates

In the previous chapters, two receiver-side strategies aimed at improving the performance of optical fibre communication systems were presented. The performance of such schemes was analysed based on two different metrics: the SNR in the case of EDC/DBP schemes, and the uncoded BER for the MLSD schemes. However, both metrics are, in general, unsuitable to predict the performance of receivers using forward error correction (FEC) [1, 2]. As discussed in this chapter, the performance of a coded system can be much more accurately predicted by means of the tools provided by *information theory*, a discipline founded by Shannon as a result of his 1948 celebrated paper "*A mathematical theory of information*" [3]. In this paper, mathematical quantities such as the *mutual information* (MI) and the *channel capacity* were defined and proved to be central in the description of the limits of information transmission through any communication channel.

It is the aim of this chapter to introduce the basic elements of information theory, in order to characterise the performance of coded optical fibre systems employing receiver schemes such as the ones discussed in chapters 3 and 4, i.e. EDC receivers, multi-channel DBP receivers and MLSD receivers. In particular, it is of paramount interest to quantify at which information rates transmission is possible with an arbitrary low BER after FEC decoding is performed (*error-free* transmission). In this chapter AIRs are discussed and quantified for each receiver scheme analysed in chapters 3 and 4.

The performance of each receiver scheme was studied within a specific channel configuration: long-haul WDM systems for EDC/DBP receivers and single-span, single-channel system for MLSD receivers. As for EDC/DBP schemes, a comprehensive study of the coded performance of long-haul spectrally-efficient WDM optical fiber transmission systems with different coded modulation decoding structures is described. AIRs were also calculated for three different square PM-MQAM formats, such as PM-16QAM, PM-64QAM, and PM-256QAM.

The AIRs performance of pragmatic encoder/decoder implementations were also considered. The four cases analysed combine hard-decision (HD) or soft-decision (SD) decoding together with either a bit-wise (BW) or a symbol-wise (SW) demapper, the last two suitable for binary and nonbinary codes, respectively. These quantities represent true indicators of the coded performance of the system for specific decoder implementations and when the modulation format and its input distribution are fixed.

As for the MLSD receivers, AIRs were quantified for the specific transmission scenario analysed in chapter 4, i.e. single-polarisation, single-channel, QPSK transmission in a single-span system.

5.1 AIRs for long-haul fibre systems using pragmatic FEC decoders

Next-generation long-haul transceivers will use powerful FEC and high-SE modulation formats, a combination known as coded modulation (CM). In order to provide reliable transmission, a FEC encoder maps blocks of information bits into longer blocks of coded bits that are sent through the channel at a nominal transmission rate. As a result, the information rate is, in general, lower than the nominal one by an amount that depends on the redundancy added by the FEC encoder, which in turn needs to be adjusted based on the quality of the channel. A key performance parameter is then the maximum rate at which an optical communication system can operate whilst maintaining reliable transmission of information.

To have an estimate of this rate, a widely used approach in the optical communication literature is based on identifying a pre-FEC BER threshold, for which a specific high-performance FEC code can guarantee an error-free performance after decoding. The code rate of such a coding scheme, multiplied by the raw transmission data rate, is used to quantify an AIR for that specific system configuration. On the other hand, information theory, founded by Shannon in his seminal paper [3], shows that quantities such as the MI can precisely indicate what is the maximum information rate at which a code can ensure an arbitrarily small error probability [4, 5]. Moreover, several recent works have shown that both the MI and the generalized mutual information (GMI) [6, 7] are more reliable indicators than the pre-FEC BER of the performance of coded optical fiber systems, regardless of the specific channel used for transmission [1, 8–13].

The *channel MI* (i.e., the MI including the channel memory) represents an upper limit on the AIRs for a given channel when a given modulation format is used along with an optimum decoder. In most cases, FEC codewords are equally likely, and thus the optimum decoder performs a maximum likelihood estimation on the received codewords. However, the implementation of such a decoder is in general prohibitive, both for complexity reasons and due to the lack of knowledge of the channel law. Instead of the optimum decoder, more pragmatic CM decoders are usually employed. Typical CM decoder implementations used in optical communications neglect the channel memory [11] and are, thus, suboptimal. Furthermore, their design involves two degrees of freedom. Each degree of freedom presents two options: hard-decision (HD) vs. soft-decision (SD) decoding and bit-wise (BW) vs. symbol-wise (SW) demapping, effectively producing four different design options.

These structures are representative of pragmatic decoders for FEC schemes employed in optical communication systems and comprehensively studied in the previous literature. SD-SW FEC schemes have been discussed in the context of optical communications for example in [10], [14, Sec. III-E]. SD-BW decoders are a more widespread choice for SD FEC and are typically used with binary low-density parity-check codes (see for example [1, 8, 9, 14–16]). As for HD FEC schemes, HD-SW decoders have been employed for Reed-Solomon codes which were adopted by the standard ITU-T G.975.1 [17] within the implementation of the so-called *super-FEC* scheme. Finally, HD binary FEC schemes such as extended Hamming codes and BCH codes were also adopted by the ITU-T G.975.1 standard. Alternative HD-BW schemes that have been recently considered for optical communications include *staircase codes* [18] and other types of so-called *generalized product codes* [19].

The channel MI is not in general an AIR for any of the four suboptimal schemes discussed above. Indeed, the adopted decoding strategy has a major impact on the AIRs, which can potentially be significantly lower than the channel MI. A common approach to calculate AIRs for specific decoder implementations is based on two steps: i) the memory of the optical fiber channel is neglected and the MI is calculated for an equivalent memoryless channel; ii) the *mismatched decoder* principle is used [20–23]. As discussed later in Sec. 5.3, each of these two methods results in a lower bound on the channel MI.

In [24] the memoryless MI was studied for coherent optical fibre systems using ring constellations. In [8, 9], the same quantity was used in an experimental scenario as a system performance metric for an SD coded system. In [11] and [1, Fig. 6], it was shown that when BW decoders are used, the GMI is a better metric to predict AIRs than the MI. The GMI has also been used to evaluate the performance of experimental optical systems in [25–27]. The memoryless MI and the GMI were also shown to be good post-FEC BER predictors for SD-SW (nonbinary) and SD-BW decoders, in [10] and [1] respectively. Finally, a study comparing SD-SW and HD-BW AIRs for polarization multiplexed (PM) quadrature-amplitude modulation (QAM) formats (PM-16QAM and PM-64QAM) was presented in [28], where electronic dispersion compensation (EDC) or digital backpropagation (DBP) are used at the receiver for a given transmission distance.

In the following, a comprehensive comparison of the AIRs of optical fibre transmission systems for different CM decoder implementations and for all transmission distances of interest for mid-range/long-haul terrestrial and transoceanic optical fiber links. The AIRs are also compared for different equalization techniques and over different PM-MQAM formats with nominal SE above 4 bits/sym per polarization such as PM-16QAM, PM-64QAM, and PM-256QAM. The results in this chapter show the design trade-offs in coded optical fiber systems where, for a given distance requirement, a compromise between transmission rates and transceiver complexity (modulation format, equalization, and decoding) must be found. To the best of our knowledge, this is the first time such an extensive study is performed for optical fiber communication systems.

5.2 Coded modulation for optical fibre systems

We consider the schematic diagram in Fig. 5.1, representing a generic multispan optical fiber communication system. Although in this work PM (4D) modulation formats are considered, for simplicity of the mathematical treatment in section 5.3, we neglect in our channel model any statistical dependence between the data transmitted on the two polarizations. Under this assumption, and for the modulation formats studied in this work (PM-16QAM, PM-64QAM, and PM-256QAM), the system under analysis can be reduced to a single-polarization (2D) one. At the transmitter, a CM encoder encodes a stream of N_b information bits $\boldsymbol{B}^{N_b} = [B_1, B_2, \ldots, B_{N_b}]$ into a sequence of N_s symbols $\boldsymbol{X}^{N_s} = [X_1, X_2, \ldots, X_{N_s}]$, each drawn from a set of M complex values $\mathcal{S} = \{s_1, s_2, ..., s_M\}$, where M is a power of 2.[1] The rate at which this operation is performed (in bits per symbol) is therefore given by

$$R = \frac{N_b}{N_s}. \tag{5.1}$$

In our analysis, we only consider the case where the symbols X_n forming a codeword $\boldsymbol{X}^{N_s}$ are independent, identically distributed (i.i.d.) random variables with equal probability $1/M$.[2]

---

[1] Throughout this chapter, boldface uppercase variables (e.g., $\boldsymbol{X}^N$) denote random vectors where the superscript indicates the size of the vector. Calligraphic letters (e.g., $\mathcal{S}$) represent sets.

[2] However, once a codebook is selected, symbols within codewords will appear as statistically dependent.

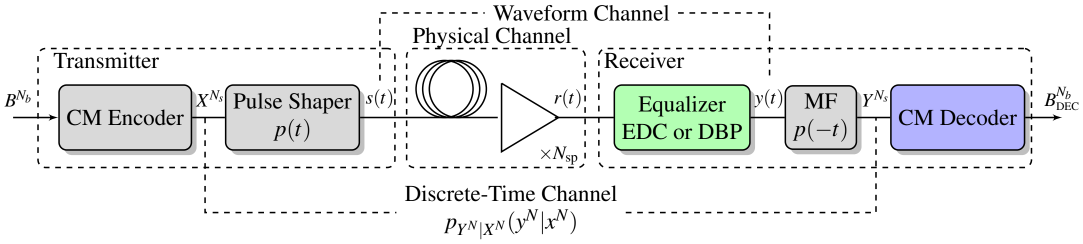


Figure 5.1: General schematic of the optical communication system analyzed in this work.

Although all CM encoders are inherently nonbinary encoders, the encoding process described above can be implemented in two different ways, as shown in Fig. 5.2. In the first implementation, shown in the top part of Fig. 5.2, the sequence of information bits is encoded using a binary FEC code and subsequently a memoryless mapper Φ is used to convert blocks of $\log_2 M$ bits into symbols of the constellation $\mathcal{S}$.[3] This implementation is naturally associated with CM decoders based on a demapper and a binary FEC decoder. The second implementation is shown in the bottom part of Fig. 5.2, where bits are first mapped into a sequence of nonbinary information symbols, which are then mapped into sequences of nonbinary coded symbols by a nonbinary FEC encoder [14, Sec. III-E], [10]. In this case, the decoding can be performed by a nonbinary FEC decoder.

We do not consider here cases where symbols are not uniformly distributed, i.e., when a probabilistic shaping on $\mathcal{S}$ is performed [31–36]. Moreover, in this study, we focus our attention on high SEs (>2 bits/sym/polarization), and thus the constellation $\mathcal{S}$ is assumed to be a square MQAM constellation where $M \in \{16, 64, 256\}$.

The symbols X_n are mapped, one every T_s seconds, onto a set of waveforms by a (real) pulse shaper $p(t)$, generating the complex signal

$$s(t) = \sum_{n=1}^{N_s} X_n p(t - nT_s). \tag{5.2}$$

The signal $s(t)$ propagates through N_{sp} spans of optical fiber (see Fig. 5.1), optically amplified at the end of each span by an erbium-doped fiber amplifier (EDFA). At the end of the fiber link, the signal is detected by an optical receiver. As shown in Fig. 5.1, the first part of the receiver includes an equalizer and a matched filter (MF), which are assumed to be operating on the continuous-time received waveform $r(t)$.[4] The equalizer performs a compensation of the most significant fiber channel impairments, either the linear ones only, as in the case of EDC, or both linear and nonlinear, as with DBP. The equalized (but noisy) waveform $y(t)$ represents the input of the detection stage

---

[3] Throughout this chapter, it is assumed that the mapping is done via the binary reflected Gray code [29, 30].

[4] The equalizer typically operates in the digital domain, but for a large enough sampling rate, the two representations are equivalent.

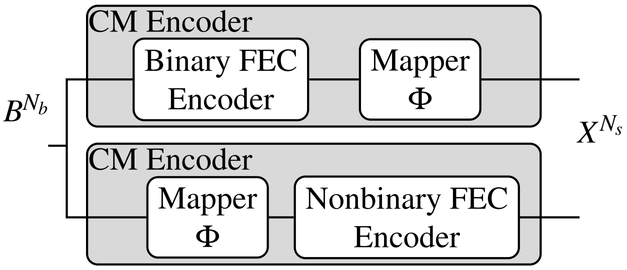


Figure 5.2: Two different implementation alternatives for the CM encoder in Fig. 5.1.

and can be therefore effectively considered as the output of the so-called *waveform channel* [37, Sec. 2.4]. Such a channel is formed by the cascade of the physical channel and the equalization block at the receiver, as shown in Fig. 5.1. The physical channel (i.e., fiber spans and amplifiers), also referred to as the nonlinear Schrödinger channel in [38], is described by the nonlinear Schrödinger equation [39, Sec. 2.3].

The receiver estimates the transmitted bits based on a set of observations $Y^{N_s} \triangleq \{Y_n\}_{n=1}^{N_s}$ where

$$Y_n = \int_{-\infty}^{+\infty} y(\tau)p(\tau - nT_s)d\tau \tag{5.3}$$

are samples extracted from the signal $y(t)$ using a MF matched to the transmitted pulse $p(t)$. As discussed in Chapter 4, and shown in [40, 41], Eq. (5.3) does not necessarily represent the optimum approach to reduce this particular waveform channel to a discrete-time one. Also, in all scenarios where residual correlated phase-noise is present due to fiber nonlinearities [23, 42], a phase-noise estimation block would improve the system performance, as shown for example in [43]. However, the focus of the following sections is on the performance of the CM encoder and decoder blocks, operating on the input and output of the *discrete-time channel*, regardless of the suboptimality of the observations Y^{N_s}.

In the following section, we discuss AIRs of the four decoding strategies shown in Fig. 5.3, representing different implementations of the CM decoder. The importance of these structures lies in the fact that they cover all the main options employing a memoryless demapper. Each BW configuration (see Figs. 5.3(b) and (d)) is character-

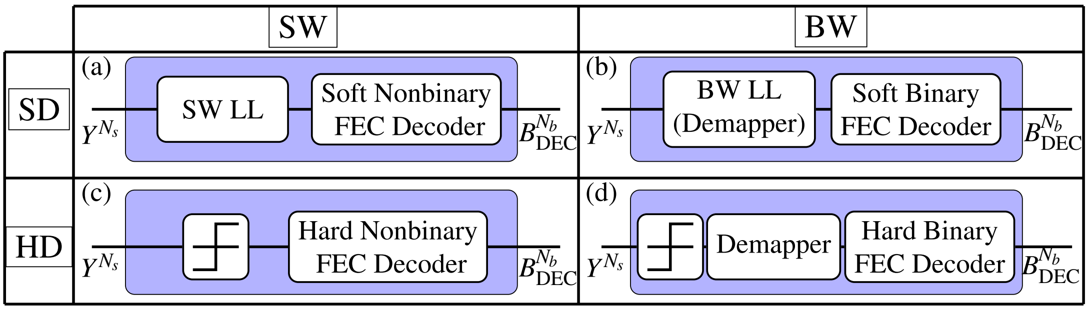


Figure 5.3: The four CM decoder implementations analised in this work.

ized by a CM decoder formed by two blocks: a memoryless demapper and a binary FEC decoder. The SW strategies (see Figs. 5.3(a) and (c)) are instead characterized by the adoption of a nonbinary decoder operating directly on symbol level metrics derived from the samples Y_n. Each of the HD schemes (see Figs. 5.3(c) and (d)) operates a symbol/bit level decision before the FEC decoder, which as a result operates on discrete quantities (*hard information*). In the SD case (see Figs. 5.3(a) and (b)), the decoder instead produces codeword estimates based on BW or SW log-likelihood (LL) values[5], which are distributed on a continuous range (*soft information*).

5.3 Information-theoretic preliminaries

5.3.1 Capacity and achievable rates

Consider an *information stable*, discrete-time channel with memory as defined in [44], and characterized by the sequence of probability density functions (PDFs)[6]

$$p_{Y^N|X^N}(y^N|x^N), \qquad N = 1, 2, \ldots \tag{5.4}$$

The maximum rate at which reliable transmission over such a channel is possible is defined by the capacity [44, eq. (1.2)]

$$C = \lim_{N\to\infty} \sup_{p_{X^N}} \frac{1}{N} I(Y^N; X^N) \tag{5.5}$$

where p_{X^N} is the joint PDF of the sequence X^N under a given power constraint. When p_{X^N} is fixed, the quantity

$$I(X^N; Y^N) = \mathbb{E}\left[\log_2 \frac{p_{Y^N|X^N}(Y^N|X^N)}{p_{Y^N}(Y^N)}\right] \tag{5.6}$$

in Eq. (5.5) is the MI between the two sequences of symbols X^N and Y^N, and

$$I_{\text{mem}} = \lim_{N\to\infty} \frac{1}{N} I(X^N; Y^N) \tag{5.7}$$

is the average *per-symbol* MI rate [4, 23], which has a meaning of channel MI. For a fixed N, Eq. (5.7) represents the maximum AIR for the channel in Eq. (5.4), and can be achieved by a CM encoder generating codewords X^{N_s} according to p_{X^N}, used along

---

[5] For the binary case, LL *ratios* are typically preferred for implementation reasons.

[6] Throughout this paper, $p_{Y|X}(y|x)$ denotes a joint conditional PDF for the random vectors Y and X, whereas a marginal joint PDF is denoted by $p_X(x)$.

with an optimum decoder.[7] Such a decoder uses the channel observations y^{N_s} to produce codeword estimates $X_{\mathrm{DEC}}^{N_s}$ based on the rule

$$X_{\mathrm{DEC}}^{N_s} = \underset{x^{N_s} \in \mathcal{S}^{N_s}}{\operatorname{argmax}} \, p_{Y^{N_s}|X^{N_s}}(y^{N_s}|x^{N_s}) \tag{5.8}$$

where the codeword likelihood $p_{Y^{N_s}|X^{N_s}}$ is calculated based on the knowledge of the channel law Eq. (5.4).

The expression of the channel law Eq. (5.4), for N large enough to account for the channel memory, remains so far unknown for the optical fiber channel despite previous attempts to derive approximated [45, 46] or heuristic [47] analytical expressions. On the other hand, brute-force numerical approaches appear prohibitive. An immediate consequence is that the exact channel MI for a given modulation format cannot be calculated. The second consequence is that the optimum receiver potentially achieving a rate $R = I_{\mathrm{mem}}$ cannot be designed. However, using the mismatched decoder approach, it is still possible to calculate nontrivial AIRs for the optical fiber channel in Fig. 5.1, when suboptimal but practically realizable CM encoders and decoders are used, such as the ones described in section 5.2 (see Fig. 5.3).

The method of the mismatched decoder to calculate AIRs for specific decoder structures originates from the works in [20], later extended to channels with memory in [21] and recently applied to optical fiber systems in, e.g., [22, 23, 28]. This approach consists in replacing, in the calculation of the channel MI, the unknown channel law with an auxiliary one, obtaining a lower bound. Moreover, such a bound represents an AIR for a system using the optimum decoder for the auxiliary channel. The tightness of such a lower bound depends on how similar the auxiliary channel is to the actual one. On the other hand, no converse coding theorem is available for the bound obtained using a given auxiliary channel. In other words, even when a mismatched decoder is used, the estimated rate is not necessarily the maximum achievable rate. Counterexamples have been shown, e.g., in [48].

Nevertheless the AIRs calculated via the mismatched decoder approach still represent an upper bound on the rates of most, if not all, coding schemes used in practice. Furthermore they are a strong predictor of the post-FEC BER of such schemes, as shown in [1, 8–10].

5.3.2 AIRs for SD CM decoders

Since each of the CM decoders presented in section 5.2 neglects the memory of the channel in Eq. (5.4), a first decoding *mismatch* is introduced. In what follows, we

---

[7] The channel can be seen as *block-wise memoryless*, and thus, codewords should be constructed using blocks of N symbols drawn independently from p_{X^N}.

discuss this mismatch using the SD-SW case (see Fig. 5.3(a)) as a representative example of all other CM decoders.

For the SD-SW, the nonbinary decoder requires SW likelihoods $p_{Y_n|X_n}$, with $n = 1,2,\ldots,N$. These N PDFs can be derived for each n by marginalizing the joint PDF in Eq. (5.4). For simplicity, however, practical implementations use a single PDF across the block of N symbols. We choose the PDF in the middle of the observation block, i.e., at time instant $n = \hat{n} = \lceil N/2 \rceil$. The marginalization of Eq. (5.4) in this case gives

$$p_{Y_{\hat{n}}|X_{\hat{n}}}(y_{\hat{n}}|x_{\hat{n}}) = \int_{\mathbb{C}^{N-1}} p_{Y^N|X_{\hat{n}}}(y^N|x_{\hat{n}})d\tilde{y}^{N-1} \tag{5.9}$$

where $\mathbb{C}$ denotes the complex field, $\tilde{y}^{N-1} \triangleq [y_1,\ldots,y_{\hat{n}-1},y_{\hat{n}+1},\ldots,y_N]$, and the conditional PDF $p_{Y^N|X_{\hat{n}}}$ in Eq. (5.9) can be expressed as

$$p_{Y^N|X_{\hat{n}}}(y^N|x_{\hat{n}}) = \frac{1}{M^{N-1}} \sum_{\tilde{x}^{N-1}\in\mathcal{S}^{N-1}} p_{Y^N|X^N}(y^N|x^N) \tag{5.10}$$

where $\tilde{x}^{N-1} \triangleq [x_1,\ldots,x_{\hat{n}-1},x_{\hat{n}+1},\ldots,x_N]$.

The choice for the single PDF to be the one in the middle of the observation block is arbitrary. However, this choice is justified by the fact that $p_{Y_{\hat{n}}|X_{\hat{n}}}(y_{\hat{n}}|x_{\hat{n}})$ will be a good approximation of all other PDFs $p_{Y_n|X_n}(y_n|x_n)$ with $n = 1,2,\ldots,N$ when N is large.

The demapper is then assuming a channel that is stationary across the block of N symbols.[8] This channel is fully determined by a PDF $p_{Y|X}(y|x)$ defined as

$$p_{Y|X}(y|x) \triangleq p_{Y_{\hat{n}}|X_{\hat{n}}}(y|x). \tag{5.11}$$

When i.i.d. symbols are transmitted, the MI for this auxiliary memoryless channel is given by

$$I_{\text{SD-SW}} = \frac{1}{M}\sum_{i=1}^{M} \int_{\mathbb{C}} p_{Y|X}(y|s_i) \log_2 \frac{p_{Y|X}(y|s_i)}{p_Y(y)} dy. \tag{5.12}$$

The SD-SW MI in Eq. (5.12) is an AIR for the SD-SW decoder structure in Fig. 5.3(a), where the demapper computes LLs $\log p_{Y|X}(y|x)$, and the FEC decoder estimates each transmitted codeword using Eq. (5.8) with a codeword likelihood given by

$$p_{Y^{N_s}|X^{N_s}}(y^{N_s}|x^{N_s}) = \prod_{n=1}^{N_s} p_{Y|X}(y_n|x_n). \tag{5.13}$$

In most cases, the channel law $p_{Y^N|X^N}$ is unknown and therefore $p_{Y|X}(y|x)$ is not available in closed form to the receiver. Also, numerical estimations of $p_{Y|X}(y|x)$ are often prohibitive. As a result, practical implementations not only ignore the memory of the channel (first mismatch), but also make an a priori assumption on the PDF $p_{Y|X}(y|x)$.

---

[8] Here we refer to *wide-sense* stationarity [49, Sec. 3.6.1].

This assumption introduces a *second* mismatch, which we discuss in what follows.

Most receivers assume a circularly symmetric Gaussian distribution for Eq. (5.11). In this case, an AIR is given by [28, eq. (2)]

$$\tilde{I}_{\text{SD-SW}} = \frac{1}{M}\sum_{i=1}^{M}\int_{\mathbb{C}} p_{Y|X}(y|s_i)\log_2\frac{q_{Y|X}(y|s_i)}{q_Y(y)}dy \tag{5.14}$$

where

$$q_{Y|X}(y|x) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|y-x|^2}{\sigma^2}\right) \tag{5.15}$$

represents the auxiliary Gaussian channel with complex noise variance σ^2, which accounts for the contributions of both ASE and nonlinear distortions.

As shown in [50, 51], the marginal PDF for the optical fiber channel is in most practical cases well approximated by a circularly symmetric Gaussian distribution.[9] This near-Gaussianity property can be attributed to the central limit theorem, as it is the result of the accumulation of many random nonlinear interference contributions.

Therefore, as pointed out in [28], we generally have

$$\tilde{I}_{\text{SD-SW}} \approx I_{\text{SD-SW}}. \tag{5.16}$$

In this case, as discussed in section 5.3.4, the AIRs of SD-SW decoders can be quite accurately estimated using the MI expression for the AWGN channel and the effective signal-to-noise ratio (SNR) at the MF output

$$\text{SNR} = \frac{\mathbb{E}\left[|X|^2\right]}{\sigma^2}. \tag{5.17}$$

In the SD-BW implementation (see Fig. 5.3(b)), for each received symbol Y the demapper generates $\log_2 M$ BW LLs [7, Ch. 3], [1]. These LLs are usually obtained assuming no statistical dependence between bits belonging to the same transmitted symbol. When such LLs are calculated based on a memoryless channel law $p_{Y|X}(y|x)$, the relevant quantity for the coded performance is the GMI [7, eq. (4.54)], [1, eq. (24)]

$$I_{\text{SD-BW}} = \sum_{k=1}^{\log_2 M} I(B_k;Y) \tag{5.18}$$

where B_k denotes the k-th bit of X and $I(B_k;Y)$ denotes the MI between transmitted bits and received symbols.

When the LLs are calculated using the auxiliary channel in Eq. (5.15) instead of the

---

[9] A deviation from a circularly symmetric Gaussian PDF can be observed, e.g., in the following cases: amplification schemes different from EDFA (such as Raman amplifiers) [45], dispersion-managed links (see for instance [23]), and for very high transmitted powers.

true channel, the GMI is lower-bounded by

$$\tilde{I}_{\text{SD-BW}} = \frac{1}{M} \sum_{k=1}^{\log_2 M} \sum_{b \in \{0,1\}} \sum_{i \in I_k^b} \int_{\mathbb{C}} p_{Y|X}(y|s_i) g_{k,b}(y) dy \tag{5.19}$$

where I_k^b is the subset of indices of the constellation $\mathcal{S}$ having the k-th bit equal to $b \in \{0,1\}$ and

$$g_{k,b}(y) \triangleq \log_2 \frac{\sum_{j \in I_k^b} q_{Y|X}(y|s_j)}{\frac{1}{2} \sum_{j=1}^{M} q_{Y|X}(y|s_j)}. \tag{5.20}$$

Similarly to the SD-SW case, for the optical fiber channel in Fig. 5.1 we have $\tilde{I}_{\text{SD-BW}} \approx I_{\text{SD-BW}}$.

5.3.3 AIRs for HD CM decoders

As illustrated in Figs. 5.3(c) and (d), the HD decoders are preceded by a memoryless threshold device casting the channel samples Y^{N_s} into a discrete set of values. In the SW case (Fig. 5.3(c)), such a device provides a sequence of hard SW estimates $\hat{X}^{N_s}$ that are passed to a nonbinary decoder.

The channel

$$P_{\hat{X}^N|X^N}(\hat{x}^N|x^N) \tag{5.21}$$

will in general show memory across multiple symbols $\hat{X}_n$. However, in analogy with Eq. (5.9), we can replace Eq. (5.21) with an equivalent memoryless channel defined by

$$P_{\hat{X}|X}(x_j|x_i) \triangleq p_{ij} \quad \text{for} \quad i,j = 1,2,...,M \tag{5.22}$$

where the p_{ij} are the SW crossover probabilities. Using the same argument on the channel memory used for the SD-SW case, the quantity

$$I_{\text{HD-SW}} = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{M} p_{ij} \log_2 \frac{p_{ij}}{\frac{1}{M} \sum_{p=1}^{M} p_{pj}} \tag{5.23}$$

represents an AIR for the HD-SW CM decoder in Fig. 5.3(c).[10]

When the HD decoder structure is preserved but a *binary* decoder is instead used (Fig. 5.3(d)), the threshold device needs to be followed by a symbol-to-bit demapper producing a sequence of pre-FEC bits estimates $\hat{B}^{N_b}$. Again, although the resulting binary channel might show memory, the HD FEC decoder typically neglects it and the most likely codeword is calculated based on each single detected bits. The auxiliary

---

[10] The rate $I_{\text{HD-SW}}$ in Eq. (5.23) is achievable with a nonbinary FEC decoder that is matched to the channel transition probabilities p_{ij}, but not necessarily with a standard nonbinary FEC decoder based on minimizing the Hamming distance.

channel law $P_{\hat{B}_k|B_k}(\hat{b}|b)$ can be in this case represented by a set of $\log_2 M$ pairs of transition probabilities

$$P_{\hat{B}_k|B_k}(\hat{b}|b) = \begin{cases} p_k & \text{for } b=0, \hat{b}=1 \\ q_k & \text{for } b=1, \hat{b}=0 \end{cases} \tag{5.24}$$

for k=1,2,...,$\log_2 M$, where p_k and q_k are the crossover probabilities for the bits $\hat{B}_k$ and B_k in k-th position within the symbols $\hat{X}$ and X, respectively. The quantity

$$I_{\text{HD-BW}} = \sum_{k=1}^{\log_2 M} I(B_k;\hat{B}_k), \tag{5.25}$$

where $I(B_k;\hat{B}_k)$ is given, after some simple algebra, by

$$\begin{aligned} I(B_k;\hat{B}_k) = 1 + \frac{1}{2}\Bigg[&\log_2 \frac{(1-p_k)(1-q_k)}{(1-p_k+q_k)(1-q_k+p_k)} \\ &+p_k \log_2 \frac{p_k(1-p_k+q_k)}{(1-p_k)(1-q_k+p_k)} \\ &+q_k \log_2 \frac{q_k(1-q_k+p_k)}{(1-q_k)(1-p_k+q_k)}\Bigg], \end{aligned} \tag{5.26}$$

represents an AIR for an HD-BW CM decoder as in Fig. 5.3(d).[11]

5.3.4 Relationships between AIRs

The relationships between the above discussed AIRs are summarized by means of the graph in Fig. 5.4. Nodes that are connected in the graph indicate the existence of an inequality between the quantities in each of the nodes. The direction of the arrows show which quantity is upper-bounding the other.

For any given input distribution, the rate I_{mem} upper-bounds all other quantities. In particular, we have

$$I_{\text{mem}} \geq I_{\text{SD-SW}} \geq \tilde{I}_{\text{SD-SW}}, \tag{5.27}$$

where the first inequality can be proven using the chain rule of the MI (see [5, Sec. 2.5.2], [24, Sec. IV], [52]). The second inequality instead reflects the additional mismatch caused by a memoryless demapper based on Eq. (5.15) rather than on Eq. (5.9). The proof of this inequality follows from the definitions Eq. (5.12) and Eq. (5.14) and is given in [21, Sec. VI]. Due to the assumption of independent bits within each transmitted

---

[11] An average binary symmetric channel (BSC) could be used instead of Eq. (5.24) as an auxiliary channel. This would result in the well-known BSC capacity which might be a pessimistic AIR for generic HD-BW decoders. However, such a quantity is a more suitable AIR for HD-BW decoders that disregard both bit position and channel asymmetry.

symbol in the calculation of Eq. (5.18), it can also be shown that [7, Sec. 4.4]

$$I_{\text{SD-SW}} \geq I_{\text{SD-BW}} \geq \tilde{I}_{\text{SD-BW}}. \tag{5.28}$$

Again, the second inequality reflects the loss of information of a mismatched demapper calculating BW LLs based on Eq. (5.15) rather than on Eq. (5.9).

Due to the data-processing inequality [5, Sec. 2.4] and the mismatch of the illustrated HD decoders to the potential channel memory, we have

$$I_{\text{SD-SW}} \geq I_{\text{HD-SW}}, \tag{5.29}$$

$$I_{\text{SD-BW}} \geq I_{\text{HD-BW}}. \tag{5.30}$$

Finally, similarly to the SD case, we have

$$I_{\text{HD-SW}} \geq I_{\text{HD-BW}}. \tag{5.31}$$

In general, nothing can be said on the relationship between $I_{\text{SD-BW}}$ and $I_{\text{HD-SW}}$. Also, no systematic inequality holds between the *mismatched* versions of the SD AIRs ($\tilde{I}_{\text{SD-SW}}$, $\tilde{I}_{\text{SD-BW}}$) and the HD AIRs ($I_{\text{HD-SW}}$, $I_{\text{HD-BW}}$). However, as already discussed in section 5.3.2, for the optical fiber channel the mismatched AIRs are expected to be very close to the AIRs obtained with perfect knowledge of the channel marginal PDF in Eq. (5.9).

When the channel is indeed AWGN, clearly

$$I_{\text{mem}} = I_{\text{SD-SW}} = \tilde{I}_{\text{SD-SW}}, \tag{5.32}$$

$$I_{\text{SD-BW}} = \tilde{I}_{\text{SD-BW}}. \tag{5.33}$$

In this case, as illustrated in 5.4, $I_{\text{SD-SW}}$ and $I_{\text{HD-SW}}$ are the maximum AIR for SD-SW and

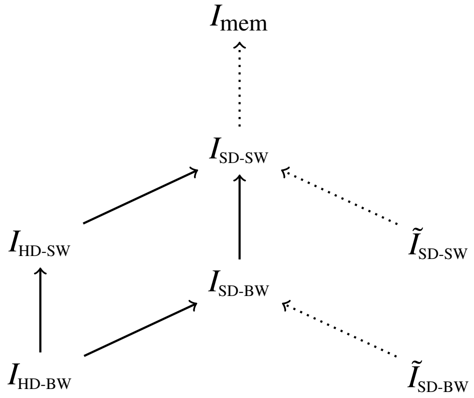
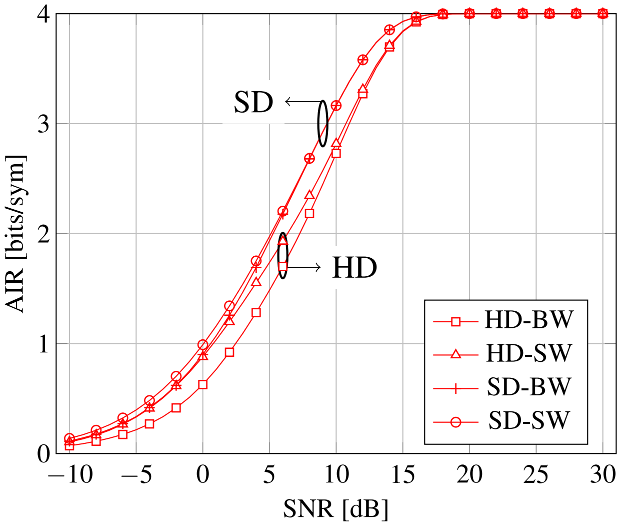
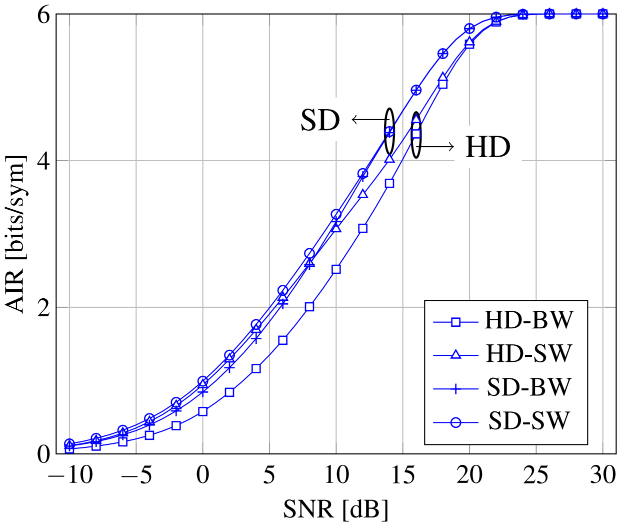


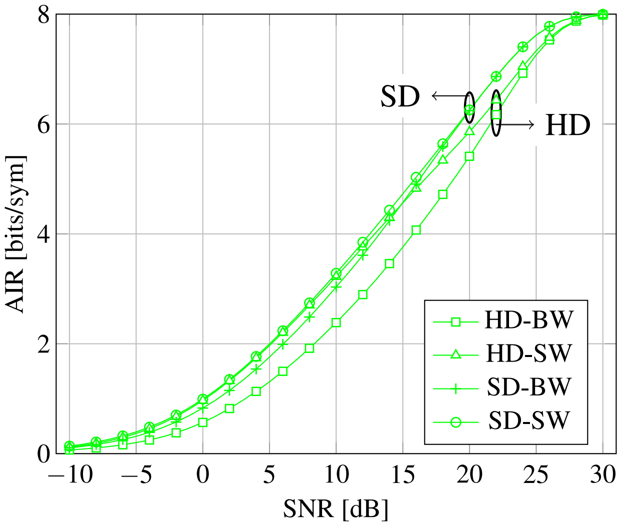
Figure 5.4: Graph showing relationships between the information-theoretic quantities presented in this chapter. Lines between nodes indicate an inequality, where the arrows point towards the upper bound. Dotted arrows indicate inequalities which become equalities for the AWGN channel.



(a) 16QAM



(b) 64QAM



(c) 256QAM

Figure 5.5: AIRs as a function of the SNR for different modulation formats in an AWGN channel.

HD-SW decoders, respectively [3], since each demapper is matched to the channel.[12] Conversely, for BW decoders, rates higher than $I_{\text{SD-BW}}$ and $I_{\text{HD-BW}}$ are still achievable (see, e.g., [48]).

In order to better illustrate the relationships discussed above, the four AIRs in Eqs. (5.12),(5.18), (5.23), and (5.25) were calculated for the AWGN channel. In Fig. 5.5, $I_{\text{SD-SW}}$, $I_{\text{SD-BW}}$, $I_{\text{HD-SW}}$, and $I_{\text{HD-BW}}$ are shown vs. the SNR in Eq. (5.17) for the three MQAM formats analised in this chapter: 16QAM, 64QAM, and 256QAM. For 16QAM, the HD AIRs are below both of the SD AIRs. It should be noted that for SD decoders, a negligible penalty is incurred by using a BW structure. As the modulation order is increased, and for low enough SNR values, it can be observed that the HD-SW AIRs match or exceed the SD-BW AIRs. Also, in this regime, the performance of these two decoders are comparable to the SD-SW one. This behaviour is clearer for a 256QAM modulation format, for which a more significant penalty is incurred by using BW demapping in an SD CM decoder, whereas the HD-SW structure performs as well as the SD counterpart. When the modulation format cardinality increases, a HD-BW decoder incurs, in general, significant penalties in AIR. Finally, the inequalities in Eqs. (5.28)–(5.31) can be seen to hold for all modulation formats shown, as expected.

5.4 AIRs using EDC receivers

In this and in the following section, numerical results based on SSFM simulations of long-haul optical fiber transmission are presented. As shown in Fig. 5.1, the simulated system consists of an optical fiber link comprising multiple standard single-mode fiber spans (parameters shown in Table 5.1), amplified, at the end of each span, by an EDFA which compensates for the span loss. At the transmitter, after the CM encoder, PM square MQAM formats (PM-16QAM, PM-64QAM, PM-256QAM) were modulated using a root raised cosine (RRC) filter $p(t)$. For each polarization of each WDM channel, independent sequences of 2^{18} symbols were transmitted. At the receiver, in order to obtain ideal equalization performance, the sampling rate at which the equalizer was operated was the same as the fiber propagation simulation (512 GSa/s).

After the MF (see Fig. 5.1) and sampling at 1 Sa/sym, AIRs calculations were performed based on the schemes shown in Fig. 5.3. In particular, we used Eqs. (5.14)–(5.15), (5.19)–(5.20), (5.23), and (5.25) to evaluate $\tilde{I}_{\text{SD-SW}}$, $\tilde{I}_{\text{SD-BW}}$, $I_{\text{HD-SW}}$, and $I_{\text{HD-BW}}$, respectively. For the calculation of $\tilde{I}_{\text{SD-SW}}$ and $\tilde{I}_{\text{SD-BW}}$ in Eqs. (5.14) and (5.19), Monte-Carlo integration was performed [53], using the 2^{18} channel samples (transmitted symbols) to estimate the variance σ^2 of $q_{Y|X}(y|x)$. In order to calculate $I_{\text{HD-SW}}$ and $I_{\text{HD-BW}}$, a Monte Carlo estimation [54, Sec. 5.6.1] of the probabilities p_{ij} and p was performed using the pairs of sequences $(X^{N_s}, \hat{X}^{N_s})$ and $(B^{N_b}, \hat{B}^{N_b})$, respectively.

---

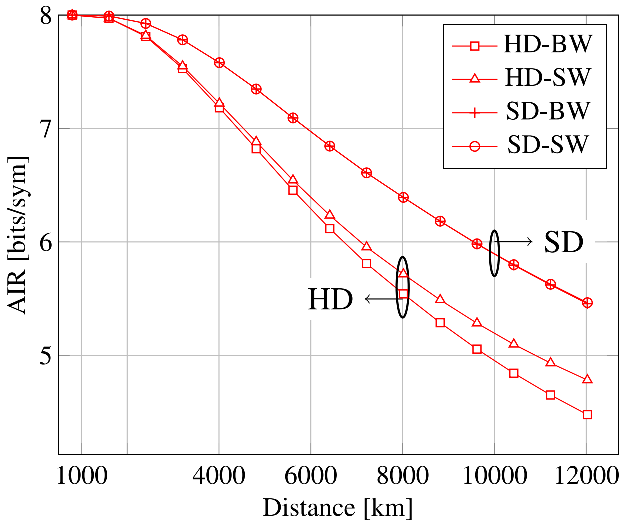
[12] In the HD-SW case, the channel seen by the nonbinary FEC decoder is the one in Eq. (5.22).

Table 5.1: System parameters

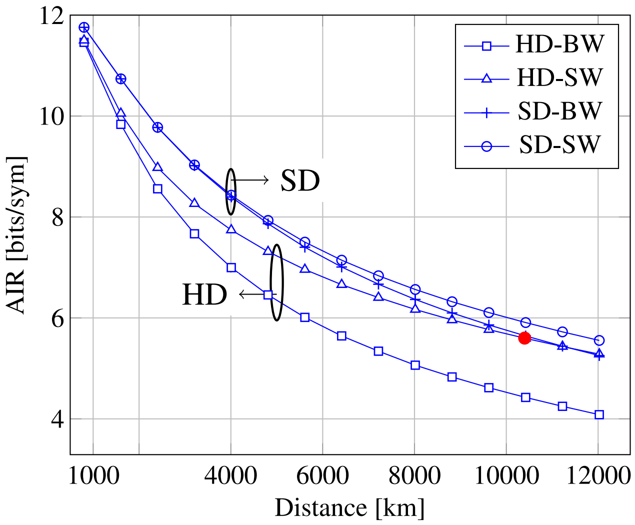
Parameter Name	**Value**
Transmitter Parameters	
WDM Channels	5
Symbol Rate	32 Gbaud
RRC Roll-Off	0.01
Channel Frequency Spacing	33 GHz
Fibre Channel Parameters	
Attenuation (α)	0.2 dB/km
Dispersion Parameter (D)	17 ps/nm/km
Nonlinearity Parameter (γ)	1.2 1/(W·km)
Fibre Span Length	80 km
EDFA Gain	16 dB
EDFA Noise Figure	4.5 dB
Numerical Parameters	
SSF Spatial Step Size	100 m
Simulation Bandwidth	512 GHz

In Figs. 5.6, three sets of results are shown for an EDC receiver. Each set shows the AIR vs. transmission distance for PM-16QAM, PM-64QAM, and PM-256QAM with the four CM decoder structures discussed in section 5.2. For each distance, equalization scheme, and CM decoder investigated, the transmitted power was optimized, resulting in different optimal powers. The investigated link lengths span the typical distances of mid-range to long-haul terrestrial links (typically 1000–3000 km), long-haul submarine (3000–5000 km), and transoceanic links (6000–12000 km).

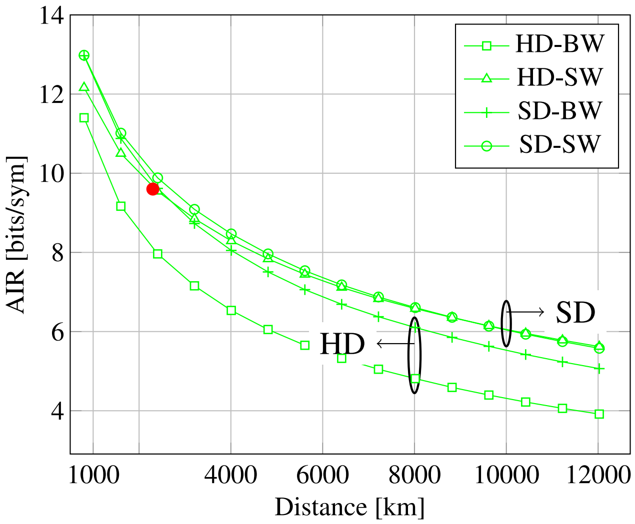
For PM-16QAM (Fig. 5.6a), SD decoders significantly outperform the HD ones, particularly for long distances. SD-BW decoders incur small penalties compared to the SD-SW implementation at all distances of interest. This can be explained by observing Fig. 5.5a, where the performance of PM-16QAM differs for SD-SW and SD-BW decoders only for very small SNR values ($\leq$2 dB). As shown in Fig. 5.6b, for PM-64QAM, SD decoders show a significant advantage over their HD counterparts (see [28] for SD-SW vs. HD-BW) and again SD-BW decoders have identical performance as the SD-SW ones at short distances. However, as the distance is increased, the AIRs of the HD-SW schemes match the SD-BW ones (see filled red circles in Fig. 5.6b and 5.6c), significantly outperforming the HD-BW rates. This trend is even more prominent for PM-256QAM (Fig. 5.6c). For this format, a crossing between the SD-BW and HD-SW AIRs can be observed at around 2300 km distance (filled red circles). More importantly, in the long distance regime, the HD-SW scheme matches the performance of the SD-SW one, with no significant penalty observed. Also, it can be noted that the HD-BW scheme shows a significant penalty ($>$3 bits/sym for long distances) compared to all other implementations.



(a) PM-16QAM



(b) PM-64QAM



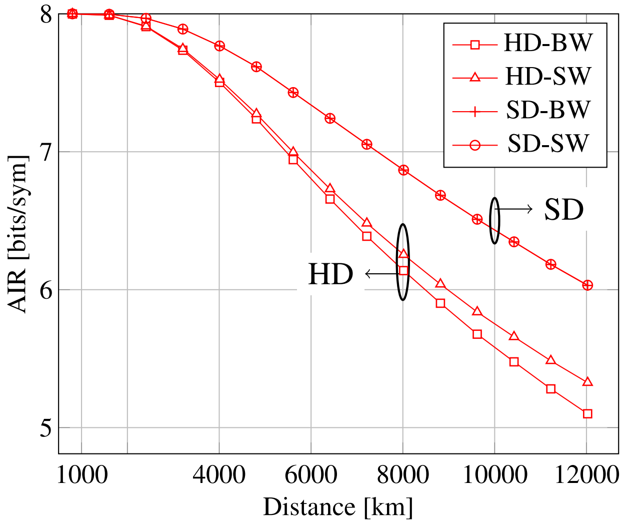
(c) PM-256QAM

Figure 5.6: AIRs as a function of the transmission distance for EDC.

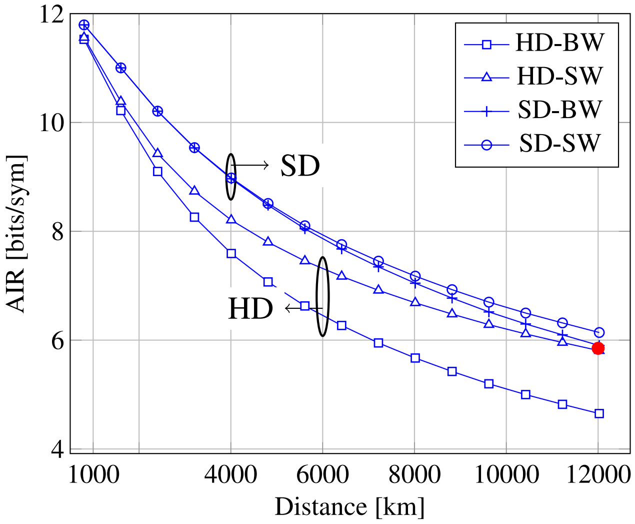
5.5 AIRs using DBP receivers

In the case where single-channel DBP is applied (Fig. 5.7), rather small AIR gains can be noticed in general, as compared to the EDC case (Fig. 5.6). This can be attributed to the fact that the compensation of the nonlinearity generated by only one channel out of the five transmitted gives only a marginal improvement of the optimum SNR at each transmission distance. However, performance differences can be noticed for higher order formats and long distances. Specifically, the distance at which the HD-SW transceiver matches the performance of the SD-BW one for PM-64QAM is increased from 10000 km to 12000 km (filled red circles in Fig. 5.6b and Fig. 5.7b), and for PM-256QAM the crossing point between HD-SW and SD-BW is moved from 2300 km to 3000 km (filled red circles in Fig. 5.6c and Fig. 5.7c).

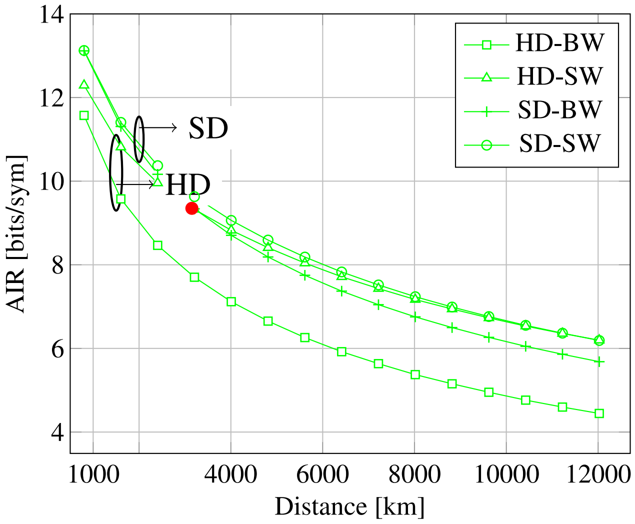
Finally, when full compensation of signal–signal nonlinear distortion is performed via full-field DBP (Fig. 5.8), a remarkable increase in the AIRs compared to the other equalization schemes can be observed for all decoding strategies and all modulation formats. Fig. 5.8a shows that, for PM-16QAM, the full nominal SE (8 bits/sym) can be achieved up to a distance of approximately 6000 km and by only using an HD-BW decoder (squares). This rate drops by only 0.5 bits/sym at 12000 km if SD decoders are used, and by an additional 0.5 bits/sym (to 7 bits/sym) when HD decoders are adopted. Fig. 5.8a also shows that when PM-16QAM is used in conjunction with full-field DBP, switching from a binary to a nonbinary scheme does not result in any significant AIR increase, as long as the same FEC decoding strategy (HD or SD) is maintained. Higher rates can be achieved using PM-64QAM (Fig. 5.8b) and PM-256QAM (Fig. 5.8c) in conjunction with SD decoders. Again, binary and nonbinary SD schemes perform identically. For these higher order modulation formats, HD-BW decoders incur significant penalties compared to SD decoders. For PM-64QAM, this penalty becomes larger than 0.5 bits/sym for distances larger than 4000 km whereas for PM-256QAM, they become larger than 0.5 bits/sym already for distances longer than 1500 km. At long distances, the penalty increases to up to 1.6 bits/sym for PM-64QAM and 2.5 bits/sym for PM-256QAM. An improvement can be obtained by using HD-SW decoders, particularly in the long-distance regime. For PM-64QAM, the AIR gap from SD decoders is reduced to 0.5 bits/sym at 12000 km. For PM-256QAM, HD-SW decoders in general largely outperform HD-BW decoders and show performances similar to SD decoders beyond distances of 7000 km, also outperforming SD-BW decoders beyond 8000 km.



(a) PM-16QAM

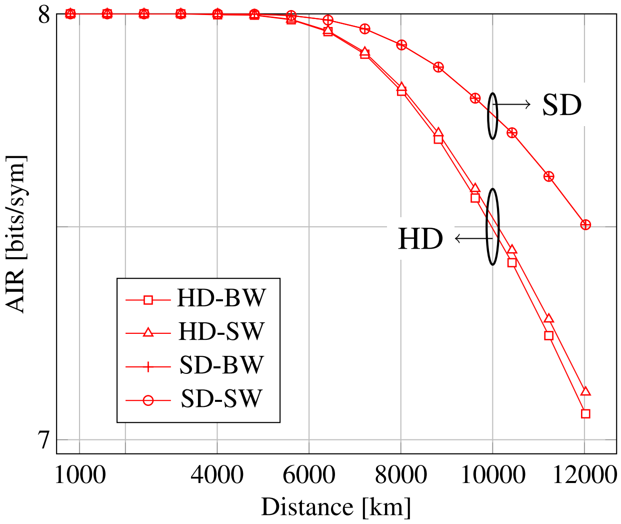


(b) PM-64QAM

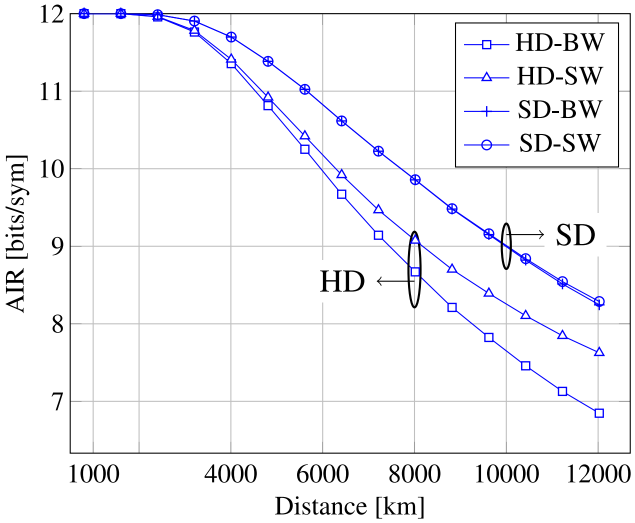


(c) PM-256QAM

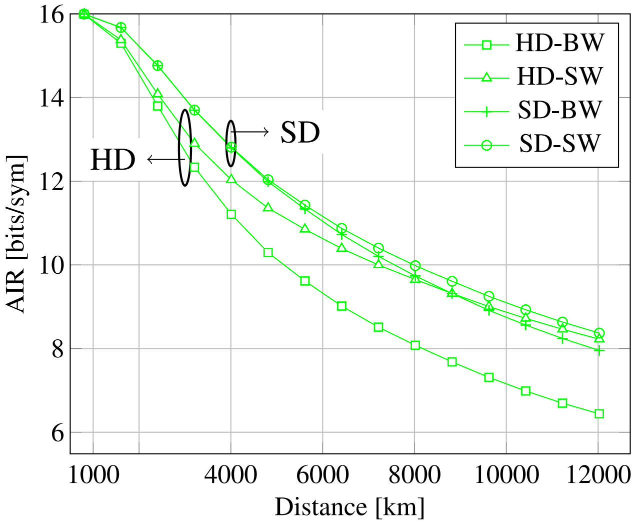
Figure 5.7: AIRs as a function of the transmission distance for single-channel DBP.



(a) PM-16QAM



(b) PM-64QAM



(c) PM-256QAM

Figure 5.8: AIRs as a function of the transmission distance with full-field DBP.

5.6 Optimised AIRs

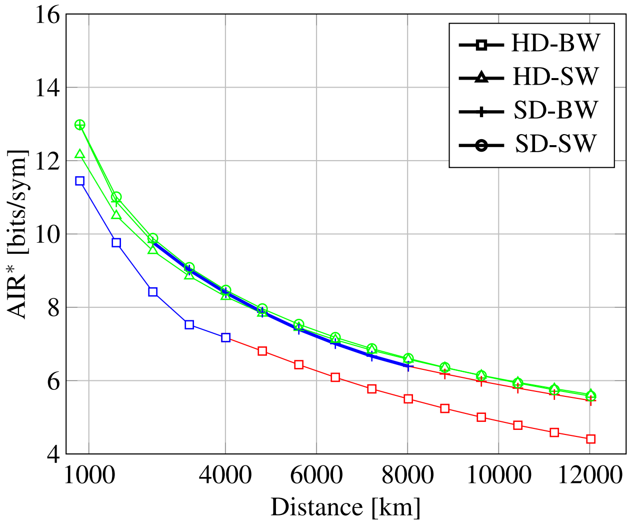
In order to highlight the performance of each decoding structure vs. the transmission distance L, in Fig. 5.9, we show the modulation format optimized AIRs, defined as

$$\mathrm{AIR}^*(L) = \max_{M\in\{16,64,256\}} \mathrm{AIR}(L,M) \tag{5.34}$$

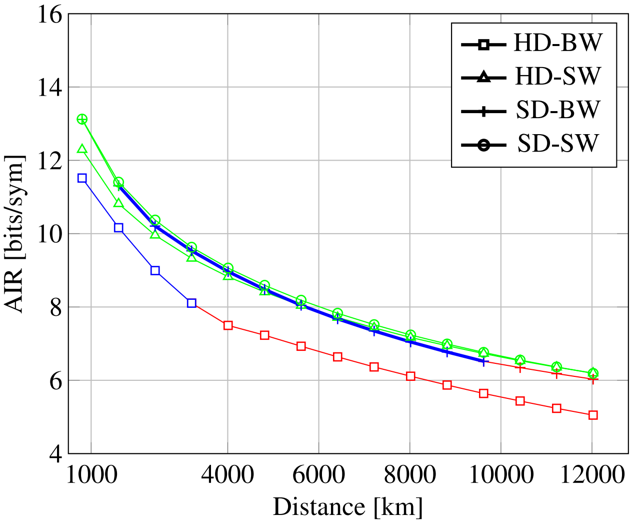
for EDC, single-channel DBP, and full-field DBP.

We observe that the set of curves shown for each equalization scheme appears as a shifted version (across the distance axis) of the other ones. This behavior is another confirmation of the fact that dispersion-unmanaged and EDFA-amplified optical fiber systems can be well described by an equivalent AWGN channel and their performance is strongly correlated to the effective SNR at the MF output. Since this SNR includes nonlinear effects as an equivalent noise source, it is improved by nonlinear compensation schemes. In the EDC case (Fig. 5.9a), except for short distances ($\leq$1000 km), HD-SW decoders have comparable performance to SD-BW and SD-SW schemes. The optimal format for both SW strategies (SD and HD) is PM-256QAM (green) at all distances, whereas for the BW schemes, PM-256QAM performs worse both for short and middle distances, where PM-64QAM (blue) is preferable, as well as in the long/ultra-long haul region, where PM-16QAM (red) is optimal. Very similar behavior is observed for single-channel DBP in Fig. 5.9b, where the optimality of PM-64QAM for BW receivers is extended to longer distances compared to their EDC counterparts.

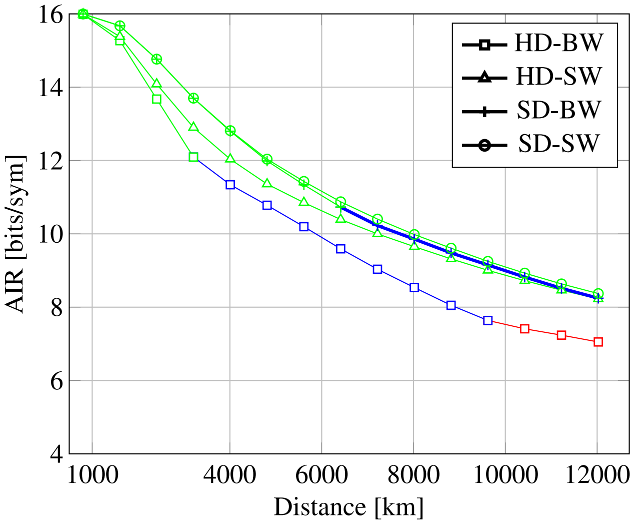
Finally, for full-field DBP (Fig. 5.9c), rates of up to 12 bits/sym can be targeted up to 5000 km, and for all decoding strategies, the optimal modulation format is PM-256QAM up to 4000 km. Also, in the ultra-long haul regime, rates above 8 bits/sym can be achieved by using PM-64QAM in conjunction with SD-BW systems without significant loss in performance compared to SD-SW or HD-SW with PM-256QAM. Overall, Fig. 5.9 also shows that HD-BW decoders perform significantly worse than all other schemes, confirming the results in [28]. Nevertheless, they can be considered as a valid low-complexity alternative for short distances or when high SNR values are available at the receiver.



(a) EDC



(b) Single-channel DBP



(c) Full-field DBP

Figure 5.9: AIRs as a function of the transmission distance for the optimal PM-MQAM format, indicated by red ($M = 16$), blue ($M = 64$) and green ($M = 256$).

5.7 AIRs of MLSD receivers

In chapter 4, the performance of MLSD receivers was characterised in terms of the uncoded BER as a function of the transmitted power for a given transmission distance. As discussed in the previous sections, the uncoded BER p can be mapped one-to-one to an AIR when HD-BW decoders are adopted. In this section, in order to better understand the significance of the BER gains shown in chapter 4, the latter is quantified in terms of AIRs.

In Fig. 5.10, the AIRs for the correlators receiver discussed in 4.3 is shown as a function of the span length L. The correlators' length is varied from 3 to 7 symbols and the MF receiver AIR is shown as a reference. The calculation of the AIRs is performed using the optimum BER at any given L and converting this value to an AIR for an HD-BW decoder using Eq. (5.26) for the specific case where an average binary transition probability p over all bit positions and transmitted bits is used. In this case, the AIR is given by the well-known binary symmetric channel BSC capacity

$$I_{\text{HD-BW}} \triangleq I(B_k;\hat{B}_k) = \log_2(M)\left[1 + p\log_2(p) + (1-p)\log_2(1-p)\right]. \tag{5.35}$$

The AIR is then expressed in bit/s using

$$\text{AIR} = I_{\text{HD-BW}} R_s \tag{5.36}$$

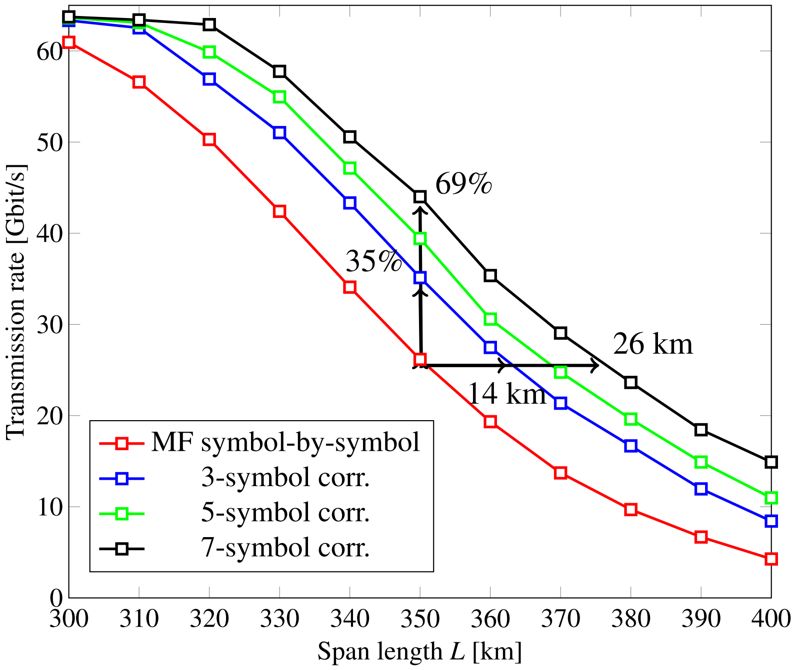


Figure 5.10: AIRs as a function of the span length L of a 32 GBaud QPSK transmission using MLSD receivers and HD-BW decoders.

where R_s=32 GBaud is the symbol rate.

The results in Fig. 5.10 show that the BER improvement observed in Fig. 4.6 translates into a substantial gain in AIRs when ML detection is performed accounting for the channel memory. For instance, the gain obtained using a 3-symbol bank of correlators compared to a MF symbol-by-symbol receiver is a 35% higher AIR at 350 km distance or a 14 km span length increase at the same reference distance. A similar beneficial effect is observed by increasing the correlators length up to 7 symbols. For such a correlator length, the AIR gain and the transmission length increase by 69% and up to 26 km, respectively.

As discussed in chapter 4, for a fixed correlators length the border effects eventually determines an optimum BER which decays as a function of L. This, in turn, causes the AIR to decay as a function of L, regardless of the transmitted power used. The alternative is represented by the MLSD receiver based on the Viterbi algorithm that was discussed in its different variants in chapter 4.

In Fig. 5.11, the AIRs are shown as functions of the span length L for the receiver schemes illustrated in Fig. 4.7, i.e. a MF followed by an MLSD receiver, and the RLPF-MLSD scheme. Both MLSD schemes considered here use a Viterbi processor with 7-symbol long states.

The black dashed line represents the reference for a receiver implementing optimum detection at every transmitted power, assuming an asympotically increasing minimum Euclidean distance between sequences. Indeed, in this case, an arbitrary low BER can be obtained and as a result the nominal bit rate $R_s \log_2 M$=64 Gbit/s is by definition

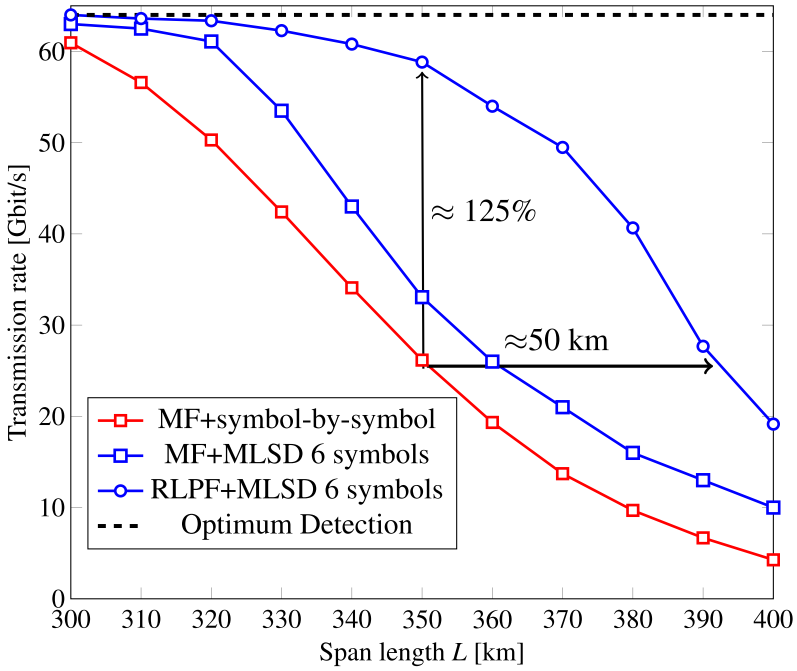


Figure 5.11: AIRs as a function of the span length (L) of a 32 GBaud QPSK transmission using MLSD receivers and HD-BW decoders.

achievable. The red curve represents the net bit rate for a MF filter receiver with detection performed on a symbol-by-symbol level. The two blue curves represent the net bit rate achievable using MLSD receivers in the two variants shown in Fig. 4.7(b) and (c). Significant rate increase for the MLSE receivers can be observed throughout the range of distances investigated. The MF-MLSD scheme achieves a gain increase in distance at 350 km compared to the MF symbol-by-symbol receiver. Interestingly, by comparing Fig. 5.10 and Fig. 5.11 it can be noticed that the MF-MLSD scheme achieves AIRs similar to the 3-symbol correlators approach, but using a longer memory. As explained in chapter 4, this similar performance can be attributed to the suboptimal statistics captured by the MF. Indeed, by using a (quasi) sufficient statistic, the RLPF-MLSD scheme is able to increase the transmission rate by 125% and the span length by 50 km with the same state length than the MF-MLSD receiver. Overall, the RLPF-MLSD receiver, shows significant AIRs improvements compared to all other schemes investigated, combining the MLSD approach to the optimised statistics extraction.

The AIRs shown in Figs. 5.10, 5.11 only applies to HD-BW decoders. Therefore improved AIRs can be obtained for instance using SD decoders. Specifically, using receivers with memory allows naturally to calculate the pdf in Eq. (5.4) which is typically done by either modifying the Viterbi algorithm in a soft-output Viterbi algorithm [55] or alternatively using the well-known Bahl, Cocke, Jelinek and Raviv or BCJR algorithm [56]. Such receivers can approach the optimal (*matched*) receiver for the channel with memory represented in Eq. (5.4) as their state length matches the memory of the channel. In this case, the channel MI (see eq. (5.7)) can be achieved and the calculation of Eq. (5.7) can be performed using the approach in [21]. This calculation is left for future work (see section 6.2).

5.8 Summary

In this chapter an analysis of the performance of optical fibre systems employing the receiver schemes discussed in chapters 3 and 4 was performed from an information-theoretic perspective.

The MI is a useful measure of the performance of a coded system and represents an upper bound on the AIRs when a given modulation format is used and optimum decoding is performed at the receiver. Conversely, the AIRs of pragmatic transceiver schemes are dictated by the specific implementation of the CM decoder. A detailed numerical study of the AIR performance for high-SE long-haul optical communication systems when these pragmatic decoders and equalization schemes, such as EDC and DBP, are employed. AIRs for MLSD schemes were also investigated for single-span systems, in the case of HD-BW FEC decoders.

The results in this chapter lead to multiple interesting conclusions on the performance of coded optical fiber communication systems using PM-MQAM modulation formats. In long-haul systems, when the equalizer enables high SNR values (through the use of full-field DBP), an SD decoder is not the only alternative to achieve high rates at long distances. On the contrary, HD nonbinary FEC schemes can, in principle, achieve similar information rates across all distances of interest. For SNR values in the low to medium range (EDC or single-channel DBP), SD decoders outperform HD ones up to medium-SE formats (PM-64QAM). However, for high-SE formats (PM-256QAM), HD-SW CM decoders can outperform SD-BW decoders. In the SD case, BW decoders do not incur significant penalties as compared to their SW counterparts, suggesting that there is no need to employ nonbinary FEC schemes. Finally, HD-BW transceivers are never desirable for high-SE systems. Nevertheless, they can represent the implementation of choice for either short-distance systems or ultra long-haul low-SE systems whenever high order modulation formats cannot be used.

In single-span systems, where nonlinear effects are the dominant cause of transmission impairment, MLSD receivers accounting for the nonlinear channel memory can significantly improve the AIRs of the system compared to the EDC case. This result, which follows from the BER results obtained in chapter 4, can be attributed to the quasi-optimality of the detection process, which avoids the generation of additional signal-ASE NLI, typical of a nonlinear zero-forcing equalisation schemes such as DBP. Thus, in principle, arbitrary high transmission rates could be achievable for every transmission distance, like for the AWGN channel case. As discussed in chapter 4, such receivers are limited by their poor complexity scaling with the receiver memory and the cardinality of the transmitted constellation. However, using only a limited amount of memory (3 symbols) already yields remarkable AIRs gains of up to 125%, and reach extensions of up to 50 km ($\approx$ +14%). Additionally, in scenarios where the linear SNR is particularly low such as the long-distance single-span transmission, the target AIRs need to lower accordingly, thus making low-cardinality modulation formats a good candidate to achieve them. This renders the implementation of (reduced-state) MLSD receivers feasible.

References

[1] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, “Replacing the soft-decision FEC limit paradigm in the design of optical communication systems,” *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, 2015.

[2] G. Liga, A. Alvarado, E. Agrell, and P. Bayvel, “Information rates of next-generation long-haul optical fiber systems using coded modulation,” *J. Lightw. Technol.*, vol. 35, no. 1, pp. 113–123, 2017.

[3] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 2006.

[6] G. Caire, G. Taricco, and E. Biglieri, “Bit-interleaved coded modulation,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[7] L. Szczecinski and A. Alvarado, *Bit-interleaved coded modulation: fundamentals, analysis and Design*. New York, NY, USA: Chichester, UK: John Wiley & Sons, 2015.

[8] A. Leven, F. Vacondio, L. Schmalen, S. ten Brink, and W. Idler, “Estimation of soft FEC performance in optical transmission experiments,” *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1547–1549, 2011.

[9] L. Schmalen, F. Buchali, A. Leven, and S. ten Brink, “A generic tool for assessing the soft-FEC performance in optical transmission experiments,” *IEEE Photon. Technol. Lett.*, vol. 24, no. 1, pp. 40–42, 2012.

[10] L. Schmalen, A. Alvarado, and R. Rios-Müller, “Predicting the performance of nonbinary forward error correction in optical transmission experiments,” in *Proc. Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA, 2016.

[11] A. Alvarado and E. Agrell, “Four-dimensional coded modulation with bit-wise decoders for future optical communications,” *J. Lightw. Technol.*, vol. 33, no. 10, pp. 1993–2003, 2015.

[12] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, “Corrections to ‘Replacing the soft-decision FEC limit paradigm in the design of optical communication systems’,” *J. Lightw. Technol.*, vol. 34, no. 2, p. 722, 2016.

[13] T. Koike-Akino, K. Kojima, D. Millar, K. Parsons, T. Yoshida, and T. Sugihara, "GMI-Maximizing constellation design with Grassmann projection for parametric shaping Pareto-efficient set of modulation and coding based on RGMI in nonlinear fiber transmissions," in *Proc. Optical Fiber Communication Conference (OFC)*, Anaheim, CA, USA, 2016.

[14] I. B. Djordjevic, M. Arabaci, and L. L. Minkov, "Next generation FEC for high-capacity communication in optical transport networks," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3518–3530, 2009.

[15] I. B. Djordjevic, M. Cvijetic, L. Xu, and T. Wang, "Using LDPC-coded modulation and coherent detection for ultra highspeed optical transmission," *J. Lightw. Technol.*, vol. 25, no. 11, pp. 3619–3625, 2007.

[16] B. P. Smith and F. R. Kschischang, "Future Prospects for FEC in Fiber-Optic Communications," *IEEE J. Quantum Electron.*, vol. 16, no. 5, pp. 1245–1257, 2010.

[17] "Forward error correction for high bit-rate dwdm submarine systems," ITU-T, Recommendation G.975.1, 2004.

[18] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, 2012.

[19] C. Häger, H. D. Pfister, A. Graell i Amat, F. Brännström, and E. Agrell, "A deterministic construction and density evolution analysis for generalized product codes," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, 2016.

[20] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.

[21] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

[22] G. Colavolpe, T. Foggi, A. Modenini, and A. Piemontese, "Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference," *Opt. Express*, vol. 19, no. 27, pp. 26 600–26 609, 2011.

[23] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, 2013.

[24] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, 2010.

[25] R. Maher, A. Alvarado, D. Lavery, and P. Bayvel, "Increasing the information rates of optical communications via coded modulation: a study of transceiver performance," *Sci. Rep.*, vol. 6, pp. 1–10, 2016.

[26] R. Rios-Müller, J. Renaudier, L. Schmalen, and G. Charlet, "Joint coding rate and modulation format optimization for 8QAM constellations using BICM mutual information," in *Proc. Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, USA, 2015.

[27] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments," *J. Lightw. Technol.*, vol. 34, no. 9, pp. 2256–2266, 2016.

[28] T. Fehenberger, A. Alvarado, P. Bayvel, and N. Hanik, "On achievable rates for long-haul fiber-optic communications," *Opt. Express*, vol. 23, no. 7, pp. 9183–9191, 2015.

[29] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar. 1953.

[30] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, 2004.

[31] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, 1984.

[32] F. R. Kschischang and S. Pasupathy, "Optimal non uniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, 1993.

[33] R. F. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York, NY, USA: Wiley, 2005.

[34] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. B. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2407–2410, 2014.

[35] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: an experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, 2016.

[36] T. Fehenberger, D. Lavery, R. Maher, A. Alvarado, P. Bayvel, and N. Hanik, “Sensitivity gains by mismatched probabilistic shaping for optical communication systems,” *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 786–789, 2016.

[37] E. Agrell, A. Alvarado, and F. R. Kschischang, “Implications of information theory in optical fibre communications,” vol. 374, no. 2062, pp. 1–18, 2016.

[38] G. Kramer, M. I. Yousefi, and F. R. Kschischang, “Upper bound on the capacity of a cascade of nonlinear and noisy channels,” in *Information Theory Workshop (ITW)*, Jerusalem, Israel, 2015.

[39] G. P. Agrawal, *Nonlinear Fiber Optics*, 3rd ed. San Diego, CA, USA: Academic, 2001.

[40] G. Liga, A. Alvarado, E. Agrell, M. Secondini, R. I. Killey, and P. Bayvel, “Optimum detection in presence of nonlinear distortions with memory,” in *Proc. European Conference on Optical Communication (ECOC)*, Valencia, Spain, 2015.

[41] N. V. Irukulapati, H. Wymeersch, P. Johannisson, and E. Agrell, “Stochastic digital backpropagation,” *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3956–3968, 2014.

[42] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “Properties of nonlinear noise in long, dispersion-uncompensated fiber links,” *Opt. Express*, vol. 21, no. 22, pp. 25 685–25 699, 2013.

[43] T. Fehenberger, M. P. Yankov, L. Barletta, and N. Hanik, “Compensation of XPM interference by blind tracking of the nonlinear phase in WDM systems with QAM input,” in *Proc. European Conference on Optical Communication (ECOC)*, Valencia, Spain, 2015.

[44] S. Verdú and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[45] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, “New bounds on the capacity of the nonlinear fiber-optic channel,” *Opt. Letters*, vol. 39, no. 2, pp. 398–401, 2014.

[46] D. Marsella, M. Secondini, and E. Forestieri, “Maximum likelihood sequence detection for mitigating nonlinear effects,” *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, 2014.

[47] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, “Capacity of a nonlinear optical channel with finite memory,” *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2862–2876, 2014.

[48] A. Martinez, L. Peng, A. Alvarado, and A. Guillén i Fàbregas, "Improved information rates for bit-interleaved coded modulation," in *IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, 2015.

[49] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge, UK: Cambridge University Press, 2014.

[50] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, 2014.

[51] A. Carena, V. Curri, G. Bosco, P. Poggiolini, and F. Forghieri, "Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1524–1539, 2012.

[52] J. Wolfowitz, "Memory increases capacity," vol. 11, no. 4, pp. 423–428, 1967.

[53] R. E. Caflisch, "Monte Carlo and quasi-Monte Carlo methods," *Acta Numer.*, vol. 7, pp. 1–49, 1998.

[54] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology and Techniques*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 2000.

[55] J. Hagenauer and P. Hoeher, "A viterbi algorithm with soft-decision outputs and its applications," in *Global Telecommunications Conference and Exhibition (GLOBECOM)*, Dallas, TX, USA, 1989.

[56] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

6

Conclusions and future work

6.1 Conclusions

Optical fibres systems, as they are currently designed, are rapidly approaching a saturation of their transmission resources. The potential to maximise transmission rates up to their fundamental limits strongly relies on the design of new digital receivers matching the properties of the nonlinear fibre channel. Such receivers will either aim to equalise the received signal from channel impairments or they will adapt to the property of the channel by optimising data detection. The research work described in this thesis investigated two specific instances of these two alternative receiver strategies: digital back-propagation (DBP) and maximum-likelihood sequence detection (MLSD).

For the first time, the performance of multi-channel DBP was analysed in the context of long-haul wideband transmission systems comparing analytical expressions and numerical results obtained through the SSFM. Closed-forms expressions for the estimation of the DBP signal-to-noise ratio (SNR) gains were derived using the Gaussian-noise model. Good agreement was found for all the regimes of interest for optical fibre systems, including large bandwidth transmission scenarios. This enables a quick assessment of the DBP SNR performance without the need to run computationally intense numerical simulations.

DBP benefit was found to be, in general, substantial compared to receivers compensating only for chromatic dispersion, with an SNR gain in excess of 10 dB for transmission distances around 1000 km. The dependence of DBP gain on the compensation bandwidth was also highlighted, showing that SNR increases faster as DBP compensation bandwidth approaches the entire transmitted bandwidth.

Although studying ideal DBP benefits is key to understand the full potential of this algorithm, real-world demonstrations often show significantly lower gains. This is due to practical constraints in the algorithm implementation or in the system characteristics. Specifically, limited digital-signal processing (DSP) computational resources and polarisation-mode dispersion (PMD) were identified as a major cause of performance impairment. Ideal implementation of DBP requires a high computational complexity, which is impractical for the current state-of-art of the electronics. As a result, reduced DSP complexity is mandatory for real-world systems. However, the increase of DBP compensation bandwidth required a higher number of DBP iterations (steps) to guarantee ideal performance or a contained penalty from the ideal performance. For instance, for a 5-channel transmission (165 GHz optical bandwidth) it was found that full-field DBP outperformed 3-channel DBP only if more than 20 steps/span were used, and single-channel DBP if more than 18 steps/span were used. Below 10 steps/span, single-channel DBP gave the best performance.

PMD impact on DBP was numerically analysed using a novel method which merges the logarithmic split-step Fourier method and the conventional wave-plate Monte Carlo approach. This method allowed to efficiently characterise PMD effects in wideband nonlinearity compensation scenarios, which would otherwise be prohibitive in terms of complexity. It was shown that using the conventional DBP algorithm in the presence of typical fibre PMD (0.1 ps/$\sqrt{\text{km}}$) prevents the achievement of the gains predicted by ideal full nonlinearity compensation. Furthermore, unlike the ideal case, PMD results in a saturation of the SNR as a function of the compensation bandwidth. As an example, for a 1 THz transmission scenario and a PMD parameter of 0.1 ps/$\sqrt{\text{km}}$, the SNR was shown to reach a value 0.8 dB below the full-field DBP gain for a compensation bandwidth of between 5 and 7 channels.

These results show that DBP current performance bottleneck lies mainly on its complexity, but that in order to unlock its full potential the PMD issue should be tackled by, for instance, designing adaptive DBP receivers able to cleverly guess the arrangement of the PMD sections along the transmission fibre. This topic is discussed in section 6.2.

As an alternative to DBP, improved detection schemes for optical fibre nonlinearity were investigated. Detection theory was applied to unrepeatered fibre systems to devise an optimal receiver strategy. Optical fibre nonlinear distortions feature a significant amount of memory even after chromatic dispersion compensation is applied. As a result, the single-span fibre channel can be modelled as a nonlinear ISI channel with additive white Gaussian noise (AWGN). In this thesis, receivers implementing detection strategies which account for the channel memory, were demonstrated, for the first time, to markedly outperform a conventional matched-filter symbol-by-symbol receiver.

Two alternative instances of maximum likelihood sequence detection (MLSD) receivers were numerically evaluated in the case of a quaternary phase shift keying (QPSK) modulation format. In the first case, a bank-of-correlators was used to detect waveforms corresponding to sequences of symbols. Despite the correlators' length extending over a small number of symbols compared to the expected channel memory, such receiver was demonstrated to achieve a BER 2.5 times lower than a matched filter receiver when using a 7-symbol long correlators.

In the second implementation, longer sequences were detected using a Viterbi processor. This was done to avoid inter-sequence interference while preserving a reasonable detection complexity. The role of sufficient statistics in the detection process was highlighted by assessing the performance of two variants of the Viterbi receiver: the first one using samples at the output of a matched filter while the second one using samples after a rectangular low pass filter (RLPF) which encloses the entire received signal bandwidth. The latter scheme, which employs a sufficient statistic, approaches the optimal detection strategy when the number of states in the Viterbi trellis matches the channel memory. The (quasi)-optimum Viterbi receiver using 6-symbol states outperformed both conventional matched filter detection and MLSD-Viterbi detection using a matched filter with a sampler at 1 Sa/sym and the same number of states. In particular, the RLPF-Viterbi receiver achieved a BER 15 times lower than the matched filter symbol-by-symbol receiver and 10 times lower than a matched filter-Viterbi receiver. Furthermore, the RLPF-Viterbi scheme closely approaches the AWGN performance of a QPSK modulation format up to a value of transmitted power of approximately 22 dBm, and a monotonically decreasing BER over the entire range of investigated transmitted powers. The BER floor observed at high transmitted powers can be explained due to the insufficient amount of memory used at the receiver. Further increasing the receiver memory is a challenging task which is ultimately discouraged by the exponential scaling of the detection complexity.

Finally, both DBP and MLSD performance was analysed in the coded regime using information theory to quantify achievable information rates (AIRs). In particular, an extensive comparative study was for the first time presented on AIRs for pragmatic coded modulation schemes, used in combination with DBP and MLSD. The coded modulation schemes studied involve three different cardinalities of the quadrature-amplitude modulation (QAM) format and four different demapper/decoder structures such as: soft-decision (SD) bit-wise (BW) (or binary), and symbol-wise (SW) (or nonbinary), and hard-decision (HD) binary and nonbinary .

The results for long-haul transmission show the AIRs gains of more than 2 bit/sym can be obtained using full-field DBP, as opposed to an electronic dispersion compensation (EDC) receiver. Single-channel DBP was instead shown to have smaller AIRs gains for a fixed transmission distance. However, fixing a target AIR, single-channel DBP was shown to extend the transmission distance by up to 1000 km. Furthermore, it was shown that when full-field DBP used SD decoders significantly outperformed HD decoders due to higher attainable SNR values. For lower SNRs, i.e. when either EDC or single-channel DBP are adopted, HD-SW decoders can closely approach the performance of SD receivers, particularly for high cardinality modulation formats (64QAM, 256QAM), and in some cases SD-BW receivers can be outperformed.

AIRs of MLSD schemes were also studied for QPSK and HD-BW decoders. For a bank-of-correlators receiver, transmission rate increases of 35% and 69% were achieved at a 350 km distance with 3-symbol and 7-symbol correlators, respectively. Alternatively, a span length increase of 14 km and 26 km was observed at the same distance for the two receiver implementations, respectively. Employing a Viterbi processor should bring an improvement by avoiding most of the inter-sequence interference. However, in the case where the Viterbi processor was fed with samples at the output of a matched filter and used 6-symbol states, a similar performance to the bank-of-correlators was observed. Indeed, a 50% increase in transmission rate and 20 km reach extension at 350 km compared to a conventional matched filter, symbol-by-symbol receiver is achieved. On the contrary, the quasi-optimum receiver strategy RLPF/Viterbi transmission rate gain was increased up to 125%, with a reach extension of 50 km ($\approx$ 15%) at the same transmission distance, compared to a matched filter/symbol-by-symbol receiver. This large improvement can be attributed to the bandwidth expansion experienced by signal at high transmitted powers. As a result, a matched filter followed by a sampler at 1 Sa/sym is unable to collect a sufficient statistic, which can be instead obtained by fast-sampling the received waveform. Collecting a sufficient statistic can be crucial for the error rate performance of a detector, as was demonstrated, for the first time, in the work described in this thesis.

The results described in this thesis reflect the detailed investigation of the performance of two main alternative strategies for the design of next-generation digital receivers, aimed to maximise transmission rates in optical fibre communication systems. Both the challenges and potential benefits have been highlighted, as well as theoretical insights that should, hopefully, guide further development of these techniques. Future research efforts in this area are supported by recent theoretical results, for example [1, 2], which showed that fundamental limit in optical fibre transmission capabilities is still far from being reached, and thus, fibre exhaust is some way away. Future work is most likely to focus on techniques to approach these limit.

6.2 Future work

In this section, some topics for further work are discussed, that naturally follow on from the research described in this thesis.

6.2.1 Towards PMD-aware DBP receivers

In this thesis, PMD was shown to be a major source of impairment of the performance of DBP. The conventional DBP algorithm disregards polarisation effects in the fibre preventing the potential full cancellation of the NLI. The residual NLI after DBP is a stochastic term depending on the specific PMD realisation and its entity can be reduced by adopting modified DBP approaches [3, 4]. In particular, initial studies were conducted in [3, 5, 6], in collaboration with Czegledi et al., where it was shown that by assuming a particular distribution of PMD sections with the only knowledge of the total Jones matrix of the fibre link, DBP SNR performance can, (on average), be improved. This works proved that PMD effects, once considered to fundamental bottleneck for nonlinearity compensation, can be at least mitigated.

Future research work on this topic could be then conducted along two main tracks:

i) Modelling of the nonlinear interaction between PMD and DBP

ii) Adaptive DBP schemes with tracking of the fibre polarisation evolution.

In order to gain full insight on how PMD interacts with fibre nonlinearity and how to make DBP more robust to PMD, research efforts could be devoted to a novel analytical model describing such interactions. The model in [7] could be extended to include the effect of a partial NLC bandwidth and to highlight, possibly in an approximated closed-form, the relationship between transmitted optical bandwidth, NLC bandwidth, PMD parameter and average DGD.

Once analytical relationships are derived, the design of an adaptive DBP receiver to track the fibre polarisation evolution should be considered. The fibre PMD sections represent a stochastic process which evolves slowly with respect to the characteristic transmission times. Thus, the design of a receiver, which estimates the distribution of these sections, appears as a realistic task. A way to implement such receiver would be via the optimisation of a cost function that achieves a global maximum (or minimum) corresponding to the perfect alignment of the polarisation sections. Such optimisation problem features as many optimisation variables as the number of PMD sections employed in the DBP algorithm. Thus, the main obstacle is represented by the complexity in evaluating this function. This problem could be addressed with the help of the results obtained from the analytical modelling of PMD, which could hopefully provide a cost function simple to compute and to optimise, such as a linear function on the optimisation variables. An adaptive DBP algorithm converging to the correct alignment of the PMD sections will perform ideal nonlinearity compensation as in the case of total absence of PMD. Extensive numerical simulations are also needed to assess the performance of such a novel algorithm.

6.2.2 Improved detection in the presence of signal-ASE noise

The detection schemes studied in this thesis were optimised for the single-channel, single-span optical fibre channel. Future work could focus on extending these strategies to the more general cases of multi-channel, multi-span systems. In such systems, in order to ensure a good performance detector, the effects of intra-channel and signal-ASE NLI need to be considered.

The effect of inter-channel NLI can be factored in by including, in the finite-state channel model of the fibre, the additional effect of transmitted symbols on channels surrounding the channel of interest. A receiver fully accounting for all channel states would then be almost unfeasible, due to complexity reasons. However, future studies on this topic could also involve possible options to reduce the number of states in the receiver while preserving acceptable performance. This can be done, for instance, by studying the topology of the received signals space when conditioned to a given channel state, and particularly their minimum Euclidean distance.

Optimising detection strategies in multi-span scenarios requires the knowledge of the stochastic properties of the signal-ASE NLI process. Results in [8] and [9] have shown how better detection strategies can be devised by accounting for signal-ASE NLI process. However, the gains compared to conventional matched filter symbol-by-symbol receivers have been demonstrated to be marginal for dispersion-unmanaged systems.

Finally, improving the performance of suboptimal detection strategies such as the matched filter, symbol-by-symbol detection is a task to be considered for future work. Indeed, different linear filters could in principle perform better than the matched filter in terms of minimum error probability. Despite the suboptimality of this approach compared to MLSD schemes, possible benefits compared to a matched filter receiver would be of particular interest as a low-complexity structure would be preserved.

6.2.3 Improved AIRs for optical fibre transmission

It seems clear that future research on exploring the limits of optical fibre communications will need to deal with channels and receivers with memory. In order to obtain tighter lower bounds on the capacity of the fibre, channel memory must be considered for the computation of the MI. This requires auxiliary channels with memory and, in general, intensive computational efforts. Future research work should be devoted to compute

MI on channel models with memory which represents an AIR for receivers matching the memory of the channel. This problem can be approached using finite-state machine models for the optical fibre channel which can simplify the calculation of the MI [10]. One such model for the single-span case can be obtained through numerical simulations, as the properties of the noise in the channel are analytically known (AWGN). Channel models with memory in the multi-span system case should also be investigated.

Potentially improved AIRs compared to conventional memoryless AIRs, would lead to testing actual implementations of demappers (with memory) and encoder/decoder pairs in order to verify their potential to achieve such rates. Demappers accounting for the channel memory are for instance based on the BCJR [11] or the soft-output Viterbi algorithm [12]. Once the demapper is matched to the channel coding design tailored to the properties of the channel could represent an interesting research direction leading to potentially significant information rate gains compared to traditional coding techniques (e.g. LDPC, product codes, etc.). The main challenges of this research area typically consists of finding a good compromise between coding gains and decoding complexity. Again, the mismatched decoding principles might be of help in devising good solutions for this problem.

References

[1] E. Agrell, “Conditions for a monotonic channel capacity,” *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 738–748, 2015.

[2] M. Secondini and E. Forestieri, “Scope and limitations of the nonlinear Shannon limit,” *J. Lightw. Technol.*, vol. 35, no. 4, pp. 893–902, Feb 2017.

[3] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, “Polarization-mode dispersion aware digital backpropagation,” in *Proc. European Conference on Optical Communication (ECOC)*, Dusseldorf, Germany, 2016.

[4] K. Goroshko, H. Louchet, and A. Richter, “Overcoming performance limitations of digital back propagation due to polarization mode dispersion,” in *International Conference on Transparent Optical Networks*, Trento, Italy, 2016.

[5] C. B. Czegledi, G. Liga, M. Karlsson, and E. Agrell, “Modified digital backpropagation accounting for polarization-mode dispersion,” in *Proc. of Optical Fiber Communication Conference (OFC)*, Los Angeles, CA, USA, 2017.

[6] C. B. Czegledi, G. Liga, D. Lavery, M. Karlsson, E. Agrell, S. J. Savory, and P. Bayvel, “Polarization-mode dispersion aware digital backpropagation,” *Opt. Express*, vol. 25, no. 3, pp. 1903–1915, 2017.

[7] Y. Gao, J. H. Ke, J. C. Cartledge, K. P. Zhong, and S. S.-H. Yam, “Implication of parameter values on low-pass filter assisted digital back propagation for DP 16-QAM,” *IEEE Photon. Technol. Lett.*, vol. 25, no. 10, pp. 917–920, 2013.

[8] N. V. Irukulapati, D. Marsella, P. Johannisson, E. Agrell, M. Secondini, and H. Wymeersch, “Stochastic digital backpropagation with residual memory compensation,” *J. Lightw. Technol.*, vol. 34, no. 2, pp. 566–572, 2016.

[9] D. Marsella, M. Secondini, and E. Forestieri, “Maximum likelihood sequence detection for mitigating nonlinear effects,” *J. Lightw. Technol.*, vol. 32, no. 5, pp. 908–916, 2014.

[10] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, 2006.

[11] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[12] M. Fossorier, F. Burkert, S. L. S. Lin, and J. Hagenauer, “On the equivalence between SOVA and max-log-MAP decodings,” *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 137–139, 1998.

A

Derivation of signal-ASE nonlinear interference term

Let us assume $S(f)$ and $N(f)$ are the power spectral densities of the propagating signal and ASE noise in a fibre span, respectively. The total power spectral density of the optical field can be then expressed as

$$G(f) = S(f) + N(f). \tag{A.1}$$

We also rewrite each term in (A.1) in terms of their normalised PSDs $s(f)$ and $n(f)$ as

$$S(f) = \frac{P}{R_s} s(f) \tag{A.2}$$

$$N(f) = \frac{P_{\text{ASE}}}{R_s} n(f) \tag{A.3}$$

where P is the transmit power of one channel, P_{ASE} is the power of the ASE noise over the channel bandwidth, and R_s is the symbol rate which we assume for simplicity to be exactly equal to the channel bandwidth (Nyquist rate).

Replacing (A.1) in (2.63) and expanding we obtain

$$
\begin{aligned}
G_{\mathrm{NLI}}(f) = \frac{16}{27}\gamma^2 \Bigg[&\frac{P^3}{R_s^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(f_1)s(f_2)s(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2 \\
&+\frac{P^2P_{\mathrm{ASE}}}{R_s^3}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1)s(f_2)s(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2\right. \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(f_1)n(f_2)s(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2 \\
&\left.+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(f_1)s(f_2)n(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2\right) \qquad \text{(A.4)} \\
&+\frac{PP_{\mathrm{ASE}}^2}{R_s^3}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} s(f_1)n(f_2)n(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2\right. \\
&+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1)s(f_2)n(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2 \\
&\left.+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1)n(f_2)s(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2\right) \\
&+\frac{P_{\mathrm{ASE}}^3}{R_s^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1)n(f_2)n(f_1+f_2-f)\theta(f_1,f_2,f)df_1\,df_2\Bigg].
\end{aligned}
$$

The integral multiplying $\frac{P^3}{R_s^3}$ is the signal-signal NLI term, whereas the residual terms involve both signal and noise. Under the assumption that $P \gg P_{\mathrm{ASE}}$, all the terms proportional to PP_{ASE}^2 or P_{ASE}^3 can be neglected compared to the terms proportional to P^2P_{ASE}. It can be noticed that three different integrals contribute to the signal-ASE NLI term proportional to P^2P_{ASE}. We denote these double-integrals in their appearing order as $I_1(f)$, $I_2(f)$ and $I_3(f)$. Since (see (2.64))

$$\theta(f_1,f_2,f) = \theta((f_1-f)^2(f_2-f)^2) \qquad \text{(A.5)}$$

clearly

$$\theta(f_1,f_2,f) = \theta(f_2,f_1,f) \qquad \forall\, (f_1,f_2,f) \in \mathbb{R}_s^3 \qquad \text{(A.6)}$$

hence $I_1(f) = I_2(f) \quad \forall f \in \mathbb{R}$. In order to check whether $I_3(f)$ is equal to the previous two terms, we apply the substitution

$$f_1 + f_2 - f = f_1^{'} \qquad \text{(A.7)}$$

which yields

$$f_1 = f_1^{'} + f - f_2. \qquad \text{(A.8)}$$

As a result, the integral $I_3(f)$ (A.4) becomes

$$I_3(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1')s(f_2)s(f_1^{'}-f_2+f)\theta((f_1'-f_2)^2(f_2-f)^2)df_1'\,df_2 \qquad \text{(A.9)}$$

which is in general different from

$$I_1(f) = I_2(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(f_1)s(f_2)s(f_1+f_2-f)\theta((f_1-f)^2(f_2-f)^2)df_1\,df_2. \tag{A.10}$$

However, if we consider the specific case

$$s(f) = rect\left(\frac{f}{B}\right) \tag{A.11}$$

$$n(f) = rect\left(\frac{f}{B_n}\right) \tag{A.12}$$

where B, B_n are the WDM and the ASE noise bandwidth respectively, the integrals become

$$I_1(f) = I_2(f) = \iint\limits_{D_1(f)} \theta((f_1-f)^2(f_2-f)^2)df_1\,df_2, \tag{A.13}$$

$$I_3(f) = \iint\limits_{D_2(f)} \theta((f_1-f)^2(f_2-f)^2)df_1\,df_2. \tag{A.14}$$

where

$$\begin{aligned} D_1(f) = \{(f_1,f_2) : f-B/2 \le f_1+f_2 \le f+B/2\} \\ \cap\{-B_n/2 \le f_1 \le B_n/2\}\cap\{-B/2 \le f_2 \le B/2\}, \end{aligned} \tag{A.15}$$

$$\begin{aligned} D_2(f) = \{(f_1,f_2) : f-B_n/2 \le f_1+f_2 \le f+B_n/2\} \\ \cap\{-B/2 \le f_1 \le B/2\}\cap\{-B/2 \le f_2 \le B/2\}. \end{aligned} \tag{A.16}$$

The two different integration domains D_1 and D_2 are illustrated in Fig. A.1 for $f = 0$. Clearly, as the two regions are different, $I_1(f) = I_2(f) \neq I_3(f)\ \ \forall f \in [-R_s/2, R_s/2]$ and, thus, follows

$$\sigma^2_{SN1} = \sigma^2_{SN2} = \int_{-R_s/2}^{R_s/2} I_2(f)df \neq \sigma^2_{SN3} = \int_{-R_s/2}^{R_s/2} I_3(f)df. \tag{A.17}$$

The overall signal-ASE NLI power is therefore given by

$$\sigma^2_{SN} = 2\sigma^2_{SN1} + \sigma^2_{SN3} \tag{A.18}$$

However, in the specific case where $B = B_n$

$$\sigma^2_{SN} = 3\sigma^2_{SN1} = 3\eta_{SS}P_{\text{ASE}}P^2 \tag{A.19}$$

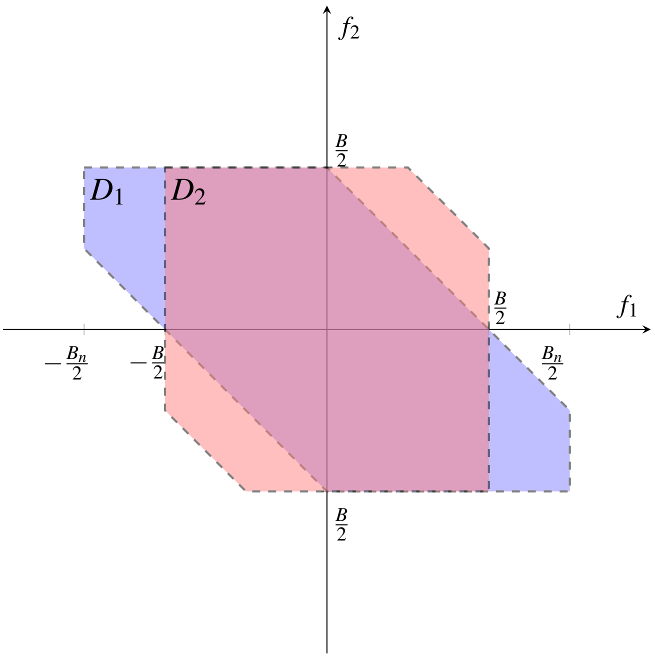


Figure A.1: Domains of integration D_1 (blue) and D_2 (red) for $f = 0$.

where

$$\eta_{SS} = \frac{1}{R_s^3}\int_{-R_s/2}^{R_s/2}\int_{-B/2}^{B/2}\int_{-B/2}^{B/2} \theta((f_1 - f)^2(f_2 - f)^2)\,df_1\,df_2\,df. \tag{A.20}$$

It is worth observing that for relatively large values of B and B_n, such as in the case of C-band or C+L-band transmission, the differences between the two integrals (A.13) and (A.14) are extremely small, due to the rapid decay of the modulus of $\theta(f_1, f_2, f)$ for typical dispersion values. Moreover, $B = B_n$ is representative of a typical transmission scenario, where WDM signal and ASE noise have the same bandwidth due to the in-line optical filtering. Therefore, the approximation $\sigma_{SN}^2 \approx 3\eta_{SS}P_{\text{ASE}}P^2$ yields, in most cases, a negligible, if not null, error.

B

Viterbi algorithm for the optical fibre channel

MLSD detection aims to find the most likely sequence of N transmitted symbols $\mathbf{x} = [x_0, x_1, ..., x_{N-1}]$ given a certain vector of observations $\mathbf{r} = [r_0, r_1, ..., r_{N-1}]$. This corresponds to the rule in (4.12).

Let us assume the waveform channel in 4.2b. For this channel the received noisy waveform can be written as

$$r(t) = \sum_{k=0}^{\infty} s(t, \sigma_{k-1}, \sigma_k) + n(t) \tag{B.1}$$

where $\sigma_k \in \{S_i\}_{i=1}^{M}$ is the channel state at time k, defined as in (4.17), $s(t, \sigma_{k-1}, \sigma_k) \in \{s_1(t - kT_s), s_2(t - kT_s), ..., s_M(t - kT_s)\}$ is a set of M waveforms of duration T_s associated with the state transition (σ_{k-1}, σ_k), and $n(t)$ is a white Gaussian noise process at least over an arbitrary bandwidth B. The channel state at time k is defined as in (4.17).

Let us assume that the sequence of samples $\mathbf{r}$ is obtained by applying the receiver in Fig. 4.7c to the channel in Fig. 4.2c, where the rectangular filter has a (double-sided) bandwidth B and the sampling rate $R_s = B$.

In this case, the output samples can be written as

$$\mathbf{r}_k = \mathbf{y}_k(\sigma_{k-1}, \sigma_k) + \mathbf{n}_k \ \forall \ k = 0, 1, ..., N-1 \tag{B.2}$$

where $\mathbf{r}_k$ and $\mathbf{n}_k$ are vectors of N_s samples defined as

$$\mathbf{r}_k \triangleq [r_{kN_s+1}, r_{kN_s+2}, ..., r_{(k+1)N_s}], \tag{B.3}$$

$$\mathbf{n}_k \triangleq [n_{kN_s+1}, n_{kN_s+2}, ..., n_{(k+1)N_s}] \quad \forall k = 0, 1, ..., N-1. \tag{B.4}$$

The sequences $\mathbf{r}$ and $\mathbf{n}$ can be clearly expressed as the concatenations

$$\mathbf{r} = [\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{N-1}], \tag{B.5}$$

$$\mathbf{n} = [\mathbf{n}_0, \mathbf{n}_1, ..., \mathbf{n}_{N-1}]. \tag{B.6}$$

It is worth noticing that the noise samples n_k are jointly Gaussian complex random variables with diagonal covariance matrix and same variance equal to N_0. For this reason and by looking at (B.2) the sequence likelihood can be factorised as

$$p(\mathbf{r}|\mathbf{x}) = \prod_{k=1}^{N_s} p(\mathbf{r}_k|\sigma_{k-1}, \sigma_k) \tag{B.7}$$

where

$$p(\mathbf{r}_k|\sigma_{k-1}, \sigma_k) = \frac{1}{(\pi N_0)^{N_s}} e^{-\frac{||\mathbf{r}_k - \mathbf{y}_k(\sigma_{k-1}, \sigma_k)||^2}{N_o}}. \tag{B.8}$$

The Viterbi algorithm then minimises the sequence log-likelihood

$$\Lambda_N = \sum_{k=1}^{N} \lambda_k(\sigma_{k-1}, \sigma_k) = \sum_{k=0}^{N-1} ||\mathbf{r}_k - \mathbf{y}_k(\sigma_{k-1}, \sigma_k)||^2 \tag{B.9}$$

using the following procedure:

1. *Channel estimation*
   Estimation of the set of vectors $\mathbf{y}_k(\sigma_{k-1}, \sigma_k)$ $\forall$ $\sigma_{k-1}, \sigma_k \in \{S_i\}_{i=1}^{M}$ using the transmission of a (long) known sequence of a data.

2. *Algorithm*

   - Initial state σ_0 initialised to arbitrary value in the state space $\{S_i\}_{i=1}^{M^{2m}}$.

   - Initial value of the survivor sequence metric Λ_0 is set to 0.

   - Time-index iteration: for $k = 0, 1, ..., N-1$.
     $\Lambda_{k+1} = 0$.
     Trellis k-th section iteration: for $s = 1, 2, ..., M^{2m}$.

  - Calculate for each *survivor* at state $\sigma_k = S_s$ the updated metric $\Lambda_{temp} = \Lambda_k + \lambda_k(\sigma_{k-1}, \sigma_k)$.
  - If $\Lambda_{k+1} < \Lambda_{temp}$ then $\Lambda_{k+1} \leftarrow \Lambda_{temp}$ and store relative sequence.
- Choose sequence of states $\hat{\boldsymbol{\sigma}} = [\hat{\sigma}_0, \hat{\sigma}_1, ..., \hat{\sigma}_{N-1}]$ that minimises Λ_N.

- Map $\hat{\boldsymbol{\sigma}}$ into corresponding sequence of input symbols $\hat{\mathbf{x}} = [\hat{x}_0, \hat{x}_1, ..., \hat{x}_{N-1}]$.

C

Acronyms

ADC	Analog-to-Digital Converter
AIR	Achievable Information Rate
ASE	Amplified Spontaneous Emission
AWGN	Additive White Gaussian Noise
BC	Bank-of-Correlators
BCH	Bose-Chaudhuri-Hocquenghem
BCJR	Bahl, Cocke, Jelinek and Raviv
BER	Bit Error Rate
BPSK	Binary Phase-Shift Keying
BW	Bit-Wise
CD	Chromatic Dispersion
CM	Coded Modulation
CMA	Constant Modulus Algorithm
CNLSE	Coupled Non-Linear Schrödinger Equation

CPE	Carrier Phase Estimation
CW	Continuous-Wave
DAC	Digital-to-Analog Converter
DBP	Digital Back-Propagation
DGD	Differential Group Delay
DPSK	Differential Phase-Shift Keying
DSP	Digital Signal Processing
DP	Dual-Polarisation
DP-QPSK	Dual-Polarised Quadrature Phase-Shift Keying
DWDM	Dense Wavelength Division Multiplexing
EDC	Electronic Dispersion Compensation
EDFA	Erbium-Doped Fibre Amplifier
FEC	Forward Error Correction
FIR	Finite Impulse Response
FWM	Four-Wave Mixing
GMI	Generalised Mutual Information
GV	Group Velocity
GVD	Group Velocity Dispersion
HD	Hard-Decision
IMDD	Intensity-Modulation Direct-Detection
IFFT	Inverse Fast Fourier Transform
IQ	In-phase Quadrature
ISI	Inter-Symbol Interference
ITU	International Telecommunication Union
LL	Log-Likelihood

LDPC	Low-Density Parity-Check Code
LMS	Least Mean Square
LO	Local Oscillator
LPF	Low-Pass Filter
LR	Line Rate
M-PSK	M-ary Phase-Shift Keying
M-QAM	M-ary Quadrature Amplitude Modulation
MAP	Maximum A-posteriori Probability
MF	Matched Filter
MI	Mutual Information
MLSD	Maximum Likelihood Sequence Detection
MZM	Mach-Zehnder Modulator
NRZ	Non-Return-to-Zero
NLC	Nonlinearity Compensation
NLI	Non-Linear Interference
NLSE	Non-Linear Schrödinger Equation
OOK	On-Off Keying
OSA	Optical Spectrum Analyser
OSNR	Optical Signal-to-Noise Ratio
OFDM	Optical Frequency Division Multiplexing
OTN	Optical Transport Network
PD	Photodiode
PDF	Probability Density Function
PM	Polarisation Multiplexed
PMD	Polarisation-Mode Dispersion

PRBS	Pseudo-Random Binary Sequence
PSD	Power Spectral Density
PSP	Principal State of Polarisation
QAM	Quadrature-Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
RLPF	Rectangular Low-Pass Filter
RRC	Root-Raised Cosine
RZ	Return-to-Zero
SD	Soft-Decision
SE	Spectral-Efficiency
SNR	Signal-to-Noise Ratio
SOP	State Of Polarisation
SPM	Self-Phase Modulation
SSFM	Split-Step Fourier Method
SSMF	Standard Single Mode Fibre
SW	Symbol-Wise
WDM	Wavelength Division Multiplexing
XPM	Cross-Phase Modulation