Study of Non-linear Optical Fibre Communications

Xianhe Yangzhang

A dissertation submitted to
University College London
for the degree of
Doctor of Philosophy

Department of Electronic & Electrical Engineering
University College London

September 24, 2020

I, Xianhe Yangzhang, 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 in the work.

Abstract

THE research presented in this thesis explores the use of non-linear Fourier transform (NFT) in long-haul optical fibre communication systems with digital coherent receivers. It aims to improve the data rate limit caused by the strong non-linear inter-channel interference at high signal powers.

The non-linear frequency-division multiplexing (NFDM) system using the continuous spectrum of NFT was systematically assessed and compared with the conventional WDM system. The per-sample capacity of NFDM was estimated via the Blahut–Arimoto algorithm using statistics from a large number of system simulations of an EDFA-amplified optical fibre link of 1600 km. NFDM shows only a marginal gain of 0.5 bit/symbol in achievable rate in comparison with the conventional WDM system in the per-sample capacity sense. However, the spectral efficiency (SE) of NFDM is not even comparable with SE of the current WDM system due to the inefficient modulation approach.

In this context, efforts were made to find a better signal modulation scheme using NFT. A dual-polarisation b-modulation scheme was proposed instead of the conventional q_c-modulation scheme. At the same time, a large number of sub-carriers up to 1050 and the dispersion pre-compensation technique were also adopted to improve the SE of the NFT-based system. The above modifications resulted in the highest net rates among existing NFT-based systems of 400 and 220 Gbit/s (7.2 and 4 bit/s/Hz) on an EDFA-amplified fibre link of 960 km in single-channel simulation and experiment, respectively.

In addition, progress was made on the modelling of spectral broadening of a Gaussian process in optical fibre using the iterative Kolmogorov-Zakharov model. The bandwidth expansion is observed to be related to the signal power, distance and initial signal bandwidth in a certain form. The relation can be used to calculate a tighter upper bound on the information-theoretic capacity of the optical fibre channel.

Impact Statement

At the time of writing, many cities around the world have gone into lockdown in response to an unprecedented pandemic. People are recommended to stay and work at home. Behind the scenes the global telecommunication infrastructure is running smoothly; ensuring that, at a time when people are physically disconnected, they remain virtually connected. The backbone of that infrastructure is the optical fibre communication network, where 99% of the data traffic run through.

It is hardly noticeable that, due to the non-linear nature of the optical fibre channel, optical fibre networks are approaching a "capacity crunch", i.e., optical fibre systems not being able to increase their data carrying capability in line with traffic demands. However, the trend of data traffic growth over the last 20 years is evidently saying so. There are currently three unique paths to address the problem, i.e., overcome the non-linearity, increase the usable bandwidth, and add more spatial dimensions. Each path has its own set of challenges and should be pursued in parallel. The research described in the thesis explored techniques to overcome the non-linearity in the optical fibre channel.

In the thesis, transmission schemes based on non-linear Fourier transform (NFT) were investigated from the initial proof-of-concept simplified system to the fully evolved experimental set-up at Nokia Bell Labs. The data rates of such transmission schemes have been increased from around 10 Gbit/s to 400 Gbit/s. Issues that stop further improvements on data rate have also been identified for future research. Additionally, other non-linear effects such as the spectral broadening was modelled in the highly non-linear regime and its implication on the achievable spectral efficiency of the optical fibre channel was also derived.

The main conclusion of the research is that, in their current form, transmission schemes based on NFT are still not beneficial to be implemented in practical systems. It is unclear whether the low data rate of NFT-based systems is due to some fundamental limits of NFT signal processing or simply just the insufficient sampling rate. Using the highly non-linear regime of optical fibre channel for information transmission remains infeasible, at least in the short term.

Acknowledgements

WITH the utmost sincerity, I wish to thank my academic supervisors, Prof. Polina Bayvel and Dr. Domaniç Lavery, for sharing their experience and insight and offering support in all aspects of life during my time at UCL. My gratitude extends equally to my industrial supervisors, Dr. Vahid Aref and Dr. Son Thai Le, for the insightful discussions and warm-hearted friendship during my 18-month secondment at Nokia Bell Labs in Stuttgart, Germany.

I would like to thank Dr. Mansoor Yousefi who introduced me to the fascinating topic of non-linear Fourier transform, and Dr. Alex Alvarado who, together with Polina, interviewed me for the PhD position and gave me great guidance and advice before leaving for an associate professorship at TU/e.

Additionally, I would like to thank my colleagues at UCL for the many enjoyable conversations over a cup of tea or a pint of lager. I would particularly like to thank Eric for he has never disappointed me for any Linux-related issues, and Tom for our shared cause for Alzheimer's disease. To my colleagues at Nokia Bell Labs in Stuttgart, thank you for creating an inviting the friendly work environment. And I will always remember Karsten's leaving words for me – *Sei brav*.

I would also like to thank my housemates Daniel, for welcoming me with open arms when I moved to London for the first time, and Boris, for many enjoyable dinners on our balcony in Stuttgart in the summer breeze.

Above all, my deepest gratitude goes to my parents, who have supported me unconditionally throughout my academic years. I also wish to thank my girlfriend Luchen for her unlimited support and inspirations.

Finally, I would like to acknowledge the financial support from the European Union's Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement No.676448.

I do not think that the wireless waves I have discovered will have any practical application.

by Heinrich Rudolf Hertz

Contents

1 Introduction **21**

1.1 A brief history of optical fibre communications 22
1.1.1 Optical amplifier 23
1.1.2 Wavelength division multiplexing 24
1.1.3 Coherent receiver 25
1.2 The challenges in optical fibre communications 27
1.2.1 Exponential increase of the global data traffic 27
1.2.2 Capacity limits of optically routed network 28
1.2.3 Roadmap to the challenges 29
1.3 Problem definition and key contributions 31
1.4 Thesis outline 32
1.5 List of publications 32

2 The fundamentals of optical fibre communications **35**

2.1 Optical transmitter 35
2.1.1 External cavity semiconductor lasers 36
2.1.2 PDM I/Q modulator 37
2.1.3 Transmitter impairments 38
2.2 Physics of optical fibre channel 39
2.2.1 Wave propagation equations – NLSE 40
2.2.2 Linear and non-linear propagation effects 41
2.2.3 The evolution of power spectral density 45
2.2.4 Polarisation effects 48
2.2.5 Split-step Fourier method for numerical simulations 51
2.3 Digital coherent receiver 53
2.3.1 90°-hybrid 54
2.3.2 Balanced photo-diodes 55
2.3.3 Receiver impairments 56
2.3.4 Typical sequence of DSP 56
2.4 Information-theoretical tools 61
2.4.1 Mutual information and channel capacity 63
2.4.2 Numerical estimation of mutual information and capacity . . 65
2.4.3 Entropy estimators 67
2.4.4 Interference channels 70

2.4.5 Models and capacity of fibre channels 74

3 Spectral broadening of Gaussian processes in optical fibre channel 81
3.1 Related work 81
3.2 The optical fibre model 83
3.3 The iterative Kolmogorov-Zakharov model 84
3.4 Characterisation of spectral broadening 85
3.5 Discussion on the spectral efficiency 88
3.6 Summary 89

4 Nonlinear Fourier Transform 90
4.1 Review of single-polarisation NFT 91
4.2 Review of dual-polarisation NFT 96
4.3 Numerical methods for the ZS system 98
4.4 Numerical examples of INFT-NFT algorithms 101
4.5 Approximating the integrable model 103
4.6 Summary 104

5 Per-sample capacity estimation of a NFDM optical network 105
5.1 Related work 105
5.2 Optical fibre networks 107
5.3 Review of NFDM 107
5.3.1 NFDM transmitter 108
5.3.2 Reducing the peak-to-average-power-ratio (PAPR) 108
5.3.3 Channel filter 109
5.3.4 NFDM receiver 109
5.4 Comparison of the achievable rates of WDM and NFDM 110
5.4.1 AIRs of WDM and NFDM 110
5.4.2 Stochastic memory 114
5.5 Impact of perturbations and non-idealities 115
5.5.1 Loss with periodic amplification 115
5.5.2 Polarisation effects and higher-order dispersion 117
5.5.3 Spectral efficiency 119
5.6 Comments on complexity 122
5.7 Summary 124

6 Dual-polarisation NFDM transmission with b-modulation 125
6.1 Related work 125
6.2 The optical fibre channel model 128
6.3 q_c- and b-modulated DP-NFDM systems 128
6.4 Optimisation of b-modulated DP-NFDM transmission 133
6.5 Co-variance matrix of q_c- and b-modulated signals 136
6.6 Some drawbacks of the current system 140
6.6.1 Transceiver distortion in noiseless back-to-back scenario . . 140

6.6.2 Channel mismatch due to non-integrability 140
6.7 *b*-modulated DP-NFDM experiment 143
6.7.1 Experimental setup and DSP 143
6.7.2 Experimental results 147
6.8 Summary 147

7 Thesis summary and future work **149**

Appendices **153**

A Probability distribution of a random variable **153**
A.1 Probability distribution of a discrete random variable 153
A.2 Probability distribution of a continuous random variable 154

B Power spectral density **155**

C The Gaussian Noise model **157**

Bibliography **159**

List of Figures

1.1 (left) Worldwide submarine optical fibre cable map provided by TeleGrography. (right) African undersea and terrestrial fibre optic cable map provided by the Network Startup Resource Center. 22
1.2 Fibre attenuation profile across wavelength with (solid) and without (dashed) water peak, adapted from [1, Fig. 7]. 23
1.3 EDFA block diagram. 24
1.4 WDM system structure. 25
1.5 Optical add-drop multiplexer implemented with fibre Bragg gratings (FBG) [2–4]. 26
1.6 Data growth Amsterdam and Seattle Exchange at roughly 46% annual increase. 28
1.7 (a) Typical ORN structure and (b) spectral efficiency (SE) estimation adapted from [5]. signal-to-noise ratio (SNR) is the ratio of the signal power and the total noise power. 30
1.8 The evolution of NFT-based systems in this thesis. 33

2.1 Generic block diagram of a coherent optical communication systems. 36
2.2 The schematic of a polarisation diversity optical transmitter unit. It consists of four digital to analogue converters (DACs), a polarisation beam splitter (PBS), two optical I/Q modulator (IQM) and a polarisation beam combiner (PBC). 36
2.3 A schematic of the external cavity laser with an external grating. . . 37
2.4 A schematic of an IQM that consists of two Mach-Zehnder modulators. 37
2.5 IQ imbalance at transmitter induced by (left) Gain Error g_{I}, g_{Q}, (middle) Offset o_{I}, o_{Q}, and (right) Quadrature Error ϕ_{IQ}. 39
2.6 power spectral densitys (PSDs) from simulations scaled by the maximum of the Gaussian noise (GN) PSD at 12 dBm. 48
2.7 PMD, adapted from [6]. 50

2.8 Split-step Fourier method for the coupled non-linear Schrödinger equation. 53
2.9 The schematics of (a) direct and (b) coherent receiver, adapted from [7]. 54
2.10 The schematic of a 90°-hybrid. 55
2.11 The typical DSP sequence, adapted from [8]. 57
2.12 2×2 MIMO structured adaptive FIR filter. 58
2.13 Optical fibre system block diagram for information-theoretic analysis [9]. 61
2.14 (a) A uniformly distributed multi-ring constellation and (b) One example of the received constellation with the transmitted phase removed. 66
2.15 A two-transmitter IC. Sc. stands for source, Enco. for encoder, Deco. for decoder and Sk. for sink. 70
2.16 Rate regions for $P_1 = P_2 = 6$ and increasing $a_{12} = a_{21}$. The legends are for (a) treating interference as noise; (b) using TDM/FDM; (c) decoding interference; (d) Sato's outer bound. 72
2.17 The AIR of the above models in a 4000 km optical fibre link. 77

3.1 PSDs from simulations scaled by the maximum of the GN PSD with same link parameters and doubled signal power of Fig. 2.6. 82
3.2 Characterisation of spectral broadening in a 20-span transmission link with different initial bandwidths of 32/64/128 GHz. The number of iterations holds for 32 GHz bandwidth. Simulations are with 16-QAM 0.01-RRC-pulse signals. 86
3.3 Characterisation of spectral broadening of signals with 32 GHz initial bandwidth at different transmission distances (20/40/80 spans). . 87
3.4 Spectra of a 32 GHz signal at (a) 12 and (b) 16 dBm. Solid, dashed and densely dotted lines are input, output, and KZ model signal spectra, respectively. 87

4.1 (a) The fundamental soliton $q(\tau) = 2\,\mathrm{sech}(2\tau)$. (b) The continuous spectrum of the fundamental soliton calculated by the NFT algorithm in the case of $\mathrm{sgn}(\beta_2) = -1$. (c) The recovered (dotted) and original (solid) fundamental soliton in the case of $\mathrm{sgn}(\beta_2) = 1$. (d) The continuous spectrum of the fundamental soliton calculated by the NFT algorithm in the case of $\mathrm{sgn}(\beta_2) = 1$. 102

5.1 Pre-equalisation methods used (left) in [10] and (right) in Sec. 5.4. 109

5.2 For the integrable model: (a) interference in WDM; (b) lack of interference in NFDM; (c) Fourier spectra at distances with maximum input-output bandwidths. For the lossy model with periodic amplification: (d) interference in NFDM. Noise is set to zero in these figures. 111

5.3 WDM and NFDM AIRs in the integrable model and focusing regime. 113

5.4 Constellation in (left) WDM and (right) NFDM in the focusing regime. 113

5.5 Transmitted and received (grey and black markers) symbols in (left) additive white Gaussian noise (AWGN) and (right) non-linear frequency-division multiplexing (NFDM), showing the SNR loss due to the stochastic memory. 115

5.6 Received constellations of (left) NFDM and (right) WDM systems in the focusing regime for the non-integrable model. 117

5.7 AIRs of WDM and NFDM in non-integrable models. The curves denoted by LOSS correspond to the lossy model with periodic amplification (no PMD). The points denoted by ALL include all perturbations (loss, PMD and third-order dispersion). 117

5.8 Received constellations of (left) NFDM and (right) WDM in the focusing regime with PMD and third-order dispersion, in dual polarisation transmission. 119

5.9 NFDM signals (left) with and (right) without pre-compensation. . . 120

5.10 Received symbols in NFDM and WDM with $N_s = 15$ at different powers. 121

5.11 The SE of NFDM and WDM with $N_s = 15$ and guard-time. For dual-polarisation transmissions in both systems, the SE and power is measured per polarisation. 122

5.12 Comparison of the time duration at the (left) input and (right) output. Here $N_u = 5$ and $N_s = 15$. The time duration at input is $T_{\text{in}} \approx 2.1$ ns, while $T_{\text{out}} \approx 22.3$ ns, which can be reduced to 11 ns as shown in Fig. 5.9. The temporal expansion ratio for $N_s = 15$ is 10. 122

6.1 Gross data rate of recently implemented NFDM systems of different transmission distances in experiments and simulations. 126

6.2 A realisation of signal at different stages of INFT-signal generation without pre-dispersion compensation. (a) OFDM spectrum as $\mathbf{u}_1$, (b) $\mathbf{b}_1$ spectrum when applying the transformation Γ_b on $\mathbf{u}$, (c) IDFT of $\mathbf{b}_1$ with tails due to the transformation Γ_b, (d) time-domain signal generated by INFT wihtout pre-dispersion compensation (PDC). (e) Transceiver digital signal processing chains. 131

6.3 Simulation diagram of (a) b- or (b) q_c-modulated DP-NFDM systems. $\mathbf{q}_b(t)$ and $\mathbf{q}_c(t)$ are the time-domain signals of each system. . . 132

6.4 Q-factor vs. signal launch power in q_c- and b-modulated DP-NFDM systems with a large guard interval $\eta = 4$ in both AWGN case and fibre transmission. The total additive noise powers are the same. 133

6.5 The impact of decreasing η on SE in q_c-modulated DP-NFDM system for different launch powers. $N_C = 210/(\eta - 1)$. 134

6.6 The impact of decreasing η on (a) Q-factor and (b) SE in b-modulated DP-NFDM system for different launch powers. $N_C = 210/(\eta - 1)$. The corresponding received constellation at the power of the maximum Q-factor (c) and SE (d). 135

6.7 Conditional differential entropy of received symbols in both q_c and b-modulated DP-NFDM systems with the assumption of Gaussian distribution, for different transmitted symbols (a) and powers (b). . . 138

6.8 The relative error between distribution-agnostic entropy estimator [11] and the Gaussian entropy for 10 adjacent sub-carriers. G stands for Gaussian entropy, DA for distribution-agnostic entropy estimation. For both methods, the result of one-dimensional entropy $h(Y_i|\mathbf{x})$ is very close. $\varepsilon = \left(\sum_{i=1}^{10} h(Y_i|\mathbf{x}) - h(Y_1, ..., Y_{10}|\mathbf{x})\right) / \sum_{i=1}^{10} h(Y_i|\mathbf{x})$. 139

6.9 'Transceiver distortion' in the b-modulated DP-NFDM system in the noiseless back-to-back scenario. The average energy of NFDM symbol is calculated by $\mathcal{P}(T_0 + T_G) = \mathcal{P}\eta N_C/W$. 140

6.10 (a) (b) Simulation diagram of non-integrable models. PA-DBP stands for path-averaged DBP, IDBP for ideal DBP, COI for channel-of-interest. (c) residual distortion of IDBP (shortened in the legend to I, preceded by total bandwidth and followed by number of channels) and PA-DBP (shortened in the legend to PA) systems. 142

6.11 A realisation of INFT symbols (a) without PDC (b) with PDC. . . . 143

6.12 Transceiver DSP chain. 144

6.13 Signal structure in one duty cycle. TS stands for training symbols. . 145

6.14 Recirculating fibre loop experiment setup. The fibre is Corning SMF-28. 146
6.15 Measured SNRs and their average (a) in terms of the designed launch power in the back-to-back scenario, (b) in terms of the launch power over 12 spans of 81.3 km SMF. Below the curve is one received constellation at signal power of -1.3 dBm, Q factor = 3.7 dB, SE ≈ 2 bit/s/Hz/pol. 148

List of Tables

3.1 Parameters of the optical fibre communication system. 83

5.1 Fibre and system parameters 110
5.2 Parameters of the periodic amplification model 116
5.3 polarisation and dispersion parameters 118
5.4 Complexity of NFDM and WDM 124
5.5 Maximum power reached by the basic DLP 124

6.1 Categorisation of recent results of NFDM systems. 127
6.2 Fibre and system simulation parameters 128
6.3 System parameters 144

Acronyms

ADC analogue to digital convertor

ADM add-drop multiplexer

AIR achievable information rate

ASE amplified spontaneous emission

ASIC application-specific integrated circuit

AWGN additive white Gaussian noise

BDFA Bismuth-doped fibre amplifier

BER bit error rate

CD chromatic dispersion

CMA constant-modulus algorithm

CNLSE coupled non-linear Schrödinger equation

COI channel-of-interest

DAC digital to analogue converter

DBP digital backpropagation

DC data centre

DFT discrete Fourier transform

DGD differential group delay

DLP discrete layer-peeling

DoF degrees-of-freedom

DP dual polarisation

DSP digital signal processing

ECSL external cavity semiconductor laser

EDFA Erbium-doped fibre amplifier

EGN enhanced Gaussian noise

ENOB effective number of bits

EPMGP expectation propagation multivariate Gaussian probability

EVM error vector magnitude

FDE frequency-domain equaliser

FDM frequency-division multiplexing

FEC forward error correction

FFT fast Fourier transform

FIR finite impulse response

FLOP floating-point operation

FMF few-mode fibre

FOC frequency offset compensation

FWM four-wave mixing

GI guard interval

GLM Gelfand-Levitan-Marchenko

GMI generalised mutual information

GN Gaussian noise

GVD group velocity dispersion

GSOP Gram-Schmidt orthogonalization procedure

IC interference channel

ICI inter-channel interference

IDFT inverse discrete Fourier transform

INFT inverse non-linear Fourier transform

IQM I/Q modulator

ISI inter-symbol interference

IST inverse scattering transform

KdV Korteweg-de Vries

KNN k nearest neighbors

KZ Kolmogorov-Zakharov

LO local oscillator

MAP maximum *a posteriori*

MCF multi-core fibre

MCM Mach-Zehnder modulator

MI mutual information

MIMO multiple-input multiple-output

NFDM non-linear frequency-division multiplexing

NFT non-linear Fourier transform

NLSE non-linear Schrödinger equation

OADM optical add-drop multiplexing

OFDM orthogonal frequency-division multiplexing

OPC optical phase conjugation

ORN optically routed networks

OSNR optical signal-noise-ratio

PA path-average

PAPR peak-to-average-power-ratio

PBC polarisation beam combiner

PBS polarisation beam splitter

PDC pre-dispersion compensation

PDF probability density function

PDM polarisation-division multiplexing

PMD polarisation-mode dispersion

POL-DEMUX polarisation-demultiplexing

PSD power spectral density

PSoP principal state of polarisation

QAM quadrature amplitude modulation

QPSK quadrature phase-shift keying

RLS recursive least-squares

RMSE root mean squared error

RRC root-raised-cosine

Rx receiver

SDM space-division multiplexing

SE spectral efficiency

SINAD signal-to-noise and distortion

SNR signal-to-noise ratio

SoP state-of-polarisation

SP single polarisation

SPM self-phase modulation

SSFM split-step Fourier method

SSMF standard single-mode fibre

TDE time-domain equaliser

TDM time-division multiplexing

TIA transimpedance amplifier

Tx transmitter

UOI user-of-interest

VoD video on demand

WDM wavelength-division multiplexing

WSS wavelength-seletive switch

WWT weak wave turbulence

XPM cross-phase modulation

ZS Zakharov-Shabat

1

Introduction

In 1975, when the world's first commercial optical fibre communication system was lit up, operating at a bit rate of 45 Mbit/s with distance up to 10 km [12], people began to share the vision of Charles Kao that in a decade's time, optical fibres would be used for all major long-distance communications. Fast-forwarding more than four decades to today, the global optical fibre communication networks have grown way beyond anyone's ambition. With in total over 3 billion km [13] of deployed fibre, worldwide, in 452 submarine and countless terrestrial cables, optical fibre communication networks carry 99% of the world's digital information. In Fig. 1.1 there are just two examples of how important optical fibres have become to a continent's and the world's communication infrastructure. Optical fibre is simply the most efficient (in both economical and functional sense) and irreplaceable transmission medium in the modern era. However, the optical fibre communication network is also facing great challenges as the data traffic grows exponentially and data transmission capacity per-fibre saturates. Research described in this thesis is aimed to further improve the data rate of optical fibre communication systems by investigating novel digital signal processing (DSP) techniques that can mitigate or remove the non-linear distortion induced by the Kerr effect in optical fibre. Before diving into the technical details of optical fibre communications, we will first walk through the evolution of optical fibre networks and the urgent challenges they face.

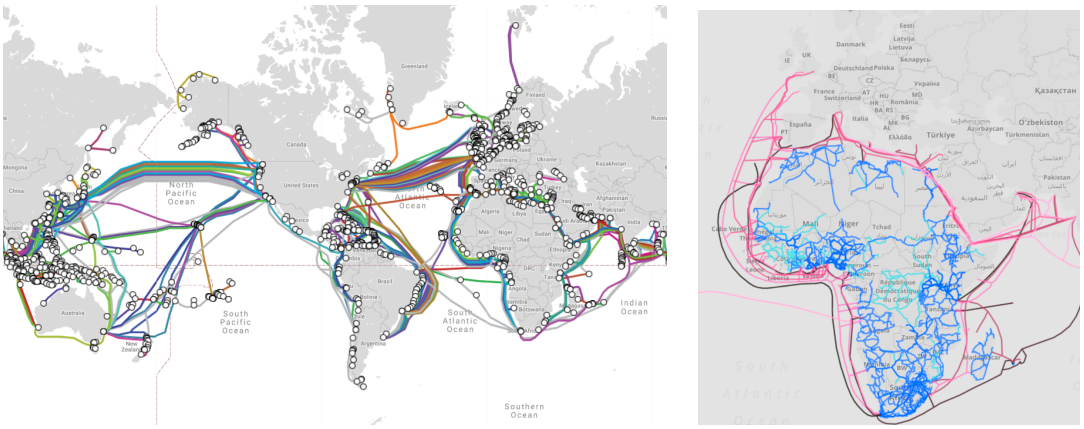

Figure 1.1: (left) Worldwide submarine optical fibre cable map provided by TeleGrography. (right) African undersea and terrestrial fibre optic cable map provided by the Network Startup Resource Center.

1.1 A brief history of optical fibre communications

In 1965, Charles Kao and George Hockham concluded, in their seminal paper [14], that the fundamental limitation on the attenuation of light travelling in glass is below 20 dB/km, which is a key threshold value for optical communications with information capacity in the excess of 1 Gbit/s. Since then, the optical fibre communication systems have experienced rapid development due to the advance of laser and optical fibre technologies. The first and second generation optical fibre systems operated at the wavelength of 800 nm and 1300 nm, respectively. The prototype of the modern optical fibre system emerged in late 1980 as the third-generation optical fibre systems. It operated at the wavelength of 1550 nm and the fibre had losses of 0.2 dB/km. Later on, the manufacture process of optical fibre improved and it opened up a rather broad low-loss window, which is now mainly referred to as C band 1530-1565 nm ($\approx$ 4.4 THz[1]). Nevertheless, the low-loss wavelength was still restricted by the attenuation peak around 1383 nm due to the OH ions; see the solid curve in Fig. 1.2. In 1998, optical fibre and cable supplier Optical Fiber Solutions (OFS) successfully developed the zero water peak fibre [15], opening up the possibility of using the E band 1360-1460 nm region ($\approx$ 15.1 THz) for long distance transmission; see the dashed curve in Fig. 1.2. Overall, the attenuation of fibre has reached roughly one twentieth of the key threshold value for optical communications over a very large bandwidth of roughly 53.5 THz.

While the reduction of attenuation of optical fibre played a significant role in pushing the optical fibre communication system into commercialization, there are

[1] $W = c/\lambda_1 - c/\lambda_2$ where $c = 3\times10^8$ m/s is the speed of light in vacuum.) and L band 1565-1625 nm ($\approx$ 7.1 THz).

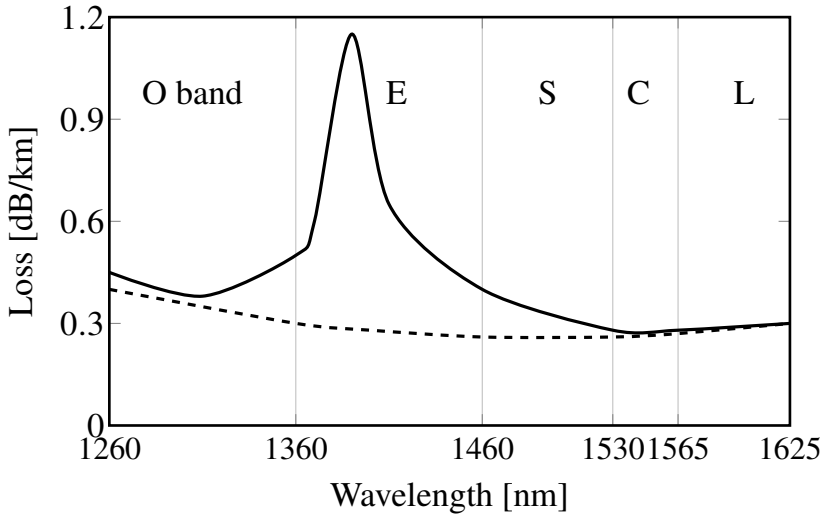

Figure 1.2: Fibre attenuation profile across wavelength with (solid) and without (dashed) water peak, adapted from [1, Fig. 7].

three technological milestones that saw the soaring of transmission data rates and the dramatic increase of transmission distance of commercial systems: Erbium-doped fibre amplifier (EDFA), wavelength-division multiplexing (WDM) and coherent technology.

1.1.1 Optical amplifier

Before the invention of the optical amplifier, long-distance optical fibre communication systems usually required electronic repeater (or regenerator) every few kilometres. A electronic repeater consists of a receiver, an electrical amplifier and a transmitter. The electronic repeater converts the optical signal to a electrical one, regenerates the signal and re-converts the electrical signal back to optical. The structure exposes three limitations of this technique: 1) highly cost-ineffective since the repeater is effectively another set of transceiver, 2) the repeater interface rate, i.e., the net bit rate that transceivers in the repeater were able to support, limits the link capacity, 3) for WDM systems, channels on different wavelengths have to be demultiplexed first, with each channel regenerated separately by its own electronics, and then multiplexed together for further transmission. The number of transceivers required at the repeater will be equal to the number of channels in the WDM system, which will incur an unacceptable cost on the system.

In spite of the above limitations, the first optical fibre sub-sea communication systems deployed across the Atlantic Ocean (TAT8) in the late 1980s was opto-electronically regenerated systems operating at 1.3 μm and carrying 280 Mb/s on each of its three fibre pairs. The next-generation sub-sea link TAT9 was then

changed to operate at 1.55 μm with doubled capacity of the previous systems, but it was still opto-electronically regenerated.

In 1987, EDFA was proposed in two seminal papers [16, 17] around the same time to tackle the limitations of the electronic repeater. It is very difficult to overestimate the importance of EDFA in modern optical communication systems due to its superb properties. To start with, the EDFA amplifies the signal without any opto-electronic conversion. Then, the amplified spontaneous emission (ASE) noise added to the signal is negligible and can normally be modelled as a Gaussian noise. Finally, the EDFA has a very large operating bandwidth that covers the entire C band and the gain spectrum is almost flat. Based on EDFA, the dual-stage optical amplifier [18] invented in 1995 paved way for cost-effective and large scale high capacity WDM systems.

A modern EDFA, typically working in the 1550 nm band, consists of a length of fibre doped with Erbium and a pump laser at 980 or 1480 nm. The pump laser supplies the energy for the amplifier, while the incoming signal stimulates emission as it passes through the doped fibre. A schematic of EDFA is shown in Fig. 1.3. With the help of EDFA, optical fibre communication systems officially entered the stage of wideband WDM systems.

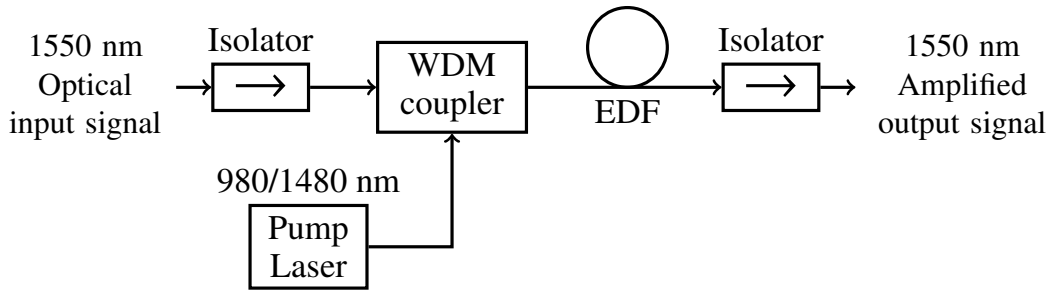

Figure 1.3: EDFA block diagram.

1.1.2 Wavelength division multiplexing

The essence of WDM system is to offer extra degrees of freedom in the order of hundreds for data transmission on different wavelengths of laser light. As shown in Fig. 1.4, an optical multiplexer is placed at the transmitter to combine several optical signals from individual transponders together and a de-multiplexer at the receiver to separate them. Between 1995 to 2000, the data rate per fibre in WDM systems has experienced 100% (due to the rapid increase of WDM channel count) annual growth rate, and continued 20% annual growth in the following years [1, Fig. 4]. The initial problem of increasing the channel count was the variation of the EDFA gain profile across wavelengths. It caused huge difference in optical signal-noise-ratio (OSNR) for different channels at the output of the link, rendering some of

the channels unusable. The effect is most significant in multi-span systems, where the variation of the EDFA gain profile accumulates over distance. In 1992, with the help of practical gain flattering filter technology, the problem was solved by an algorithm [19, 20] that calculated the appropriate launch powers for all channels in a WDM system so that their output OSNR are equal. This algorithm is commonly referred to as the optical pre-emphasis.

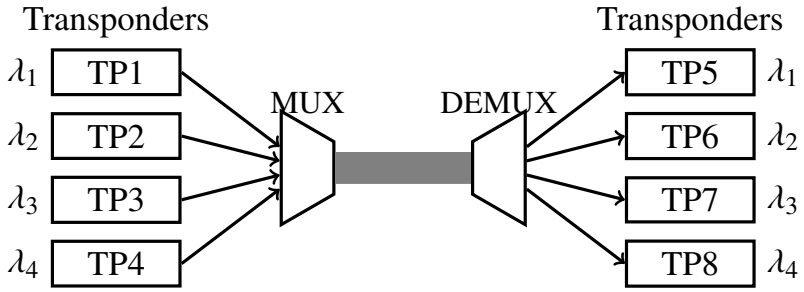

Figure 1.4: WDM system structure.

The other benefit of WDM systems comes from the optical networking and bandwidth management perspective. In the mid-1990s, as the capacity of single wavelength in WDM systems has surpassed the capacity of electronic switches and routers, multiple electronic signals were aggregated into a single wavelength. It became desirable to manage the bandwidth directly in the optical domain.

The device that aggregates, dis-aggregates, and switches data paths in the optical domain "colourlessly" and "directionlessly" is commonly referred to as the optical add-drop multiplexing (OADM). "Colourless" refers to the ability of OADM to flexibly allocate any wavelength or colour to any port, while "directionless" indicates that the OADM allows wavelengths to be routed to any direction available on the node without physical rewiring of fibres [21]. A typical fixed OADM system was initially just a set of optical de-multiplexer and multiplexer with a reconfiguration mechanism in between, e.g., a fibre Bragg grating with optical circulators [2, 3] as shown in Fig. 1.5. The modern OADM has evolved into a remotely re-configurable one with the help of wavelength selective switching module. It provides multi-degree mesh connectivity, flexible wavelength allocation, dynamic gain equalisation to improve the OSNR and maximize reach.

1.1.3 Coherent receiver

Initially studied in the 1980s [22], coherent receivers were resurrected in 2000s after 15 years of being unnoticed. It became an ideal solution for the troublesome problems caused by chromatic dispersion (CD) and polarisation-mode dispersion (PMD). At the same time, CMOS processing speed was catching up with

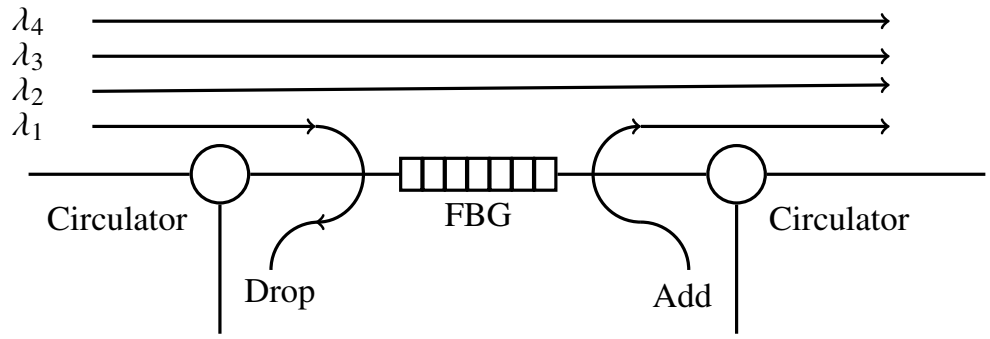

Figure 1.5: Optical add-drop multiplexer implemented with fibre Bragg gratings (FBG) [2–4].

10 GBaud symbol rate. With these advances, digital coherent receivers could bring 4 times increase in spectral efficiency (SE) through quadrature modulation (I/Q) and polarisation-division multiplexing (PDM).

Before the introduction of coherent receiver, CD was compensated inline by concatenating two types of fibre with opposite signs of dispersion so that the overall CD at the output of a link is nearly zero but the "local" CD everywhere along the link is enough to suppress non-linear effects. This technique was referred to as dispersion management [23, 24]. When the coherent receiver was re-introduced in 2000s, it provided access to the full optical field in digital form and simplified the link design by using only one type of fibre for the entire link and compensating the chromatic dispersion entirely in the digital domain. This gives significant advantage over the dispersion-managed fibre link by reducing the network complexity and costs. It also opened up the possibility of digital compensation of the PMD effect and fibre non-linear distortion. The research focus of this thesis, digital compensation techniques of non-linear effects, would not be possible without coherent receivers.

The commercialisation of digital coherent receiver is very successful in long-haul and even metro optical networks. The coherent receiver is also attractive for the optical access network due to its high receiver sensitivity [25, Sec. II]. However, because of its significantly higher cost than direct-detection receiver, digital coherent receiver has not prevailed in the short-reach optical links, where the connectivity and low cost is more important than high data rate. The high cost of a digital coherent receiver comes from the need of a local oscillator (LO) laser at the receiver, two 90°-hybrid, four balanced photodiodes, four analogue to digital convertors (ADCs) and an application-specific integrated circuit (ASIC) to run the powerful digital signal processing (DSP) for transmission impairment compensation. There have been many endeavours to develop simplified coherent receiver targeting optical access

network, e.g., [25]. Nevertheless, the cost of simplified coherent receiver has not dropped below the threshold for massive deployment. Although there are numerous quasi-coherent solutions [26,27] that use direct-detector and advanced DSPs and are claimed to be low-cost, the author believes that in the future, coherent receivers will become cost-efficient enough to replace any other types of transponders completely, over all distance of optical fibre links.

The deployment of coherent receivers also changed network planning. At the age of direct-detection with dispersion managed links, optical communication systems had to be simulated using computationally expensive split-step Fourier method (SSFM) to achieve an acceptable accuracy in performance prediction in network management. For dispersion-unmanaged links with coherent receivers, semi-analytical models such as the GN model [28, 29] or the enhanced GN model [30] can predict the system performance quite accurately with little computational effort. Such advantages have made them a great tool in flexible optical networks to provide quick assessments of certain path's transmission quality [31]. However, for novel transceiver and link designs, SSFM is still indispensable.

1.2 The challenges in optical fibre communications

In the previous section, we have walked through the great progress and the enabling technologies of optical fibre communication systems in the last 55 years. The achievable data rate transmitted in a single fibre has increased from 45 Mb/s to over 100.5 Tbit/s [32]. The applications running on this global fibre communication infrastructure has evolved from just sending simple messages to streaming videos or even virtual reality. The optical fibre communication systems have also evolved from point-to-point links to networks with diverse structures, especially due to the emerge of data centres in recent years. For a while, it seems that our work of building fast and reliable optical fibre networks has been accomplished. However, global data traffic trends in the past decade have suggested otherwise. In this section, we will look closely into the exponential increase of the global data traffic and the saturating capacity of current optical fibre networks.

1.2.1 Exponential increase of the global data traffic

Several review papers [1, 33–35] have reported various growth rate of data traffic around the globe ranging from 30%-70%. The measurement are usually taken at regional broadband internet providers, or at their internet exchanges, e.g., the Seattle Internet Exchange or the AMS-IX. Fig. 1.6 shows the historical data traffic volume gathered at AMX-IX and the Seattle Internet Exchange in the period of 2002-2019. By simply connecting the start and end point, we can observe an annual growth rate

of 46%. Assuming the same growth rate for the future prediction, we can easily calculate the exhaust time of current system [1]

$$C_{\text{system}} = C_{\text{current}}(1 + r_{\text{growth}})^{T_{\text{exhaust}}}$$

where C_{current} is the current traffic load on the system, r_{growth} is the annual traffic growth rate, and T_{exhaust} is the time it takes to fully exhaust the installed system capacity C_{system}. For example, if the current system loading is 10% and the annual growth rate is 46%, we have only 6 years until it is fully loaded.

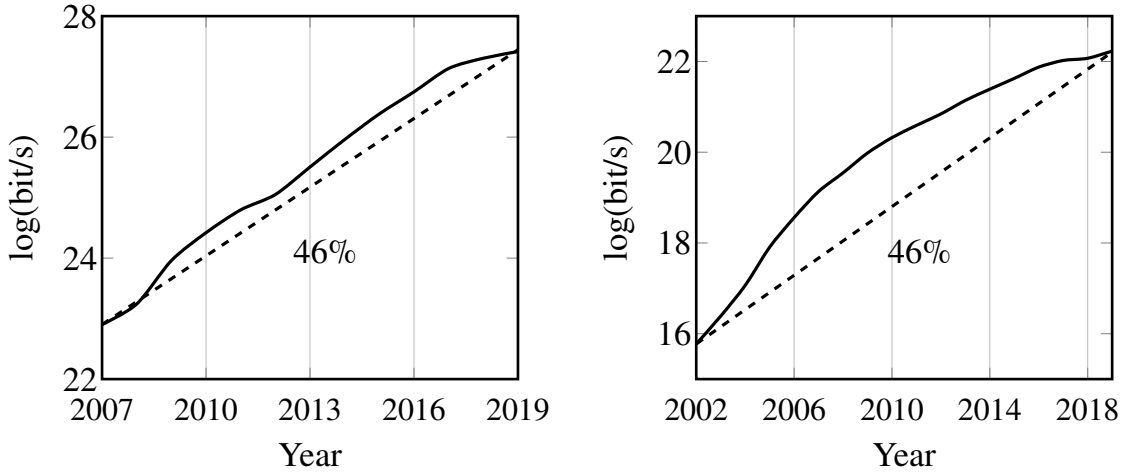

Figure 1.6: Data growth Amsterdam and Seattle Exchange at roughly 46% annual increase.

Of course, extrapolation to the future is by no means an accurate prediction. The great growth of data traffic is mainly attributed to the upsurge of video on demand (VoD) service such as Netflix and YouTube. As the VoD service matures, we can already see from Fig. 1.6 that the data traffic growth is slowing down in the recent years. However, the commercialisation of virtual reality technology or other unknown applications might give another even greater drive to the data traffic. The optical fibre networks must prepare for the future challenges. Also note that due to the uprising of the data centre (DC) technology, the frequently accessed content are often cached within the metro network, resulting in a 75% of the traffic terminated within the metro network in 2017 as opposed to 25 % in 2012 [35]. This traffic flow shift alleviates the pressure on the backbone optical networks but requires more capacity for the DC-DC interconnect and DC-user connections.

1.2.2 Capacity limits of optically routed network

Given the challenge of accommodating the ever-growing data traffic in optical fibre networks, network operators would ideally like to upgrade the optical fibre networks

at a matching growth rate in terms of link capacity. However, the system capacities of optical fibre links has been experiencing only 20 % annual increase since 2002. Furthermore, we are approaching to the information-theoretic capacity limit of standard single-mode fibre channel. A comprehensive study [5] on the capacity limit of long-haul optically routed networks (ORN) has shown that the capacity limit is mainly imposed by the inter-channel interference caused by the fibre non-linear Kerr effect. An ORN is normally defined as an optical fibre network running WDM with reconfigurable OADM such as the network demonstrated in Fig. 1.7 (a). In such a network, signals travel through the network via a path specified by a routing algorithm from their respective transmitters to their respective receivers. Different channels might partially share their paths, distort each others' signals and head to their own destinations. Therefore, a receiver only have access to its own channel and cannot remove the distortion caused by the other co-propagating channels. As a result, the SE of optical fibre link does not increase monotonically with SNR as shown in Fig. 1.7 (b). In addition, as the transmission distance increases, the SE at the optimal SNR also decreases. This non-linear phenomenon has been sometimes termed as the "non-linear Shannon limit" [36]. Although the limit is theoretically disputable [37] for it assumes Gaussian distribution on the input signal and considers only the conventional WDM scheme, no practical optical fibre communication system has managed to surpass it.

1.2.3 Roadmap to the challenges

With the exponential explosion in global data traffic and saturating capacity of optical fibre communication system, we can undoubtedly forsee the "capacity crunch" [5] of optical fibre network. To overcome the "capacity crunch", one can resort to Shannon's capacity limit for the AWGN channel [38],

$$C = W \log(1 + \mathrm{SNR}), \quad \text{bit/s} \tag{1.1}$$

where C, W and SNR denote the channel capacity, transmission bandwidth and signal-to-noise ratio, respectively. Strategies can be derived from (1.1)

- *Improve SNR with a variety of non-linear transmission schemes*, e.g. support vector machine, Volterra transfer function, digital backpropagation (DBP), optical phase conjugation (OPC) and non-linear Fourier transform (NFT). Among them, OPC and NFT both have theoretical foundation for the possibility of reducing inter-channel interference [39, 40], hence promise a significant improvement in the data rate. However, they require rather substantial changes in the network infrastructure. Other above-mentioned techniques

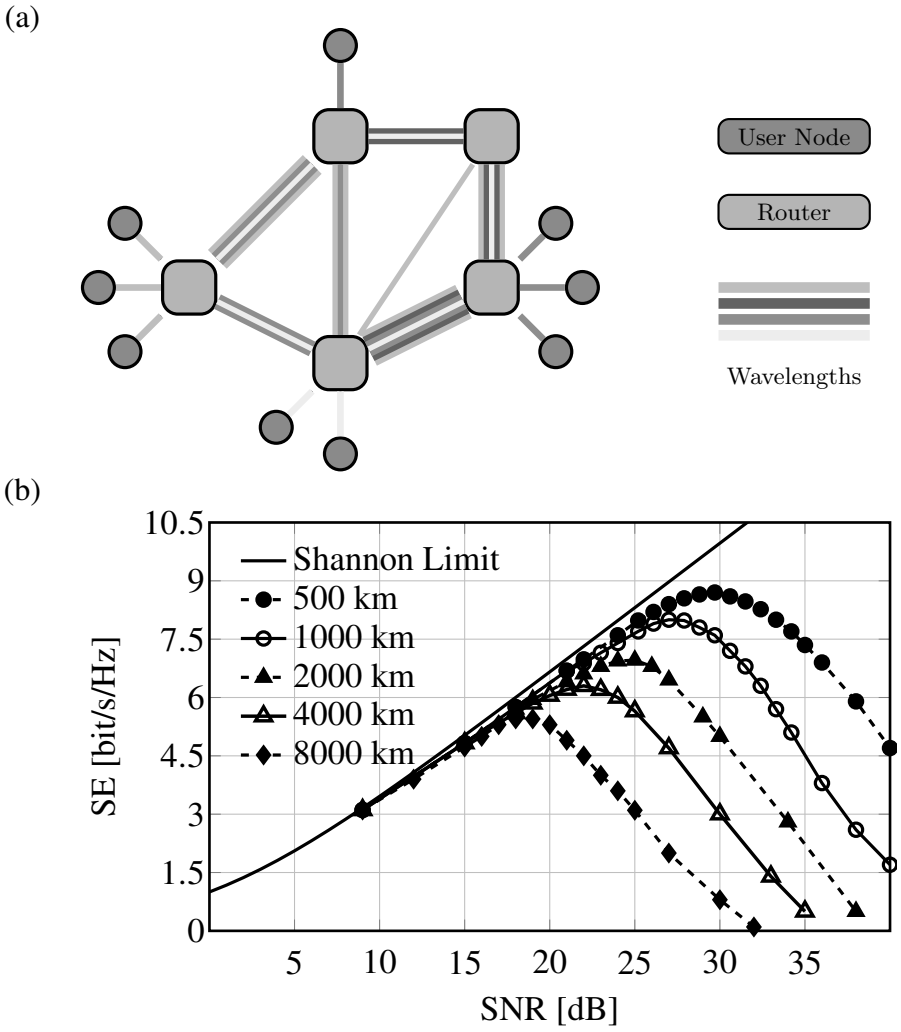

Figure 1.7: (a) Typical ORN structure and (b) spectral efficiency (SE) estimation adapted from [5]. SNR is the ratio of the signal power and the total noise power.

demand no changes in the infrastructure and are more straightforward to implement, but their gains in SNR are also limited. This will be explained in Sec. 2.4.4.

- *Utilise larger transmission bandwidth W.* The current transmission bandwidth is constrained by the low-loss windows (O~L band in Fig. 1.2, 1260 nm~1625 nm, i.e. 53.5 THz) of standard single-mode fibre (SSMF) and the availability of optical amplifiers at different wavelengths. With the recent advances in fibre-doped amplifier technologies, extended transmission bands become possible [41]. Conventional EDFA covers C+L bands (1530 nm - 1625 nm $\approx$ 11.5 THz), Bismuth-doped fibre amplifier (BDFA) operates across O and S bands (~1260 nm - 1530 nm, i.e., 42 THz) [42, 43]. U band is normally avoided due to its high transmission loss. Ideally, a 12-fold in-

crease in network capacity could be achieved by the full utilization of the low-loss window (O~L bands) as opposed to C band only, but it has been argued [1, Sec. III. C.] that the ultra-wideband transmission would not live up to our expectation but only provide roughly a 5-fold increase in network throughput at a price of a more complicated network management.

- *Exploit the spatial dimension.* As SSMF has reached a very high level of maturity, global efforts have focused on exploring new types of fibre that make use of spatial degrees-of-freedoms (DoFs) in the same glass strand [44, Sec. 3], which is generically referred to as space-division multiplexing (SDM). Besides the most obvious fibre bundle solution, other types of fibre, such as multi-core fibre (MCF), few-mode fibre (FMF) and combined few-mode MCF, have all demonstrated experimentally high data rate transmission in both short and long ranges [45–48]. Some successful studies on the non-linear propagation in FMF and MCF have also been reported in [49, 50]. In spite of the impressively high data rate achieved by MCF or FMF, ideal spatial parallelization requires not only new types of fibres but also integrated transceiver arrays, spatial and wavelength selective switches and multi-core/mode amplifiers, whose progress lags behind the fast development pace of new fibres. Note that, MCF or FMF does not seem to have clear advantages over fibre bundle. The fair comparison of these SDM regimes will eventually come down to the cost efficiency, development of in-line optical SDM amplifier, and transceiver integration.

1.3 Problem definition and key contributions

The research described in this thesis focuses on the first challenge listed above: improve SNR with non-linear transmission schemes. And in particular, two questions are of interest, 1) whether it is possible to achieve a monotonically increasing SE in the ORN, in contrast to the saturating behaviour shown in Fig. 1.7, 2) if such a transmission scheme exists, whether it is viable in practice.

The research described in this thesis made contributions in three aspects. To begin with, progress is made in finding a tighter information-theoretic upper bound on the SE of the optical fibre channel through accurate and fast estimation of the spectral broadening over large range of distance and power. This is evidence against a monotonically-increasing SE in an ORN. Then, we explored the potential of NFT-based transmission schemes in both simulations and experiments. Results include 1) the estimation of per-sample capacity of NFT-based optical network in both ideal and realistic fibre infrastructure. Particularly, in the ideal fibre infrastructure,

the NFDM scheme showed monotonically increasing SE over the simulated signal power range. 2) the development of dual-polarisation b-modulated NFT scheme using continuous spectrum. The proposed system achieved the highest data rate among existing NFT-based systems in both simulation and experiment. Finally, we developed an alternative method to assess a large-scale NFDM network without running the time-consuming simulation of NFDM network.

The thesis has three original equations of my invention, i.e. (3.4), (3.6) and (6.9). The rest of the mathematics involved here serves as a necessary background as well as the guidance for the reproduction of relevant results.

1.4 Thesis outline

The remainder of the thesis is organized as follows. Chapter 2 introduces the prerequisites of optical fibre communication systems, including the structure of coherent transceivers, mathematical fundamentals of optical fibre channel, and the information-theoretical tools for the evaluation of various digital signal processing schemes.

Chapter 3 describes the results on the spectral broadening of Gaussian processes in optical fibre channel. Iterative Kolmogorov-Zakharov model was used to efficiently estimate the power spectral density at the output of the optical fibre. It corresponds to the publication IV listed in Sec. 1.5.

Chapter 4 gives a brief introduction of NFT and its inverse. Numerical methods for NFT and inverse non-linear Fourier transform (INFT) are detailed and the chapter is closed with a review of recent NFT-based optical communication systems.

Chapter 5 and 6 are comprehensive investigations of NFT-based communication systems, from the simplest and ideal case – one symbol per channel use and lossless – to the realistic single-channel experimental verification with data rate of 220 Gbit/s. They correspond to the publications I, II, III and VII listed in Sec. 1.5. The evolution of NFT-based communication system in this thesis is depicted in Fig. 1.8.

The thesis is summarised in Chapter 7 with outlook into future research.

1.5 List of publications

Some of the research leading to this thesis has resulted in the following publications:

Journal Articles

I X. Yangzhang, D. Lavery, P. Bayvel and M. I. Yousefi, “Impact of perturbations on non-linear frequency-division multiplexing," *in Journal of Lightwave*

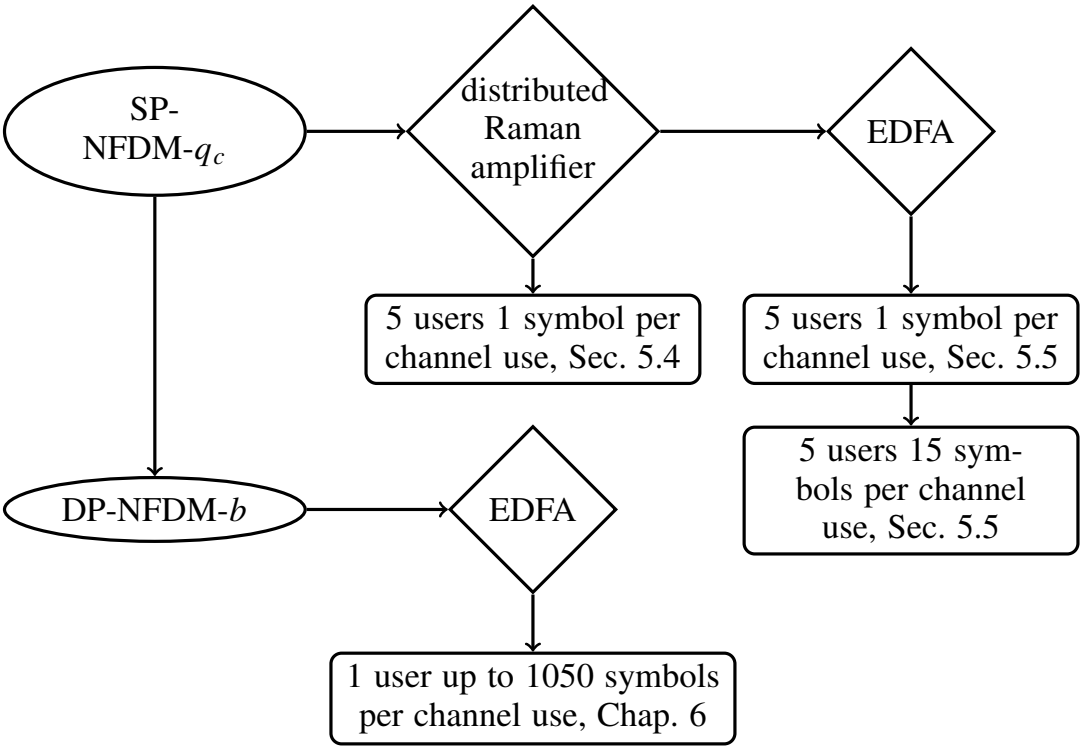

Figure 1.8: The evolution of NFT-based systems in this thesis.

Technology, vol. 36, no. 2, pp. 485–494, Jan. 2018.

II X. Yangzhang, V. Aref, S. T. Le, H. Buelow, D. Lavery and P. Bayvel "Dual-Polarization Non-Linear Frequency-Division Multiplexed Transmission With *b*-Modulation," *in Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1570-1578, Mar. 2019.

III X. Yangzhang, S. T. Le, V. Aref, H. Buelow, D. Lavery and P. Bayvel, "Experimental Demonstration of Dual-Polarization NFDM Transmission With *b*-Modulation", *IEEE Photonics Technology Letters*, vol. 31, no. 11, pp. 885-888, Apr. 2019.

Conference Papers

IV X. Yangzhang, M. I. Yousefi, A. Alvarado, D. Lavery and P. Bayvel "Nonlinear frequency-division multiplexing in the focusing regime," *in Proceeding of Optical fibre Communication Conference*, Los Angeles, CA, USA, Mar. 2017, Paper Tu3D. 1.

V X. Yangzhang, V. Aref, S. T. Le, H. Buelow and P. Bayvel, "400 Gbps dual-polarisation non-linear frequency-division multiplexed transmission with b-

modulation," *in Proceeding of European Conference on Optical Communication*, Rome, Italy, Sep. 2018, Paper We4F.1.

VI X. Yangzhang, V. Aref, H. Buelow and P. Bayvel, "Spectral Broadening of Gaussian Process in Optical Fibre and Implication on the Spectral Efficiency," *in Proceeding of European Conference on Optical Communication*, Rome, Italy, Sep. 2018, Paper Th2.38.

VII X. Yangzhang, D. Lavery and P. Bayvel, "Inter-channelInterference in Non-linear Frequency-division Multiplexed Networks on Fibre Links with Lumped Amplication," *in Proceeding of Conference on Lasers and Electro-Optics/Europe – European Quantum Electronics Conference*, Munich, Germany, Jul. 2019, Paper CI-P.8.

2

The fundamentals of optical fibre communications

This chapter establishes the fundamentals of optical fibre communication systems and introduces information-theoretical tools for the evaluation of various digital signal processing schemes. The goal is to provide firm technical ground for the later contributional chapters so that their results can be reproduced with the help of equations in this chapter. Readers should feel free to skip this chapter and will be referred back later when more details are needed. We start with the fundamentals of optical fibre communications, covering the aspects of system components, component physics, signal impairments, and focus only on the coherent transmission system. A coherent transmission system consists of an optical transmitter, an optical fibre channel, and a digital coherent receiver. A generic block diagram of such a system is shown in Fig. 2.1. In the following sections, we explain the physical mechanisms of each block.

2.1 Optical transmitter

A typical optical transmitter in a coherent optical communication system is comprised of four parts: a digital signal processing unit, DACs, a PDM IQM and a semiconductor laser as shown in Fig. 2.2. To generate an optical signal, the DSP unit provides the sampled complex-valued waveform in the base-band, then four DACs

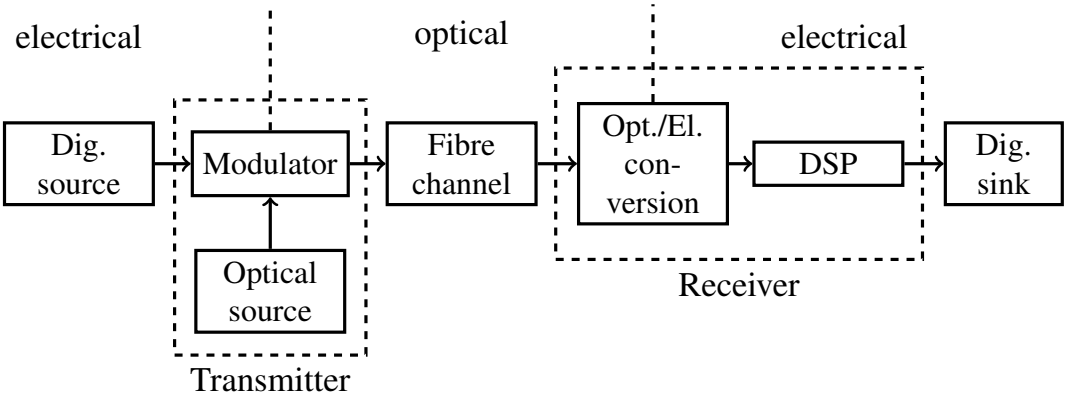

Figure 2.1: Generic block diagram of a coherent optical communication systems.

convert the signals into analog ones, corresponding to the in-phase and quadrature signals in x- and y-polarisations, to drive the modulator to modulate the monochromatic optical signal from the laser. This section explains the working principles of the optical parts in such a transmitter, i.e., the semiconductor laser and the PDM IQM. Particularly for the semiconductor laser, the external cavity semiconductor laser (ECSL) will be taken as an example because it will be used in the later experiment in Sec. 6.7.

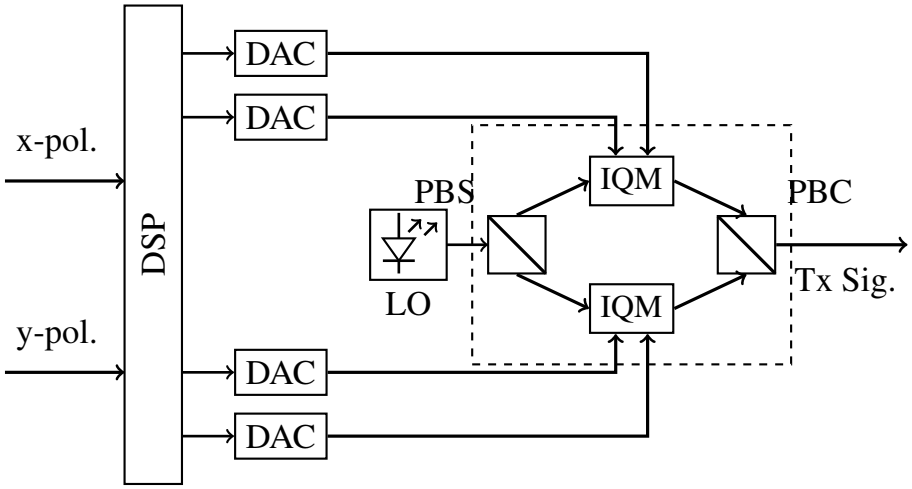

Figure 2.2: The schematic of a polarisation diversity optical transmitter unit. It consists of four DACs, a PBS, two optical IQM and a PBC.

2.1.1 External cavity semiconductor lasers

Here we give a short introduction of the ECSL assuming a basic understanding of the principles of semiconductor laser operation [51, Chap. 3]. The ECSL is a type of semiconductor laser that couples the laser cavity to an external cavity to reflect a portion of the escaped light back into the laser cavity. A typical way to realise

the external cavity is an external grating as shown in Fig. 2.3. To provide a strong coupling between the laser cavity and the external cavity, one normally puts an anti-reflection coating on the laser facet facing the grating to reduce its reflectivity. The external cavity in an ECSL provides two advantages when compared with other types of single-mode semiconductor lasers, i.e., wavelength tunability through rotating the grating (typically 50 nm) and low line-width ($\sim$ 100 kHz). The longer resonator increases the damping time of the intracavity light and thus allows for lower phase noise and a smaller emission linewidth. The wavelength tunibility is achieved by turning the external grating that functions as a tunable bandpass filter. Wavelength tunability is a desirable feature for lasers used in WDM communication systems.

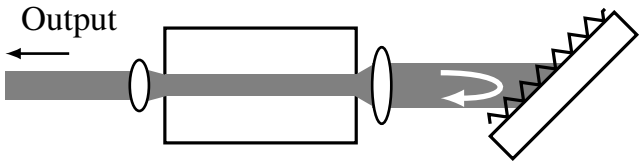

Figure 2.3: A schematic of the external cavity laser with an external grating.

2.1.2 PDM I/Q modulator

A PDM-IQM is depicted in the dashed box in Fig. 2.2. A PBS split the light source into two orthogonal polarised lights, each going into an optical IQM. The two signals are then modulated by the IQM, and subsequently combined into one by a PBC. The output of the PBC will then be fed into the optical fibre as the transmitted optical signal. We open up the individual IQM and look closely into its working principle. As shown in Fig. 2.4, an IQM is comprised of a $\pi/2$ phase shift and two

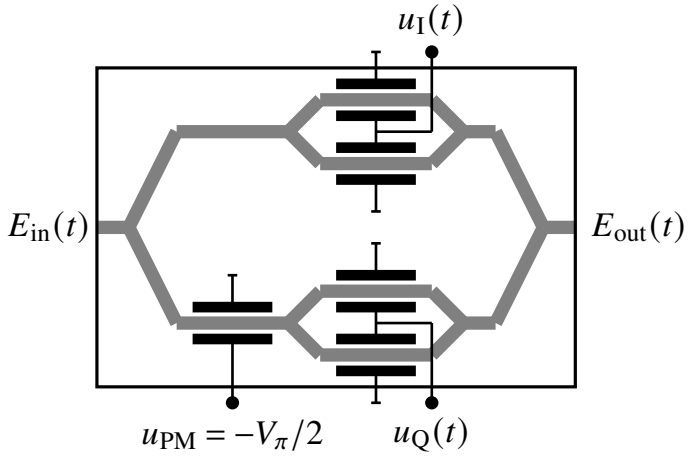

Figure 2.4: A schematic of an IQM that consists of two Mach-Zehnder modulators.

Mach-Zehnder modulators (MCMs) for in-phase and quadrature components. The MCMs are operating in push-pull mode and biased at minimum transmission point. The transfer function of an individual MCM is

$$\frac{E_{\mathrm{Mout}}(t)}{E_{\mathrm{Min}}(t)} = \cos\left(\frac{u(t)}{2V_\pi}\pi\right), \tag{2.1}$$

where $E_{\mathrm{Min}}(t)$ is the optical field from the laser source, $E_{\mathrm{Mout}}(t)$ is the output signal of MCM, and $u(t)$ is the electrical base-band signal. Together with the $\pi/2$ phase shift, the overall transfer function of an IQM is

$$\frac{E_{\mathrm{out}}(t)}{E_{\mathrm{in}}(t)} = \frac{1}{2}\cos\left(\frac{u_{\mathrm{I}}(t)}{2V_\pi}\pi\right) + e^{j\frac{\pi}{2}}\frac{1}{2}\cos\left(\frac{u_{\mathrm{Q}}(t)}{2V_\pi}\pi\right), \tag{2.2}$$

where $E_{\mathrm{in}}(t)$ is the input optical signal from the laser source, $E_{\mathrm{out}}(t)$ is the output optical signal of the IQM, and $u_{\mathrm{I}}(t)$ and $u_{\mathrm{Q}}(t)$ are the electrical in-phase and quadrature base-band signals.

2.1.3 Transmitter impairments

The various parts mentioned above in an optical transmitter introduce unwanted distortions to the signal. A careful analysis of these distortions will help to estimate the transceiver noise and to design DSP techniques to compensate them. The following discussion will form the basis of part of the DSP chain in Fig. 6.12.

DAC resolution The DAC introduces quantisation noise, timing error from its clock jitter, and other distortion particularly for broadband signal. The combined effect is measured through the signal-to-noise and distortion (SINAD) ratio

$$\mathrm{SINAD} = \frac{P_{\mathrm{signal}} + P_{\mathrm{noise}} + P_{\mathrm{distortion}}}{P_{\mathrm{noise}} + P_{\mathrm{distortion}}}, \tag{2.3}$$

where P_{signal}, P_{noise}, and $P_{\mathrm{distortion}}$ stand for the power of signal, noise and distortion, respectively. A more commonly-used figure of merit is the effective number of bits (ENOB). It can be quickly converted from SINAD by

$$\mathrm{ENOB} = \frac{\mathrm{SINAD\ (dB)} - 1.76}{6.02}, \quad \mathrm{bit}. \tag{2.4}$$

Note that the ENOB/SINAD also depends on the bandwidth of the input signal [52].

Laser noise Even though a laser is biased at a constant current with negligible fluctuations, its output still fluctuates in intensity, phase, and frequency [51, Sec. 3.3.3]. The noises originate from two fundamental mechanisms: spontaneous emission and electron-hole recombination (shot noise). We are mostly concerned with the phase

noise of the laser that causes a finite spectral line-width when lasers are operated at a constant current. The fluctuation of central frequency of laser Δv is described by the changes in optical phase θ [53, Eq. (4)]

$$\Delta v = \frac{1}{2\pi}\frac{d\theta}{dt}. \tag{2.5}$$

In numerical simulation, the laser phase noise is normally modelled as a GN with a variance of $2\pi dt\Delta f$, where Δf is the laser line-width.

I/Q imbalance and delay skew If the IQM at transmitter is improperly biased, the signals at I- and Q-ports have phase/gain mismatch in between. In addition, the I/Q delay skew refers to the timing mismatch caused by the difference in the physical path length of the circuitry. These effects can be summarised into one equation [54, Eq. (1)]

$$d(t) = \left[g_{\mathrm{I}}\,\mathrm{Re}\left\{s(t-\frac{\tau_{\mathrm{IQ}}}{2})\right\} + o_{\mathrm{I}}\right] + \left[g_{\mathrm{Q}}\,\mathrm{Im}\left\{s(t+\frac{\tau_{\mathrm{IQ}}}{2})\right\} + o_{\mathrm{Q}}\right] e^{j(\frac{\pi}{2}+\phi_{\mathrm{IQ}})}, \tag{2.6}$$

where $s(t)$ and $d(t)$ are the ideal and distorted optical signals, g_{I} and g_{Q} the I/Q gains, o_{I} and o_{Q} the offsets in I/Q paths, τ_{IQ} the timing mismatch and ϕ_{IQ} the quadrature phase error. Each of the effects on the constellation is visualised in Fig. 2.5. Compensating the IQ imbalance is normally the first step in the receiver (Rx) DSP, and will be used in the experiment described in Sec. 6.7.

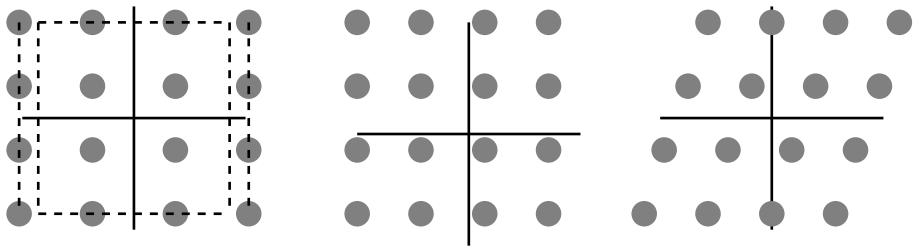

Figure 2.5: IQ imbalance at transmitter induced by (left) Gain Error g_{I}, g_{Q}, (middle) Offset o_{I}, o_{Q}, and (right) Quadrature Error ϕ_{IQ}.

2.2 Physics of optical fibre channel

In this section, the mathematics of optical fibre channel and its effects on the pulse propagation are introduced, with the hope that if one wishes to simulate optical fibre systems in later sections, one can simply write the simulation codes according to the equations in this section. The non-linear Schrödinger equation (NLSE), the coupled non-linear Schrödinger equation (CNLSE), and the Manakov-PMD equation will be described in the mentioned order as we include more effects in the channel model

and approach the realistic optical fibre channel.

2.2.1 Wave propagation equations – NLSE

The NLSE describes the wave propagation of a single-polarisation signal in a idealised single-mode, single-polarisation optical fibre with dispersion. We present the NLSE without deriving it from the Maxwell's equations.

$$\frac{\partial A}{\partial z} = -\beta_1 \frac{\partial A}{\partial t} - j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{\alpha}{2}A + j\gamma|A|^2 A, \tag{2.7}$$

where $j = \sqrt{-1}$, $A = A(z,t)$ denotes the slowly-varying complex envelope of the signal in base-band, z and t are the distance along the fibre and the time, respectively. γ is the non-linear coefficient in the unit of $(\mathrm{W}\cdot\mathrm{km})^{-1}$. α represents the attenuation coefficient of optical fibre in the unit of km^{-1} or dB/km, with the conversion relation α (dB/km) = 4.343α (km^{-1}). For example, the industry standard Corning SMF-28 ultra optical fibre has an attenuation coefficient of 0.18 dB/km, which means the optical signal loses 4.61% of its power every kilometre of propagation. $\beta_1(\omega)$ or its inverse $v_g = 1/\beta_1(\omega)$ specifies the velocity at which a specific spectral component travels in the fibre. This velocity v_g (m/s) is commonly referred to as the *group velocity*. It can also be calculated from the refractive index n of fibre's core $v_g = c/(n + \omega\frac{dn}{d\omega})$. Since the group velocity is frequency-dependent, the different spectral component of an optical pulse will not arrive simultaneously at the optical fibre output, leading to a pulse broadening in the time domain for unchirped pulses. The extent of pulse broadening in a fibre of length L can be estimated by

$$\Delta T = L\beta_2 W, \tag{2.8}$$

where W is the signal bandwidth, and $\beta_2 = d\beta_1/d\omega$ in the unit of (s^2/m) is the group velocity dispersion (GVD). The pulse broadening behavior could be changed into contraction if the pulse has an initial frequency chirp. In the case of linearly chirped Gaussian pulse, it has the form of

$$A(0,t) = \exp\left(-\frac{1+jC}{2}\frac{t^2}{T_0}\right), \tag{2.9}$$

where T_0 is the half-width of the Gaussian pulse and C is the chirp parameter. For $C < 0$, the chirped Gaussian pulse will contract first and broaden again depending on the propagation distance.

Another commonly-used parameter for dispersion is the *dispersion parameter*

$$D = -\frac{2\pi c}{\lambda^2}\beta_2, \quad \text{(ps/km-nm)}, \tag{2.10}$$

where c denotes the speed of light in vacuum, λ represents the corresponding wavelength of the spectral component. If the bandwidth of the optical signal is described in terms of the range of wavelengths $\Delta\lambda$ rather than the range of frequencies, it is convenient to use dispersion parameter to estimate the pulse broadening as shown below

$$\Delta T = DL\Delta\lambda. \tag{2.11}$$

In addition, it is a common practice to simplify (2.7) by introducing a reference frame moving together with the central frequency component of the signal at the speed of $v_g(\omega_0)$. This can be accomplished by a simple change of variables

$$t' = t - \beta_1 z, \;\; z' = z, \tag{2.12}$$

and after cleaning up, the NLSE writes

$$\frac{\partial A}{\partial z} = - \underbrace{j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2}}_{\text{chromatic dispersion}} - \underbrace{\frac{\alpha}{2}A}_{\text{attenuation}} + \underbrace{j\gamma|A|^2 A}_{\text{non-linearity}}. \tag{2.13}$$

(2.13) will be our master equation for single-polarisation channel analysis.

2.2.2 Linear and non-linear propagation effects

From (2.13) we can identify three propagation effects in the optical fibre channel: attenuation, chromatic dispersion and non-linearity. A common way to study these effects is by setting the other terms to zero and solving the residual equation analytically. Here we will look at the effect of chromatic dispersion and non-linearity more closely.

Chromatic dispersion Setting $\alpha = 0,\ \gamma = 0$ in (2.13), the residual equation writes

$$\frac{\partial A}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2}. \tag{2.14}$$

Let $\tilde{A}(z,\omega)$ be the Fourier transform of time domain signal $A(z,t)$ given by

$$\tilde{A}(z,\omega) \triangleq \int_{-\infty}^{\infty} A(z,t)e^{-j\omega t}dt. \tag{2.15}$$

The residual equation in frequency domain reads

$$\frac{\partial \tilde{A}}{\partial z} = j\frac{\beta_2 \omega^2}{2}\tilde{A}, \tag{2.16}$$

whose solution is

$$\tilde{A}(z,\omega) = \tilde{A}(0,\omega)\exp\left(j\frac{\beta_2\omega^2}{2}z\right). \tag{2.17}$$

The chromatic dispersion alone is just an all-pass filter effect.

Non-linearity Setting $\alpha = 0,\ \beta_2 = 0$, the residual equation reads

$$\frac{\partial A(z,t)}{\partial z} = j\gamma|A(z,t)|^2 A(z,t). \tag{2.18}$$

It is customary to treat the equation in the frequency domain. The residual equation in the frequency domain reads

$$\begin{aligned}
\frac{\partial \tilde{A}(z,\omega)}{\partial z} &= j\gamma \tilde{A}(z,\omega) \circledast \tilde{A}^*(z,-\omega) \circledast \tilde{A}(z,\omega), \qquad (2.19)\\
\text{(suppress } z) \quad &= j\gamma \int_{-\infty}^{\infty} \tilde{A}(s)\tilde{A}^*(s-\omega)ds \circledast \tilde{A}(\omega),\\
&= j\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{A}(\omega-p)\tilde{A}(s)\tilde{A}^*(s-p)dsdp,\\
(\omega_1 = \omega - p\,, \omega_2 = s) \quad &= j\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{A}(\omega_1)\tilde{A}(\omega_2)\tilde{A}^*(\omega_1+\omega_2-\omega)d\omega_{12},\\
&\triangleq N(\tilde{A})(z,\omega),
\end{aligned}$$

where $\circledast$ denotes the convolution operation and * represents complex conjugation. It can be seen from the above equation that the non-linear term in NLSE causes complex interactions between different frequency components of the signal. To have a detailed insight of the non-linear interaction, one can rewrite the residual equation in the discrete Fourier domain for T-periodic signals. Slightly different from the Fourier transform, the Fourier series of a T-periodic signal is defined as

$$\tilde{A}_k \triangleq \frac{1}{T}\int_{T/2}^{T/2} A(t)e^{-j\frac{2\pi kt}{T}}dt. \tag{2.20}$$

In the discrete Fourier domain, the residual equation (2.18) becomes [55,

Eq. (4.3.2)] [56, Eq. (19)]

$$\begin{aligned}
\frac{\partial \tilde{A}_k}{\partial z} &= \sum_{\ell}\sum_{m} \tilde{A}_\ell \tilde{A}_m \tilde{A}^*_{\ell+m-k} \\
&= \left(\sum_{(\ell=k) \text{ or } (m=k)} + \sum_{(\ell\neq k) \text{ and } (m\neq k)} \right) \tilde{A}_\ell \tilde{A}_m \tilde{A}^*_{\ell+m-k} \\
&= \left(\sum_{\ell=k} + \sum_{m=k} - \sum_{(\ell=k) \text{ and } (m=k)} + \sum_{(\ell\neq k) \text{ and } (m\neq k)} \right) \tilde{A}_\ell \tilde{A}_m \tilde{A}^*_{\ell+m-k} \\
&= 2j\gamma \tilde{A}_k \sum_{m} |\tilde{A}_m|^2 - j\gamma |\tilde{A}_k|^2 \tilde{A}_k + j\gamma N_k(\tilde{A})(z) \\
&= \underbrace{2j\gamma \tilde{A}_k \sum_{m\neq k} |\tilde{A}_m|^2}_{\text{XPM}} + \underbrace{j\gamma |\tilde{A}_k|^2 \tilde{A}_k}_{\text{SPM}} + \underbrace{j\gamma N_k(\tilde{A})(z)}_{\text{FWM}},
\end{aligned} \tag{2.21}$$

where k is the discrete frequency index with a frequency interval $\omega_0 = 1/T$. $\tilde{A}_k$ is the kth coefficient of the Fourier series of the periodic time domain signal, and $N_k(\tilde{A})(z)$ is given below

$$\begin{aligned}
N_k(\tilde{A})(z) &\triangleq \sum_{\ell mn \in nr_k} \tilde{A}_\ell(z) \tilde{A}_m(z) \tilde{A}^*_n(z), \\
nr_k &\triangleq \left\{ (\ell, m, n) | \ell + m = n + k, \ell \neq k, m \neq k \right\}.
\end{aligned} \tag{2.22}$$

As underlined in (2.21), the non-linear interactions are categorised into three types, i.e., self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM). Based on the frequency components that each type of non-linear interaction involves, the naming is quite self-evident. Also note that all types of non-linear interactions only induce phase-shift on the signal.

Perturbation solution We have separately inspected the effects of chromatic dispersion and non-linearity, the combined effect will only be more complicated. For an arbitrary initial condition, NLSE does not necessarily have a closed-form solution. Therefore, one usually looks at its perturbation solution in frequency domain to the first-order. The NLSE in frequency domain reads

$$\frac{\partial \tilde{A}(z,\omega)}{\partial z} = -\frac{\alpha(\omega)}{2} \tilde{A}(z,\omega) - j\frac{\beta_2 \omega^2}{2} \tilde{A}(z,\omega) + j\gamma N(\tilde{A})(z,\omega). \tag{2.23}$$

According to Duhamel's formula [57], (2.23) can be written as an integral

$$\tilde{A}(z,\omega) = e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\tilde{A}(0,\omega) + j\gamma\int_0^z e^{(z-z')(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})}N(\tilde{A})(z',\omega)dz'. \tag{2.24}$$

Ignoring the second term, we have the zero-order perturbation solution of (2.24)

$$\tilde{A}^{(0)}(z,\omega) = e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\tilde{A}(0,\omega). \tag{2.25}$$

Substitute the zero-order solution (2.25) into the non-linear term in (2.24)

$$\begin{aligned}\tilde{A}^{(1)}(z,\omega) =& e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\tilde{A}^{(0)}(0,\omega) + j\gamma\int_0^z e^{(z-z')(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})}N(\tilde{A}^{(0)})(z',\omega)dz' \\ =& e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\tilde{A}^{(0)}(0,\omega) + j\gamma\int_0^z e^{(z-z')(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \\ & e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega_1^2}{2})z'}\tilde{A}(0,\omega_1)e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega_2^2}{2})z'}\tilde{A}(0,\omega_2) \\ & \left(e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega_3^2}{2})z'}\tilde{A}(0,\omega_3)\right)^*\delta(\omega_1+\omega_2-\omega_3-\omega)dz'd\omega_{123}.\end{aligned} \tag{2.26}$$

Substitute the following equations

$$\Omega_{123\omega} = \frac{\beta_2}{2}(\omega_1^2+\omega_2^2-\omega_3^2-\omega^2) \tag{2.27}$$

$$\delta_{123\omega} = \delta(\omega_1+\omega_2-\omega_3-\omega) \tag{2.28}$$

into (2.26), it becomes

$$\begin{aligned}\tilde{A}^{(1)}(z,\omega) =& e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\Bigg\{\tilde{A}^{(0)}(0,\omega) + j\gamma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{A}(0,\omega_1)\tilde{A}(0,\omega_2)\tilde{A}^*(0,\omega_3) \\ & \delta_{123\omega}\left(\int_0^z e^{-\alpha z'+jz'\Omega_{123\omega}}dz'\right)d\omega_{123}\Bigg\}.\end{aligned} \tag{2.29}$$

Simplifying by substituting the following

$$H(\Omega_{123\omega},z) = \frac{e^{-z\alpha+jz\Omega_{123\omega}}-1}{-\alpha+j\Omega_{123\omega}} \tag{2.30}$$

into (2.29), we finally have the first-order perturbation solution of the NLSE

$$\tilde{A}^{(1)}(z,\omega) = e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\Bigg\{\tilde{A}^{(0)}(0,\omega)+ j\gamma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\Omega_{123\omega},z)\tilde{A}(0,\omega_1)\tilde{A}(0,\omega_2)\tilde{A}^*(0,\omega_3)\delta_{123\omega}d\omega_{123}\Bigg\}. \quad (2.31)$$

In the conventional optical communication systems, signals are transmitted in the relatively low-power and hence linear regime. The signal represented by the second term in (2.31) is often treated as unwanted additive GN whose power can be easily estimated using the above equation. Based on the perturbation approach and the Gaussian signal assumption, the well-known GN model for optical fibre communication was developed [58] and is widely-used for fast estimation of quality of transmission.

2.2.3 The evolution of power spectral density

In the previous section, we have seen the effects of the optical fibre channel on the input signal either in time or frequency domain. Due to the non-linear nature of the channel, it is also important to see the channel effects on the PSD (defined in Appx. B) as new frequency components are generated during propagation. This seemingly trivial effect has a huge impact on the information-theoretic analysis of the optical fibre channel. In a linear channel, where the spectrum does not change, for a band-limited digital signal with a duration D and bandwidth W, we consider the signal of having $2DW$ discrete DoFs. All the information must be encoded on these fixed number of DoFs, and the channel capacity analysis is also based on them too. However, if the spectrum changes significantly during propagation, it is not possible to digitise the signal properly. The signal varies in dimensions and no analysis is valid at that point.

There are two methods available to estimate the signal spectrum at any point of the transmission link: 1) solve the NLSE with the perturbation approach and calculate the signal spectrum with Gaussian assumption on the input signal. This is the basis of the GN model [58], 2) convert the NLSE into kinetic equation and solve it to the first-order in non-linearity with quasi-Gaussian (at most 4-point cumulants are significant [56, Eq. (10)]) assumption on the input signal. This is the basis of the Kolmogorov-Zakharov (KZ) model [56].

Gaussian Noise Power Spectral Density

Recall the first-order perturbation solution of NLSE in the frequency domain

derived in the previous subsection

$$\tilde{A}^{(1)}(z,\omega) = e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\Bigg\{\tilde{A}^{(0)}(0,\omega)+ j\gamma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\Omega_{123\omega},z)\tilde{A}(0,\omega_1)\tilde{A}(0,\omega_2)\tilde{A}^*(0,\omega_3)\delta_{123\omega}d\omega_{123}\Bigg\}. \quad (2.32)$$

We denote the PSD as $S(z,\omega)$ and calculate it (see Appx. C) according to its definition with the Gaussian assumption on the input signal $\tilde{A}^{(0)}(0,\omega)$

$$\begin{aligned} S(z,\omega)^{GN} =& \mathbb{E}\left[\tilde{A}^{(1)}(z,\omega)\tilde{A}^{(1)*}(z,\omega)\right] \\ =& e^{-\alpha z}S^{(0)}(0,\omega)+ \\ & 2e^{-\alpha z}\gamma^2\int |H(\Omega_{123\omega},z)|^2 S(0,\omega_1)S(0,\omega_2)S(0,\omega_3)\delta_{123\omega}d\omega_{123}. \end{aligned} \quad (2.33)$$

If EDFAs are present in the link that consists of N spans with span length L_{sp}, the GN PSD changes to [56, Eq. (58)]

$$S(z,\omega)^{\text{GN}} = S^{(0)}(0,\omega) + 2\gamma^2\int |\tilde{H}(\Omega_{123\omega},z)|^2 S(0,\omega_1)S(0,\omega_2)S(0,\omega_3)\delta_{123\omega}d\omega_{123} \quad (2.34)$$

$$\tilde{H}(\Omega_{123\omega},z) = H(\Omega_{123\omega},L_{\text{sp}})\frac{\sin(z\Omega_{123\omega}/2)}{\sin\left(L_{\text{sp}}\Omega_{123\omega}/2\right)}e^{-j\Omega_{123\omega}(z-L_{\text{sp}})}. \quad (2.35)$$

Kolmogorov-Zakharov Power Spectral Density

The Kolmogorov-Zakharov model of optical fibre was developed in the framework of weak wave turbulence (WWT). To roughly define the weak non-linear regime, it was proposed [56, 59] to use the ratio of non-linear and linear parts of the NLSE Hamiltonian (preserved quantity)

$$H(z) \triangleq j\int_{-\infty}^{\infty}\left(|\partial_t A(t,z)|^2 - \frac{\gamma}{2}|A(t,z)|^4\right)dt. \quad (2.36)$$

The ratio of non-linear and linear parts, denoted as κ, can be then written as

$$\kappa = \frac{\int_{-\infty}^{\infty}\frac{\gamma}{2}|A(t,z)|^4 dt}{\int_{-\infty}^{\infty}|\partial_t A(t,z)|^2 dt}. \quad (2.37)$$

When $\kappa < 0.1$, the system is in the pseudo linear regime. $\kappa > 0.1$ indicates the non-linear regime. It is assumed in the application of WWT that $\kappa << 1$ [60].

We rewrite the important steps in the derivation of the kinetic equation of NLSE in its discrete version [56]

$$\frac{dS_k(z)}{dz} = \mathbb{E}\left[\tilde{A}_k^*(z)\frac{d\tilde{A}_k(z)}{dz} + c.c.\right], \qquad c.c. \text{ for complex conjugate} \tag{2.38}$$
$$= \mathbb{E}\left[\tilde{A}_k^*(z)\left(-\frac{\alpha(z,\omega)}{2}\tilde{A}_k(z) - j\frac{\beta_2\omega^2}{2}\tilde{A}_k(z) + j\gamma N(\tilde{A})_k(z)\right) + c.c.\right],$$

and after many steps described in [56, Eq. (40)]

$$= 2\gamma \sum_{\ell mn \in nr_k} \mathrm{Re}\left(e^{jz\tilde{\Omega}_{\ell mnk}} S_{\ell mnk}(0)\right) +$$
$$4\gamma^2 \sum_{\ell mn \in nr_k} \int_0^z \cos\left(\tilde{\Omega}_{\ell mnk}(z-z')\right) T_{\ell mnk}(z')dz', \tag{2.39}$$

where

$$\tilde{\Omega}_{\ell mnk} = j\alpha - \Omega_{\ell mnk}, \tag{2.40}$$
$$\Omega_{\ell mnk} = \frac{\beta_2\omega_0^2}{2}(\ell^2 + m^2 - n^2 - k^2) \tag{2.41}$$
$$T_{\ell mnk}(z) = S_\ell(z)S_m(z)S_n(z) + S_\ell(z)S_m(z)S_k(z)$$
$$- S_\ell(z)S_n(z)S_k(z) - S_m(z)S_n(z)S_k(z). \tag{2.42}$$

(2.39) is the kinetic equation for $S_k(z)$. It is again a cubic non-linear equation similar to the NLSE. Now the rapidly varying time-domain signals are averaged out and the PSD evolves very slowly along distance. Hence, the perturbation approach will work better in this regime than in the coherent signal regime as in the GN model. Solving (2.39) to its first-order in perturbation, we obtain [56, Eq. (56)]

$$S_k^{KZ}(z) = e^{-\alpha z}S_k(0) + 2\gamma^2 \sum_{\substack{\ell \neq k \\ m \neq k}} |H_{\ell mnk}(z)|^2 T_{\ell mnk}(0)\delta_{\ell mnk}, \tag{2.43}$$
$$H_{\ell mnk}(z) = \frac{-1 + e^{-z\alpha + jz\Omega_{\ell mnk}}}{-\alpha + j\Omega_{\ell mnk}}. \tag{2.44}$$

Similarly, EDFAs can also be included in the KZ PSD, it changes to [56, Eq. (62)]

$$S_k^{KZ}(z) = S_k(0) + 2\gamma^2 \sum_{\substack{\ell \neq k \\ m \neq k}} |\tilde{H}_{\ell mnk}(z)|^2 T_{\ell mnk}(0)\delta_{\ell mnk}, \tag{2.45}$$
$$\tilde{H}_{\ell mnk}(z) = H_{\ell mnk}(L_{\mathrm{sp}}) \frac{\sin(z\Omega_{\ell mnk}/2)}{\sin\left(L_{\mathrm{sp}}\Omega_{\ell mnk}/2\right)} e^{-j\Omega_{\ell mnk}(z - L_{\mathrm{sp}})}. \tag{2.46}$$

Comparison between the estimated output PSDs from the GN and KZ models and the actual output PSD from simulations is plotted in Fig. 2.6. Wave propagations were simulated in the NLSE (4.4) for illustrative purpose only. The input PSD is obtained by averaging spectra of 4000 randomly generated signals, denoted as S_0. The random signals are 32 GBaud root-raised-cosine (RRC)-pulses with a roll-off factor of 0.5. Their average power is 12 dBm. Each of the signals is sent through a 1600 km EDFA-amplified optical fibre link with 80 km span length. The received output spectra are averaged to obtain the output PSD, denoted as S_{SIM}.The input PSD in Fig. 2.6 has an average κ of 0.49, indicating that the signals are in the nonlinear regime. Fig. 2.6 shows that the GN model significantly (with 1.3 times power as the input PSD) overestimate the PSD at the output, while the PSD from the KZ model is reasonably close to the simulated PSD. Although in WWT the assumption is $\kappa << 1$, input PSDs with intermediate κ values do not break down the KZ model. Similar phenomenon was also observed in [60, Fig. 3].

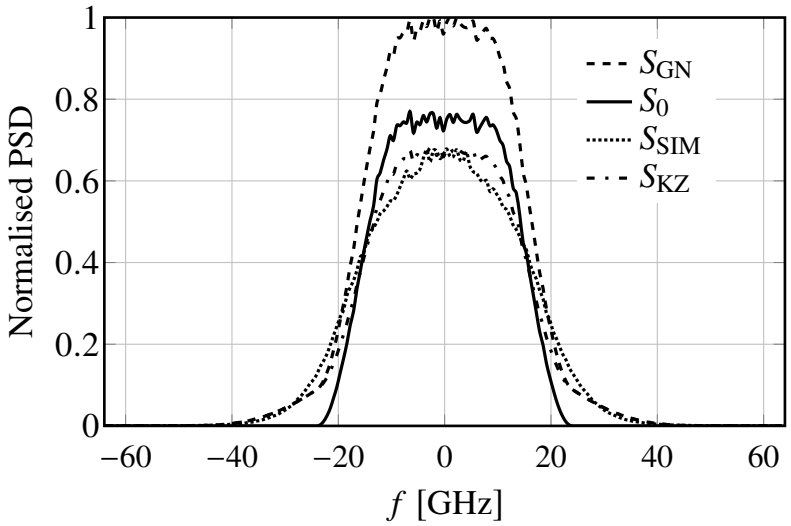

Figure 2.6: PSDs from simulations scaled by the maximum of the GN PSD at 12 dBm.

With the above basic understanding of the KZ model, Chap. 3 will describe the extension of KZ model to estimate the spectral evolution in optical fibre over a broader range of system parameters.

2.2.4 Polarisation effects

In a standard single-mode fibre (e.g. SMF-28), signals can be present at two orthogonal polarisations. They excite the fundamental HE_{11} mode of the fibre and their transverse spatial distribution does not change during propagation [55, Eq. (3.11)]. However, the two signals often experience different group velocities and change their state-of-polarisation (SoP) in a random fashion because the refractive index in

the fibre core is often cylindrically asymmetric (called birefringence) due to imperfection in core shape and external stress. This is commonly referred to as the PMD. In addition, the signals on the orthogonal polarisations also interfere with each other in a non-linear way. These related polarisation effects and their impact will be under investigation later in Sec. 5.5.2 and Chap. 6 when dual-polarisation signals are transmitted. The mathematical foundation is laid out here first. The coupled vector signal propagation in optical fibre is governed by the CNLSE [6, Eq. (1)]

$$\begin{aligned}\frac{\partial \mathbf{A}}{\partial z} =&(-\frac{\alpha}{2}-\boldsymbol{\beta_1}\frac{\partial}{\partial t}-j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}+\frac{\beta_3}{6}\frac{\partial^3}{\partial t^3})\mathbf{A}+\\ &j\gamma\left[|\mathbf{A}|^2\mathbf{I}-\frac{1}{3}(\mathbf{A}^H\sigma_3\mathbf{A}\sigma_3)\right]\mathbf{A}\\ =&(\hat{\mathbf{D}}+\hat{\mathbf{N}})\mathbf{A}\end{aligned} \tag{2.47}$$

where $\mathbf{A} = [A_x(z,t)\ A_y(z,t)]^T$ contains signals in the x- and y-polarisations as a function of distance z and time t. In this equation, we consider only the first-order PMD and ignore the difference in attenuation and second- and third-order dispersion at arbitrary polarisation. Hence, α, β_2 and β_3 are scalars, $\boldsymbol{\beta_1}$ is a matrix related to PMD and will be explained later. $\hat{\mathbf{D}}$ and $\hat{\mathbf{N}}$ are just a shorthand to the linear and non-linear operators, respectively. H represents the Hermitian transpose. The Pauli spin matrix [61, Eq. 2.6] σ_3 is given by

$$\sigma_3 = \begin{bmatrix}0 & -j\\ j & 0\end{bmatrix}. \tag{2.48}$$

Due to its distributed and stochastic nature, the PMD is modelled by dividing the fibre into N_{sec} sections. One normally considers first-order PMD in each section, with arbitrary rotation in the principal state of polarisation (PSoP) between adjacent sections as illustrated in Fig. 2.7. The PSoPs is a pair of orthogonal polarisation states that, in the absence of dispersion and non-linearity, induce only first-order PMD of a pulse, preserving it from any high-order PMD [61]. The PMD in each section is written as [6, Eq. (4)]

$$\beta_1(z) = \beta_1^{(k)} = \mathbf{R}^{(k)}(\frac{\zeta_k}{2}\sigma_1)\mathbf{R}^{(k)H},\ \text{ for }\ z_{k-1}\le z\le z_k,\ \ \sigma_{\mathbf{1}} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \tag{2.49}$$

$$\mathbf{R}^{(k)} \triangleq \mathbf{R}(\theta_k,\phi_k) = \begin{bmatrix}\cos\theta_k\cos\phi_k - j\sin\theta_k\sin\phi_k & -\sin\theta_k\cos\phi_k + j\cos\theta_k\sin\phi_k\\ \sin\theta_k\cos\phi_k + j\cos\theta_k\sin\phi_k & \cos\theta_k\cos\phi_k + j\sin\theta_k\sin\phi_k\end{bmatrix} \tag{2.50}$$

where $\mathbf{R}$ is a unitary matrix that belongs to the special unitary group of degree two [62, Eq. (3.38)], denoted by $SU(2)$, with the general form $\begin{bmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{bmatrix}$. The first column of $\mathbf{R}^{(k)}$ is a PSoP in section k represented by a unit vector on the Poincaré sphere with azimuth and ellipticity angles $2\theta_l$ and $2\phi_k$, the second column can be obtained from the general structure of $SU(2)$. ζ_k is the differential group delay (DGD) per unit length in the unit of ps/m in section k and assumed to be constant within the section. The accumulated DGD per section is denoted by τ_k in the unit of ps and has a Maxwellian distribution [63].

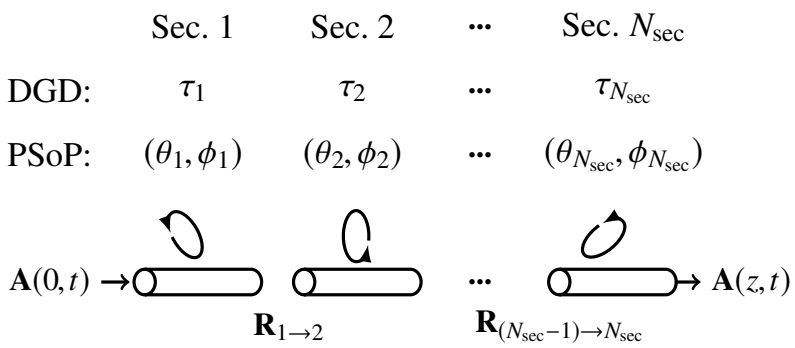

Figure 2.7: PMD, adapted from [6].

To further the analysis, we need to first define three types of length: the non-linear length, the beat length and the correlation length. To start with, the non-linear length is defined as $L_{\text{NL}} = 1/(\gamma P_0)$ where P_0 is the initial signal power. For example, if $\gamma = 2$ (W·km)$^{-1}$ and P_0 is in the range of $2 \sim 4$ mW, the non-linear length L_{NL} is around 100 km. Then the beat length is defined as $L_{\text{B}} = \lambda/|n_{\text{effx}} - n_{\text{effy}}|$ where n_{effx} and n_{effy} are the effective refractive indices in orthogonal polarisations, respectively. For a linear polarised light travelling through a birefringent fire, its SoP changes in a periodic manner from linear to elliptical and back to linear over the beat length. Typically, $|n_{\text{effx}} - n_{\text{effy}}| \sim 10^{-7}$, $L_{\text{B}} \sim 10$ m for $\lambda \sim 1\mu$m. Finally, we define the correlation length ℓ_{c} as the length scale at which the fibre birefringence changes its magnitude and orientation (with typical values in the range of 10 - 100 m). The correlation length is often used as PMD section length in simulation.

To characterize the PMD effect of a fibre link, optical fibre manufacturers often measure the total DGD τ_{tot} at the end of the optical fibre. The τ_{tot} is related to the DGD at each section through [64, Eq. (C5)]

$$\mathbb{E}[\tau_{\text{tot}}^2] = \mathbb{E}[\tau_1^2] + ...\mathbb{E}[\tau_k^2]... + \mathbb{E}[\tau_{N_{\text{sec}}}^2], \tag{2.51}$$

where τ_{tot} and τ_k are all assumed to be Maxwellian distributed. For a Maxwellian

distribution, its mean μ and variance σ^2 can be connected by $\sigma^2 = \mu^2(3\pi - 8)/8$ [65, p. 149], i.e., $\mathbb{E}[\tau_k^2] = 3\pi/8 \cdot \mathbb{E}[\tau_k]^2$. On a fibre product information sheet, the total DGD is specified by the PMD parameter D_p in the unit of ps/km$^{1/2}$ through the relation [55, Eq. (3.4.4)]

$$\mathbb{E}[\tau_{\text{tot}}^2] = D_p^2(\exp(-z/\ell_c) + z/\ell_c - 1). \tag{2.52}$$

For $z \gg \ell_c$, $\mathbb{E}[\tau_{\text{tot}}^2] = D_p^2 z$. Under the assumption that τ_k has the same distribution, i.e., $\mathbb{E}[\tau_{\text{tot}}^2] = N_{\text{sec}}\mathbb{E}[\tau_k^2]$, we obtain [64, Eq. (C7) is incorrect]

$$\mathbb{E}[\tau_k] = \sqrt{8\mathbb{E}[\tau_{\text{tot}}^2]/(3\pi N_{\text{sec}})} = \sqrt{8/(3\pi)}\sqrt{\ell_c} D_p. \tag{2.53}$$

The typical value of D_p ranges from 0.01 to 10 ps/km$^{1/2}$.

In fibre links where the non-linear length L_{NL} is much larger than the beat length L_{B}, the non-linear operator in (2.47) can be averaged on the Poincaré sphere and changes to the Manakov-PMD equation [66]

$$\frac{\partial \mathbf{A}}{\partial z} = (-\frac{\alpha}{2} - \boldsymbol{\beta_1}\frac{\partial}{\partial t} - j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3})\mathbf{A} + j\frac{8}{9}\gamma|\mathbf{A}|^2\mathbf{A}, \tag{2.54}$$

The main difference between CNLSE and Manakov-PMD is that the non-linear term in Manakov-PMD equation is independent of the launch polarisation.

2.2.5 Split-step Fourier method for numerical simulations

Depending on the purpose of the research, the numerical simulations involved could be any of the above three equations: NLSE, CNLSE or Manakov-PMD equation. They can be all solved by the SSFM [67, Chap. 2.4]. The SSFM is used throughout the thesis in simulations of optical fibre channel and, thus, will be introduced in details in this section. The essence of SSFM is to break the optical fibre channel into very short pieces so that in each of the section NLSE can be solved by calculating the linear and non-linear parts sequentially. In this section, the SSFM for the most complicated channel model CNLSE with PMD will be described so that for any other case one can simply remove irrelevant effects. Typically, the CNLSE is solved in circular polarisations because the non-linear operator $\hat{\mathbf{N}}_c$ in (2.47) is diagonal in this basis [6]. The input field becomes

$$\mathbf{A}_c = \begin{bmatrix} A_+ \\ A_- \end{bmatrix} = \mathbf{R}_c^H \mathbf{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix} \mathbf{A}. \tag{2.55}$$

The linear operator per section becomes

$$\hat{\mathbf{D}}_c^{(k)} = \mathbf{R}_c^H \mathbf{R}^{(k)} \left[\left(-\frac{\alpha}{2} - j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} + \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3} \right) \mathbf{I} - \frac{\zeta_k}{2}\frac{\partial}{\partial t}\sigma_1 \right] \mathbf{R}^{(k)H} \mathbf{R}_c, \quad \text{for } z_{k-1} \le z \le z_k. \tag{2.56}$$

The non-linear operator per section reduces to

$$\begin{aligned} \hat{\mathbf{N}}_c(z) =& j\gamma \mathbf{R}_c^H \left[\mathbf{A}_c^H \mathbf{R}_c^H \mathbf{R}_c \mathbf{A}_c \mathbf{I} - \frac{1}{3}(\mathbf{A}_c^H \mathbf{R}_c^H \sigma_3 \mathbf{R}_c \mathbf{A}_c)\sigma_3 \right] \mathbf{R}_c \\ =& j\frac{2\gamma}{3} \begin{bmatrix} |A_+(z)|^2 + 2|A_-(z)|^2 & 0 \\ 0 & 2|A_+(z)|^2 + |A_-(z)|^2 \end{bmatrix}. \end{aligned} \tag{2.57}$$

In each section, we have the exact solution of (2.47) [67, Eq. (2.4.6)]

$$\begin{aligned} \mathbf{A}_c(z_{k+1}, t) =& \exp\left\{ \Delta z \left(\hat{\mathbf{D}}_c^{(k)} + \hat{\mathbf{N}}_c(z_k) \right) \right\} \mathbf{A}_c(z_k, t) \\ \approx& \exp\left\{ \Delta z \hat{\mathbf{D}}_c^{(k)} \right\} \exp\left\{ \Delta z \hat{\mathbf{N}}_c(z_k) \right\} \mathbf{A}_c(z_k, t). \end{aligned} \tag{2.58}$$

The approximation is based on the Campbell Baker Hausdorff formula [68, Chap. 1] for non-commuting operators. It is accurate to the second order in the step size Δz. Upon implementation, the linear operator is often performed in the frequency domain

$$\begin{aligned} \tilde{\mathbf{A}}_c(z_{k+1}, \omega) \approx& F\{\exp\left\{ \Delta z \hat{\mathbf{D}}_c^{(k)} \right\}\} \tilde{\mathbf{A}}_{cn}(z_k, \omega) \\ =& \exp\left\{ \Delta z F\{\hat{\mathbf{D}}_c^{(k)}\} \right\} \tilde{\mathbf{A}}_{cn}(z_k, \omega), \end{aligned} \tag{2.59}$$

where $\tilde{\mathbf{A}}_{cn}(z_k, \omega)$ is the signal in frequency domain after the non-linear step and F stands for the Fourier transform, and [6, Eq. (9)]

$$\begin{aligned} \exp\left\{ \Delta z F\{\hat{\mathbf{D}}_c^{(k)}\} \right\} =& \exp\left\{ (-\frac{\alpha}{2} + j\frac{\beta_2}{2}\omega^2 + \frac{\beta_3}{6}\omega^3)\Delta z \right\} \times \\ & \begin{bmatrix} \cos(\frac{\omega\zeta_k}{4}) + j\sin 2\phi_k \sin(\frac{\omega\zeta_k}{4}) & j\cos 2\phi_k e^{-j2\theta_k} \sin(\frac{\omega\zeta_k}{4}) \\ j\cos 2\phi_k e^{j2\theta_k} \sin(\frac{\omega\zeta_k}{4}) & \cos(\frac{\omega\zeta_k}{4}) - j\sin 2\phi_k \sin(\frac{\omega\zeta_k}{4}) \end{bmatrix}. \end{aligned} \tag{2.60}$$

In practice, the rotation matrices could be implemented more easily in time domain. For each section, to simulate the PMD effect, we randomly generate the angles ϕ and θ from $[0, \pi]$ uniformly, and the DGD per section τ_k from a Maxwellian distribution with mean $\sqrt{8/(3\pi)} D_p \sqrt{\ell_c}$, then for each sample τ_k in section k we

have $\zeta_k = \tau_k/\ell_c$. Schematically, the SSFM is illustrated in Fig. 2.8

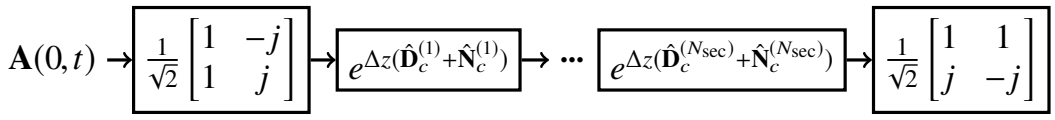

Figure 2.8: Split-step Fourier method for the coupled non-linear Schrödinger equation.

To improve the accuracy of the approximation in (2.58), we can use the following procedure [67, Eq. (2.4.9)]

$$\mathbf{A}_c(z_{k+1},t) \approx \exp\left\{\frac{\Delta z}{2}\hat{\mathbf{D}}_c^{(k)}\right\}\exp\left\{\frac{\Delta z}{2}[\hat{\mathbf{N}}_c(z_k)+\hat{\mathbf{N}}_c(z_{k+1})]\right\}\exp\left\{\frac{\Delta z}{2}\hat{\mathbf{D}}_c^{(k)}\right\}\mathbf{A}_c(z_k,t). \tag{2.61}$$

It includes the non-linear effect in the middle of the segment and employs the trapezoidal rule to approximate the integrated non-linear effect, resulting in an accuracy to the third order of the step size Δz. All simulations throughout the thesis use (2.61) for SSFM.

The correct choice of step size is of vital importance and ensures that the simulation does not generate numerical artifacts. Simulations of optical fibre links without PMD in this thesis follow the logarithmic step-size distribution proposed in [69]. The number of steps per fibre span is calculated via

$$N_{\text{sec}} = \frac{3}{4}L_e^2\gamma^2P_{\text{tot}}^2 10^{R/10} \tag{2.62}$$

where P_{tot} is the total signal power, $L_e = (1-e^{-2\alpha L_{\text{sp}}})/2\alpha$ is the effective span length, and R is the spurious tone suppression ratio that normally takes the value 10. The logarithmic step sizes per span are then generated via

$$\Delta z_n = -\frac{1}{2\alpha}\ln\left[\frac{1-n(1-e^{-2\alpha L_{\text{sp}}})/N_{\text{sec}}}{1-(n-1)(1-e^{-2\alpha L_{\text{sp}}})/N_{\text{sec}}}\right], \quad n = 1,2,\ldots,N_{\text{sec}}. \tag{2.63}$$

Simulations with PMD adopt a fixed step size, i.e., the correlation length ℓ_c.

2.3 Digital coherent receiver

The details of digital coherent receiver are relevant here because they are used in all simulations and experiment throughout the thesis. Since its resurrection in the 2000s [70, 71], the digital coherent receiver has been dominant in the optical communication system in long-haul, sub-sea and even metro networks. The relative high cost of a coherent receiver is the only reason that it has not replaced direct detection in shorter ranges, such as data centre network and passive optical networks. For comparison, the schematics of a direct detector and a coherent receiver

are shown in Fig. 2.9. A direct detector has only one optical band-pass filter, one

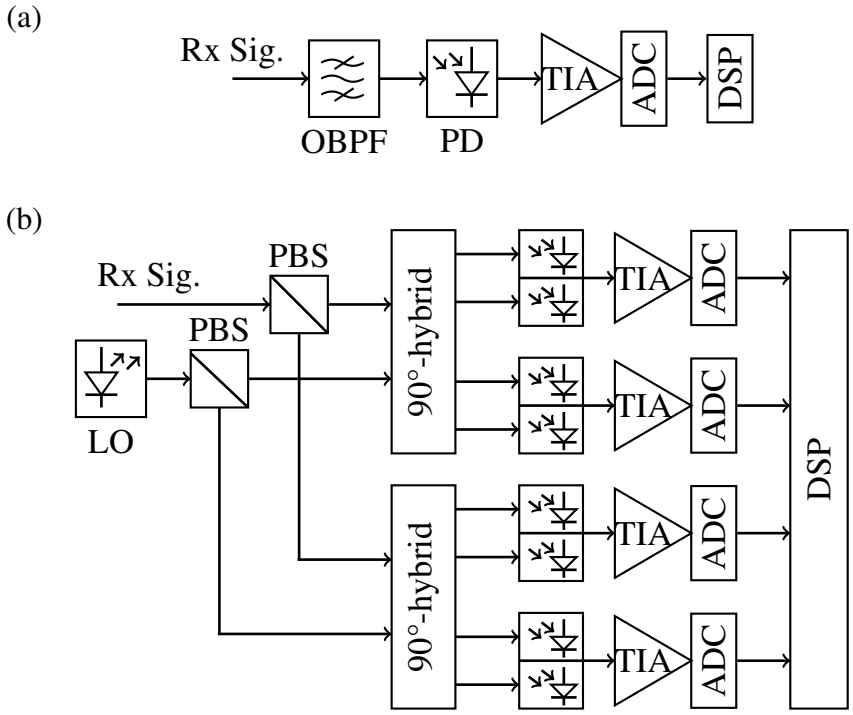

Figure 2.9: The schematics of (a) direct and (b) coherent receiver, adapted from [7].

photo-diode, an ADC and an optional DSP unit. In contrast, a polarisation-diversity digital coherent receiver is roughly four times more complicated than a direct detector. The optical front-end includes two PBSs, one local oscillator (normally an ECSL), and two 90° hybrids. Every two outputs of a 90° hybrid will be received by two photo-diodes and combined into the in-phase or quadrature component of the distorted signal in each polarisation in the base-band. The analogue signals are then amplified by trans-impedance amplifiers, and subsequently digitized by ADC for the later DSP unit. A digital coherent receiver linearly maps the incoming optical signals into four electrical signals in the base-band. The four signals represent the in-phase and quadrature signals in both orthogonal polarisations, respectively. In the following subsections, the working principle of a 90° hybrid and the typical sequence of DSP chain in a coherent receiver are described.

2.3.1 90°-hybrid

A 90°-hybrid functions to mix the incoming signal with the LO as well as the 90°-shifted LO. As shown in Fig. 2.10, a 90°-hybrid has four 3 dB couplers and one 90° phase shift. The two 3 dB couplers right after the input ports E_s and $E_{\ell o}$ function as beam splitters, providing two copies of the input signal in half of its power. The rest two 3 dB couplers mix the two inputs and produce the sum and difference of

them. The transfer function of a 90°-hybrid is

$$\begin{bmatrix} E_1(t) \\ E_2(t) \\ E_3(t) \\ E_4(t) \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -j \\ 1 & j \end{bmatrix} \begin{bmatrix} E_{\mathrm{s}}(t) \\ E_{\ell o}(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) - E_{\ell o}(t)) \\ \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) + E_{\ell o}(t)) \\ \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) - jE_{\ell o}(t)) \\ \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) + jE_{\ell o}(t)) \end{bmatrix}, \tag{2.64}$$

where $E_{\mathrm{s}}(t)$ and $E_{\ell o}(t)$ are the received optical signal and the optical field of the local oscillator.

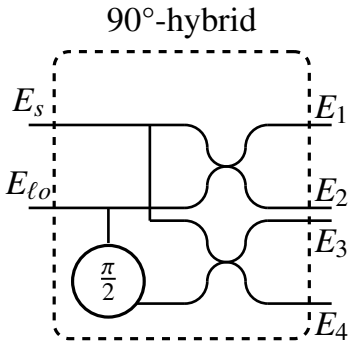

Figure 2.10: The schematic of a 90°-hybrid.

2.3.2 Balanced photo-diodes

The balanced photo-diodes are used in the coherent receiver to remove the direct current component and maximise the mixing between the received signal and the LO. One pair of balanced photo-diodes is responsible to output either the in-phase or quadrature signal on one polarisation. For example, in the homodyne detection scheme, the outputs of the upper pair of 90°-hybrid are

$$\begin{bmatrix} E_1(t) \\ E_2(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) - E_{\ell o}(t)) \\ \frac{1}{\sqrt{2}}(E_{\mathrm{s}}(t) + E_{\ell o}(t)) \end{bmatrix}. \tag{2.65}$$

Absorbing these input lights, the photo-diodes generate the following photo-currents

$$I_1(t) = \frac{R}{2}[P_{\mathrm{s}}(t) + P_{\ell o} + 2\sqrt{P_{\mathrm{s}}(t)P_{\ell o}}\cos\{\phi_{\mathrm{sig}}(t) - \phi_{\ell o}(t)\}], \tag{2.66}$$

$$I_2(t) = \frac{R}{2}[P_{\mathrm{s}}(t) + P_{\ell o} - 2\sqrt{P_{\mathrm{s}}(t)P_{\ell o}}\cos\{\phi_{\mathrm{sig}}(t) - \phi_{\ell o}(t)\}], \tag{2.67}$$

where $P_{\mathrm{s}}(t)$ and $P_{\ell o}(t)$ are the signal and LO powers, respectively, $\phi_{\mathrm{sig}}(t)$ and $\phi_{\ell o}(t)$ are the phase of the received and LO signals. Furthermore, R is the responsivity of

the photo-diode given by

$$R = \frac{q\eta}{\hbar\omega_s}, \tag{2.68}$$

where q is the electron charge, η the quantum efficiency of the photodiode, $\hbar$ the reduced Planck's constant and ω_s the central angular frequency. Combining $I_1(t)$ and $I_2(t)$, we obtain the output of the balanced photodiodes

$$\begin{aligned} I(t) &= I_1(t) - I_2(t) \\ &= 2R\sqrt{P_s(t)P_{\ell o}(t)}\cos\{\phi_{\text{sig}}(t) - \phi_{\ell o}(t)\}. \end{aligned} \tag{2.69}$$

2.3.3 Receiver impairments

Digital coherent receiver introduces distortions into the received digital signal that are similar to the transmitter impairments discussed in Sec. 2.1.3. These distortions can be partially compensated by the receiver DSP as well. The following discussion on the receiver distortion will form the basis of part of the DSP chain in Fig. 6.12.

Transmitter-alike impairments Similar to transmitter, the receiver also has laser noise, ADC quantization noise, I/Q imbalance and delay skew from the imperfection in the optical 90°-hybrid, balanced photodiodes, or transimpedance amplifiers (TIAs). These effects can be compensated altogether at the receiver DSP unit.

Clock mismatch The clock mismatch is due to the two asynchronous clocks at transmitter and receiver sides. The clocks run at the same frequency but may sample at different time. The sampling mismatch introduces a non-trivial additional distortion that needs to be accounted for.

2.3.4 Typical sequence of DSP

The DSP unit in a coherent receiver is the most crucial and powerful part of the receiver. As stated in Sec. 1.1.3, the resurrection of the digital coherent receiver has not only improved the data rate on current optical fibre links, but also changed the link design, moving all the impairment compensation (e.g., the dispersion-managed link) into the DSP unit. In a dispersion-unmanaged link, the received signal is contaminated by the aforementioned transceiver and channel impairments. A typical sequence of DSP operations is depicted in Fig. 2.11.

IQ imbalances compensation The phase and gain mismatch between I- and Q-ports from the transceiver, as well as the timing mismatch caused by the difference in the physical path length of the circuitry, are compensated in this stage. The most well-known compensation technique, also the one used later in Sec. 6.7, is the Gram-Schmidt orthogonalization procedure (GSOP) [72, Sec. II]. Denote the received I/Q signals as $r_I(t)$ and $r_Q(t)$, and the compensated I/Q signals as $I(t)$ and

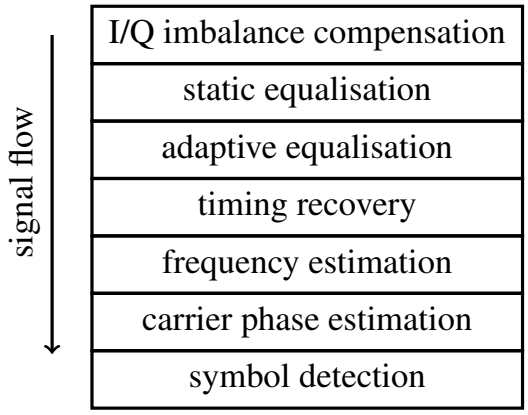

Figure 2.11: The typical DSP sequence, adapted from [8].

$Q(t)$, we have the relation

$$I(t) = \frac{r_I(t)}{\sqrt{P_I}}, \quad Q(t) = \frac{r_Q(t) - \rho r_I(t)/P_I}{\sqrt{P_Q}}, \tag{2.70}$$

where $P_I = \mathbb{E}[r_I^2(t)]$, $P_Q = \mathbb{E}[r_Q^2(t)]$, $\rho = \mathbb{E}[r_I(t) \cdot r_Q(t)]$, and $\mathbb{E}[\cdot]$ denotes expectation over time.

Static equalisation The static equaliser is responsible for the compensation of CD and fibre non-linearity in the optical fibre channel. Static filters are normally time-invariant and have a large number of filter taps.

For the CD compensation, the most commonly-used CD equaliser is the frequency-domain equaliser (FDE), which is simply (2.17). The complexity of such a FDE is $O(\log N)$ [7, Sec. IV.A], where N is the number of taps. In contrast, the complexity of a CD time-domain equaliser (TDE) is $O(N)$.

For the non-linearity compensation, one of the most investigated algorithms is the DBP. It transmits the received digital signal back through a virtual fibre to reverse all the deterministic effects during propagation. Ignoring the PMD effect, we can obtain the virtual propagation by reversing (2.54)

$$\frac{\partial \mathbf{A}}{\partial z} = (\frac{\alpha}{2} + j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2} - \frac{\beta_3}{6}\frac{\partial^3}{\partial t^3})\mathbf{A} - j\frac{8}{9}\gamma|\mathbf{A}|^2\mathbf{A}. \tag{2.71}$$

To investigate its maximum potential gain in transmission systems, DBP is often performed in simulation using the SSFM with fine step size. In a fully-loaded WDM system, a single-channel DBP has the potential gain in peak-SNR of 0.5 dB, while the joint three- or five-channel DBP has roughly 1 and 1.2 dB, respectively [73]. However, when it comes to its implementation in ASICs, the high complexity of SSFM-DBP has to be reduced with various methods [74, 75], often compromising

its effectiveness.

Apart from the DBP algorithm of high complexity, several other approaches are proposed to tackle the fibre non-linearity, such as the Volterra equaliser [76], NFT-based transmission scheme [77–79], machine learning methods [80], maximum *a posteriori* (MAP) [81] etc..

Adaptive equalisation Adaptive equalisers target the stochastic and time-varying impairments caused by the transceiver and fibre birefringence. They often take the form of a two by two multiple-input multiple-output (MIMO) finite impulse response (FIR) filter as shown in Fig. 2.12. We consider first input sequence of

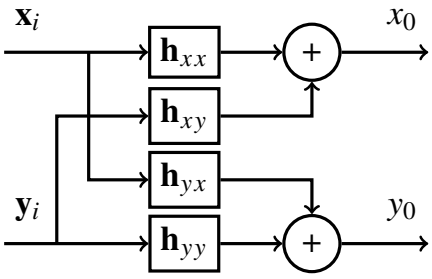

Figure 2.12: 2×2 MIMO structured adaptive FIR filter.

one-sample-per-symbol. Let the column vectors $\mathbf{h}_{xx}$, $\mathbf{h}_{xy}$, $\mathbf{h}_{yx}$ and $\mathbf{h}_{yy}$ of length N denote the N tap weights of the FIR filter. $\mathbf{x}_i$ and $\mathbf{y}_i$ are sliding blocks of N input samples fed into the filter in the form of a row vector

$$\mathbf{x}_i(n) = [x_i(n)\ x_i(n-1)\ \ldots\ x_i(n-N)],$$
$$\mathbf{y}_i(n) = [y_i(n)\ y_i(n-1)\ \ldots\ y_i(n-N)].$$

We denote $[\mathbf{h}_{xx}^H\ \mathbf{h}_{xy}^H]$ as $\mathbf{h}_x^H$, $[\mathbf{h}_{yx}^H\ \mathbf{h}_{yy}^H]$ as $\mathbf{h}_y^H$, and $[\mathbf{x}_i(n)\ \mathbf{y}_i(n)]^T$ as $\mathbf{u}_i$, the output of the adaptive filter is then

$$x_0(n) = \mathbf{h}_x^H \mathbf{u}_i(n),$$
$$y_0(n) = \mathbf{h}_y^H \mathbf{u}_i(n). \tag{2.72}$$

There are in general three types of algorithms to calculate the tap weights of an adaptive filter, i.e., the constant-modulus algorithm (and its variations), the decision-directed least-mean-square algorithm, and the recursive least-squares (RLS) algorithm. They all have a cost/error function and use stochastic gradient descent to update the tap weights until they converge. The principle of the RLS algorithm is described here because it is used later in Sec. 6.7.1 due to its faster convergence [82].

The cost function of a standard RLS algorithm is

$$J_{x/y}(n) = \sum_{i-1}^{n} \lambda^{n-i} |\epsilon_{x/y}(i)|^2$$
$$= \sum_{i-1}^{n} \lambda^{n-i} \left| R_2 - |\mathbf{h}_x^H(n)\mathbf{u}_i(i)|^2 \right|^2, \tag{2.73}$$

where $\lambda < 1$ is the forgetting factor, usually set to 0.999, and R_2 is the square radius of the transmitted symbols x_{sym}, calculated by $R_2 = \mathbb{E}[|x_{\text{sym}}|^4]/\mathbb{E}[|x_{\text{sym}}|^2]$. Because the cost function is not quadratic in array weight vector $\mathbf{h}_{x/y}$, an intermediate data vector $\mathbf{z}_{x/y}$ has to be introduced [83]. The detailed updating process is written out in Alg. 1 at the end of the chapter.

Timing recovery The timing error, or clock mismatch, results from the asynchronous clocks at the transmitter and receiver. The most widely-used method is the Gardner's method [84]. One damaging drawback of the Gardner's algorithm is that it fails for the Nyquist pulse shaped signal, which is commonly used in the high spectral efficiency communication systems. A modification was, therefore, proposed [85]. Let $\mathbf{x_i} = [x_i(-1)\ x_i(0)\ \ldots\ x_i(N-1)]$ be the tow-fold oversampled complex input signal of length N, the Gardner's algorithm estimates the timing phase error using

$$\tau_{\text{err}} = \text{Re}\left\{ \sum_{n=0}^{N/2-1} [x_i(2n-1) - x_i(2n+1)] x_i^*(2n) \right\}. \tag{2.74}$$

The modified Gardner's estimator reads

$$\tau_{\text{err}} = \text{Re}\left\{ \sum_{n=0}^{N/2-1} [P_i(2n-1) - P_i(2n+1)] P_i(2n) \right\}, \text{ where } P_i(n) = |x_i(n)|^2. \tag{2.75}$$

In addition, a fourth-power frequency-domain algorithm was proposed in [86]

$$\tau_{\text{err}} = \sum_{k=-N'/2}^{N'/2} \text{Im}\left\{ \left(\sum_{\ell=0}^{N'} \tilde{X}_i(\ell) \tilde{X}_i^*(\ell + k + N_{\text{DFT}} - N') \right) \times \right.$$
$$\left. \left(\sum_{\ell=0}^{N'} \tilde{X}_i^*(\ell + N_{\text{DFT}} - N') \tilde{X}_i^*(\ell + k) \right)^* \right\}, \tag{2.76}$$

where $\tilde{X}_i(k)$ is the kth element of the N_{DFT}-period discrete Fourier transform (DFT) of $\mathbf{x}_i$. $N' < N$ is a power-of-two integer that can be used to adjust the complexity of the algorithm. The 4th power frequency domain method shows the best performance over others and is insensitive to the roll-off factor of the RRC pulse

shape [86].

Frequency offset estimation An intradyne coherent system means that the local oscillators at transceiver are not frequency and phase locked. The resulted frequency offset of the received signal is tracked in the DSP unit. To formulate the problem mathematically, the received symbol $y(k)$ at time instance k can be written as

$$y(k) = x(k)\exp\big(j[\theta(k) + k2\pi\Delta f T_{\text{sym}}]\big) + n(k), \tag{2.77}$$

where $x(k)$ is the transmitted symbol at time instance k, θ_k and $n(k)$ are the phase and additive noise, and T_{sym} is the symbol duration. The frequency offset Δf is to be estimated. There are in general two classes of frequency offset compensation (FOC) algorithms, i.e., blind estimation methods that explore certain properties of the transmitted signal, or the training-aided estimations that are based on the periodically inserted training symbols. A frequency domain blind estimation method is described here because it is later used in Sec. 6.7.1.

The useful signal property we can explore for FOC purpose is that any quadrature amplitude modulation (QAM) constellation satisfies the moment conditions $\mathbb{E}[x^2(n)] = \mathbb{E}[x^3(n)] = 0, \mathbb{E}[x^4(n)] \neq 0$ because of its $\pi/2$-rotationally invariant symmetry property [87]. Therefore, the spectrum of $\mathbf{x}^4$ exhibits a spike at the frequency $4\Delta f$. The frequency offset can be written as

$$\Delta f = \frac{1}{4}\arg\max_f \left|\text{DFT}\{\mathbf{x}^4\}\right|^2, \; -\frac{R_s}{2} < f < \frac{R_s}{2}, \tag{2.78}$$

where R_s stands for the sampling rate.

Carrier phase estimation As explained earlier in Sec. 2.1.3, the carrier phase noise is a result of the constant fluctuation of the central frequency of the laser. To model the problem, it is often assumed that the central frequency of the laser does not change but the phase of the laser fluctuates. The task of tracking the carrier phase on the received signal can be formulated as estimating the $\theta(k)$ in $y(k) = x(k)\exp(j\theta(k)) + n(k)$. The notation is the same as in (2.77).

Here, we described a simple pilot-aided maximum likelihood carrier phase estimation method [88] that is also used in Sec. 6.7.1. Multiple pilot symbols are inserted periodically in the signal in a frequency slightly larger than the line-width of the laser. Since the carrier phase noise is a slowly-varying process, it can be assumed to be constant within one block of pilots and payloads. Let $\mathbf{p} = [p(1)\ p(2)\ \ldots\ p(L)]$ be the transmitted training symbol vector and $\mathbf{v} = [v(1)\ v(2)\ \ldots\ v(L)]$ be the corresponding received symbol vector. The carrier phase noise within the current block can be estimated by the maximum likelihood

estimator

$$\theta(k) = \arg(\mathbf{v}^H \mathbf{p}). \tag{2.79}$$

The detailed derivation of the maximum likelihood estimator can be found as early as in 1986 [89]. The maximum likelihood estimator is proven to work better with BPSK signals than any other high-order QAM signals. An improvement has been proposed in [90] with the cost of higher complexity.

2.4 Information-theoretical tools

So far, we have reviewed the physics of many components in an optical communication system, the impairments they have on the signal, and methods to compensate them. To evaluate the overall system performance in terms of achievable information rate (AIR), the framework for information-theoretic analysis needs to be introduced and they are heavily used in the system analysis in Chap. 5 and 6.

The information in a digital communication system experiences a change of information carriers, i.e., from bits to symbols, symbols to waveform, and backwards to bits. AIR for a symbol-wise decoder is commonly referred to as the mutual information (MI), while the generalised mutual information (GMI) refers to the AIR for a bit-wise decoder. These transition of information carriers and the difference between MI and GMI are demonstrated in Fig. 2.13.

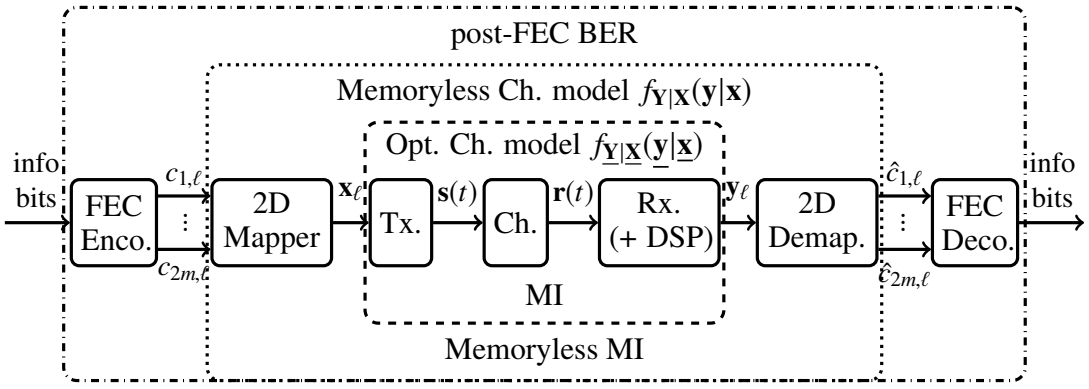

Figure 2.13: Optical fibre system block diagram for information-theoretic analysis [9].

The notations (adopted from [9]) used in the figure are explained as follows. As a general rule, random variables are denoted by capital letters such as X, random vectors by boldface letters such as $\mathbf{X}$, random vectors of length k by boldface letters with superscript such as $\mathbf{X}^k$, and random matrices by underline boldface letters such as $\underline{\mathbf{X}}$. Their realizations are respectively denoted by x, $\mathbf{x}$, $\mathbf{x}^k$ and $\underline{\mathbf{x}}$. A conditional probability density function (PDF) (defined in Appx. A) is written in the form of

$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$. Let $\underline{\mathbf{c}}$ be the realization of the encoded bits for L sequential symbols

$$\underline{\mathbf{c}} = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,L} \\ \vdots & \vdots & \ddots & \vdots \\ c_{2m,1} & c_{2m,2} & \dots & c_{2m,L} \end{bmatrix}.$$

In an optical fibre communication system, each of the codeword $\mathbf{c}_\ell = [c_{1,\ell}\ c_{2,\ell}\ \dots\ c_{2m,\ell}]^T$ corresponds to two complex symbols on the two polarisations of light denoted by $\mathbf{x}_\ell = [x_{1,\ell}\ x_{2,\ell}]^T$ with $\ell = 1, 2, ..., L$. The symbols $x_{1,\ell}$ and $x_{2,\ell}$ are drawn uniformly from a discrete constellation $\mathcal{X}$ with cardinality $M = 2^m = |\mathcal{X}|$. Common choices of constellation include 16QAM. The symbols of an entire block are denoted as $\underline{\mathbf{x}} = [\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_L]$. The same notation rule applies for the output symbol $\mathbf{y}_\ell \in \mathbb{C}^2$, which is then de-mapped into $\hat{\mathbf{c}}_\ell = [\hat{c}_{1,\ell}\ \hat{c}_{2,\ell}\ \dots\ \hat{c}_{2m,\ell}]^T$.

Modern communication systems usually use orthonormal bases $\{g(t - nT)\}_{n\in\mathbb{N}}$, for instance, the set of sinc functions (the Nyquist pulse shape), to generate band-limited continuous waveform signal $\mathbf{s}(t) = [s_1(t)\ s_2(t)]$ from the input symbols $\underline{\mathbf{x}}$. As an example, $s_1(t)$ of duration D can be generated by sinc functions with bandwidth W

$$s_1(t) = \sum_{\ell=0}^{L-1} x_{1,\ell}\ \mathrm{sinc}(2Wt - \ell), \tag{2.80}$$

where $L = 2DW$. At the same time, $x_{1,\ell}$ can be recovered from $s_1(t)$ via the inner product (also known as the matched filter)

$$x_{1,\ell} = \int_{-\infty}^{\infty} s_1(t)\ \mathrm{sinc}(2Wt - \ell)dt, \tag{2.81}$$

or sampling with a Dirac function at the rate of 2W.

At the receiver, the continuous waveform signal $\mathbf{r}(t)$ is discretized into intermediate samples $\underline{\mathbf{u}_\ell}$ in the form of

$$\underline{\mathbf{u}_\ell} = \begin{bmatrix} u_{1,\gamma_o(\ell-1)+1} & \dots & u_{1,\gamma_o\ell} \\ u_{2,\gamma_o(\ell-1)+1} & \dots & u_{2,\gamma_o\ell} \end{bmatrix}, \tag{2.82}$$

where γ_o is the oversampling rate. The discretisation process is done either via sampling with a Dirac function at the rate of $2\gamma_o$W, or equivalently the inner product

$$u_{1,k} = \int r_1(t)\ \mathrm{sinc}(2\gamma_o Wt - k)dt,\ k = 1, 2, ..., \gamma_o L. \tag{2.83}$$

Then $\underline{\mathbf{u}}_\ell$ enters the DSP unit to calculate the final symbol value $\mathbf{y}_\ell$. The need of oversampling is due to the signal spectral broadening induced either by phase noise [91] or the fibre non-linearity [92]. The spectral broadening is one of the fundamental problems in the information theory of optical fibre channel as during transmission, the original information is spread over more dimensions than the initial signal dimension of 2DW.

In the context of optical fibre communications, systems often operate at a signal power where the spectral broadening is not significant. Since the power-rate curve (e.g., Fig. 1.7) often shows a peak around certain signal power, the engineering community is not much interested in how the channel behaves in the highly nonlinear regime. However, the exact capacity of optical fibre channel over any range of input powers remains an actively pursued problem for information theorists. In this thesis, the spectral broadening effect is also largely ignored for simulations and experiment except for the work described in Chap. 3, which aims to characterise the effect over a large range of powers precisely and efficiently.

For all the optical fibre systems studied in simulations or experiments here, we are interested in the MI and the capacity of the channel $f_{\mathbf{Y}|\mathbf{X}}(\underline{\mathbf{y}}|\underline{\mathbf{x}})$. The random matrices $\underline{\mathbf{Y}}$ and $\underline{\mathbf{X}}$ can be also written in vector form as $\mathbf{Y}^{2L} = [\mathbf{Y}_1^L\ \mathbf{Y}_2^L]$ and $\mathbf{X}^{2L} = [\mathbf{X}_1^L\ \mathbf{X}_2^L]$. Doing so has no impact on the calculation of MI. Note that, as shown in Fig. 2.13, the channel model includes also the DSP unit. Therefore, the MI depends on the chosen DSP algorithm, and maximising the MI of a such channel model over various input distributions cannot produce the capacity of the optical fibre channel, but only a lower bound. In general, the optical fibre channel has multi-dimensional input and output, with inter-symbol interference among the consecutive time-domain symbols caused by channel memory. This makes it extremely difficult to rigorously derive its channel capacity. The concept of MI and its relation with channel capacity are explained in the following section.

2.4.1 Mutual information and channel capacity

Mutual information was developed by Claude Shannon in his seminal work [38] on the capacity of communication channels. The most-studied channel is the discrete memoryless channel. This is a channel for which the input and output are each sequences of letters from finite alphabets and for which the output letter at a given time depends statistically only on the corresponding input letter [93, Sec. 4.1]. For the single-dimensional discrete time memoryless channel, MI is defined as

$$I = I(X;Y) \triangleq \mathbb{E}_{X,Y}\left[\log_2 \frac{f_{Y,X}(y,x)}{f_Y(y) f_X(x)}\right], \tag{2.84}$$

where $\mathbb{E}_{X,Y}$ represents expectation over both X and Y. For the memoryless channel denoted by $f_{Y|X}(y|x)$ with the input variable of distribution $f_X(x)$, the largest data transmission rate achieved by various channel codes is $I(X;Y)$. The channel capacity of such a channel is then defined as

$$C \triangleq \sup_{f_X(x)} I(X;Y), \tag{2.85}$$

where sup means the supremum of $I(X;Y)$ over all the possible distributions of the input variable X.

MI can also be defined via entropy. Let $h(X)$ denote the entropy of the continuous random variable X with probability density function $f_X(x)$, then $h(X)$ is defined as

$$h(X) \triangleq -\mathbb{E}_X\Big[\log_2 \ f_X(x)\Big]. \tag{2.86}$$

Similarly, the conditional entropy $h(Y|X)$ is defined as

$$h(Y|X) \triangleq -\mathbb{E}_{Y,X}\Big[\log_2 f_{Y|X}(y|x)\Big]. \tag{2.87}$$

Finally, MI can be written in terms of entropy as

$$I(X,Y) = h(X) - h(X|Y) = h(Y) - h(Y|X). \tag{2.88}$$

Note that for discrete random variables with probability mass functions, entropy is commonly denoted by capital letter H and calculated using the same definition, only replacing the integral with a finite sum. As a historical remark, Shannon defined the entropy of a continuous distribution simply by replacing the sum in the entropy of a discrete distribution with an integral [38, Part III, Sec. 20]. The continuous entropy is often criticised for the lack of transformation invariance [94] and its contradiction with variance when measuring uncertainty [95, Example 1]. The lack of transformation invariance causes direct contradiction with the fact the deterministic processing cannot increase Shannon entropy, i.e., $h(f(X)) \leq h(X)$ where $f(\cdot)$ is a deterministic and invertible function. Therefore, it was proposed in [94] to use the Kullback-Leibler divergence [95, Eq. (2.3)] to measure the relative entropy of a continuous distribution with respect to a well-understood distribution such as the uniform distribution. Unfortunately, I only discovered the disadvantage of the Shannon's continuous entropy by the time of writing the thesis. Since we are only interested in the relative relations between the continuous entropy of different distributions, the Shannon definition for entropy will continue to be used.

In the case of a channel with memory [93, Chap. 4], one has to consider multiple consecutive symbols at the input and output, the channel capacity is defined as

$$C_{\text{mem}} \triangleq \lim_{k\to\infty} \sup_{f_{\mathbf{X}^k}(\mathbf{x}^k)} \frac{1}{k} I(\mathbf{X}^k;\mathbf{Y}^k), \tag{2.89}$$

where $\mathbf{X}^k = [X_1\ X_2\ ...\ X_k]$. The complexity of estimating the mutual information of multi-dimensional input and output is so high such that in reality one rarely considers detection scheme of sequence longer than 3. If the signal detection scheme for a channel with memory is based on the block-wise memoryless assumption, the following relation holds

$$\frac{1}{k} I(\mathbf{X}^k;\mathbf{Y}^k) \leq C_{\text{mem}}. \tag{2.90}$$

2.4.2 Numerical estimation of mutual information and capacity

The numerical estimation of MI falls broadly into three situations, 1) the exact channel model $f_{\mathbf{Y}^k|\mathbf{X}^k}(\mathbf{y}^k|\mathbf{x}^k)$ is known, 2) the channel model $f_{\mathbf{Y}^k|\mathbf{X}^k}(\mathbf{y}^k|\mathbf{x}^k)$ needs to be estimated via a histogram with data gathered through exhaustive system simulations, 3) the channel model $f_{\mathbf{Y}^k|\mathbf{X}^k}(\mathbf{y}^k|\mathbf{x}^k)$ is approximated with a known distribution, e.g., the Gaussian distribution. When estimating the MI and capacity of optical fibre channel $f_{\underline{\mathbf{Y}}|\underline{\mathbf{X}}}(\underline{\mathbf{y}}|\underline{\mathbf{x}})$ later in Sec. 5.4 and 6.4, we also needed to make two assumptions to simplify the procedure, which is otherwise often not feasible. Firstly, the symbols on two polarisations of light are assumed independent after DSP. Secondly, the sub-channels on each polarisation are assumed memoryless after DSP. We use the second method in Sec. 5.4 and the third method in Sec. 6.4. In both cases, the optical fibre channel is treated as a single-dimensional discrete memoryless channel denoted as $f_{Y|X}(Y|X)$.

Consider the continuous input random variable X of certain distribution $f_X(x)$. At each discrete time instance k, the random variable generates a random sample $x_k \in \mathbb{C}$, forming a symbol stream $\mathbf{x} = [x_1\ x_2\ ...\ x_L]$. The power of the input symbol is defined as $P = \mathbb{E}[|X|^2] = \|\mathbf{x}\|_2^2/L = (x_1^2 + x_2^2 + ... + x_L^2)/L$. The corresponding discrete memoryless channel outputs symbols are $\mathbf{y} = [y_1\ y_2\ ...\ y_L]$.

MI estimation with a histogram To estimate $f_{Y|X}(y|x)$ for a channel with continuous input and output space, ideally one has to send every point on the complex plane through the channel for a large number of times to gather statistics via histogram. However, as there are infinite number of points on the complex plain, we can only choose finite discrete points for X under the the power constraint $\mathbb{E}[|X|^2] \leq P$. On the receiving side, building a histogram will also require the discretization of Y into fine intervals so that the frequencies of Y falling into each interval can be counted.

For this purpose, a popular choice of the input constellation $\mathcal{X} = [x_{c1}\ x_{c2}\ ...\ x_{cM}]$

is the uniform ring constellation [5] as shown in Fig. 2.14(a). The constellation point x_{ci} can be easily represented by its radius and phase $x_{ci} = re^{j\phi}$. The rationale behind is that a bi-dimensional Gaussian constellation achieves the capacity of a complex-input complex-output AWGN channel and a discrete multi-ring constellation is its easiest approximation. Since the optimal input distribution for the channel under estimation is unknown, it is sensible to start with a input distribution that works well for the AWGN channel.

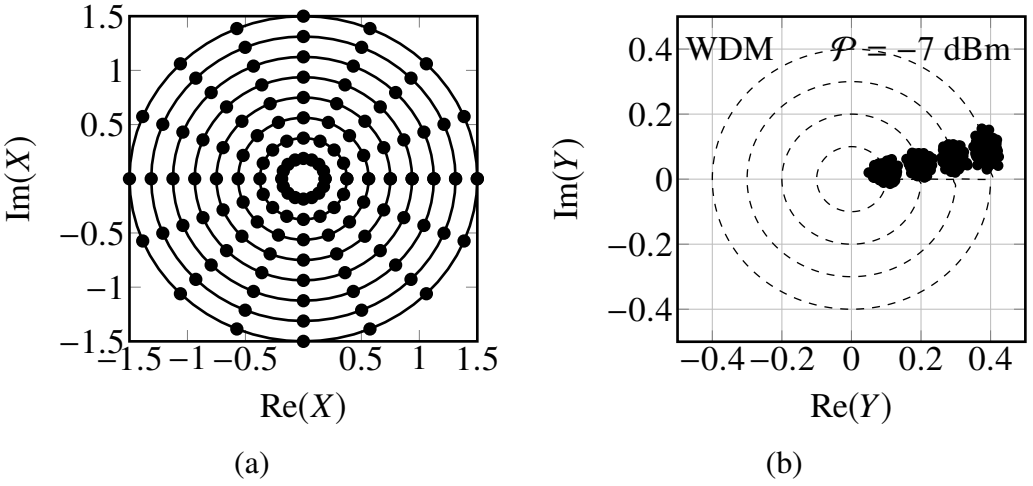

Figure 2.14: (a) A uniformly distributed multi-ring constellation and (b) One example of the received constellation with the transmitted phase removed.

To further simplify the procedure, one can assume that the channel model is rotational-invariant, i.e., $f_{Y|X}(y|x) = f_{Y|X}(ye^{j\phi}|xe^{j\phi})$, $\phi \in \mathbb{R}$. Under this assumption, we can just gather statistics of the transmitted constellation points on one specific phase. In practice, we send every points in the constellation but remove the transmitted phase upon receiving the symbol so that the received noisy symbols cluster around the constellation points on the positive real axis as shown in Fig. 2.14(b). The deviation from the positive real axis in this example is caused by the XPM in (2.21). Due to the rotational invariance of the channel, it is convenient to work in the polar coordinate. The channel model can be rewritten as $f_{R_Y\Phi_Y|R_X\Phi_X}(r_y, \phi_y|r_x, \phi_x)$, where R_X and Φ_X are the radius and phase of X such that $X = R_X e^{j\Phi_X}$ and the same for R_Y and Φ_Y. To simplify the notation without loss of generality, let us ignore the deviation from the positive real axis for a moment and denote the obtained histogram as $g_{R_Y\Phi_Y|R_X}(r_y, \phi_y|r_x)$, then $f_{R_Y\Phi_Y|R_X\Phi_X}(r_y, \phi_y|r_x, \phi_x)$ can be obtained via

$$f_{R_Y\Phi_Y|R_X\Phi_X}(r_y, \phi_y|r_x, \phi_x) = g_{R_Y\Phi_Y|R_X}(r_y, \phi_y - \phi_x|r_x). \tag{2.91}$$

The MI can be calculated via

$$\begin{aligned} I(X;Y) &= I(R_Y, \Phi_Y; R_X, \Phi_X) \\ &= \mathbb{E}_{R_Y,\Phi_Y,R_X,\Phi_X}\left[\log_2 \frac{f_{R_Y\Phi_Y|R_X,\Phi_X}(r_y,\phi_y|r_x,\phi_x)}{\sum_{R_X,\Phi_X} f_{R_Y\Phi_Y|R_X\Phi_X}(r_y,\phi_y|r'_x,\phi'_x) f_{R_X\Phi_X}(r'_x,\phi'_x)}\right]. \end{aligned} \tag{2.92}$$

With the estimated PDF $f_{R_Y\Phi_Y|R_X\Phi_X}(r_y,\phi_y|r_x,\phi_x)$, the channel capacity can also be estimated by optimising MI over different input distributions. The classical algorithm for the optimisation is the Arimoto-Blahut algorithm [96,97]. To simplify the notation, we write out the algorithm using the $f_{Y|X}(y|x)$ and $f_X(x)$ in Alg. 2 at the end of the chapter without giving any proof.

MI estimation with a known distribution Although the above histogram-based method is theoretically sound, it is quite computationally expensive and errors are often introduced via the discretisation of Y. In most of the applications, the detection scheme is based on the assumption that the conditional channel PDF is a Gaussian distribution. Therefore, it is sensible to calculate the MI with the same assumption. The input distribution will be the same as the one used in the corresponding system, e.g., often the uniform distribution or the Maxwell-Boltzmann distribution from probabilistic constellation shaping. To mathematically formulate the problem, we denote the conditional Gaussian distribution as $G_{Y|X}(y|x) \sim \mathcal{N}(x, \sigma^2 \underline{\mathbf{I}_2})$, where $\underline{\mathbf{I}_2}$ is the 2×2 identity matrix and σ^2 is the noise variance per real dimension. The MI can be written as

$$I(X;Y) = \sum_{X,Y} f_X(x) G_{Y|X}(y|x) \log_2 \frac{G_{Y|X}(y|x)}{\sum_X G_{Y|X}(y|x') f_X(x')}. \tag{2.93}$$

2.4.3 Entropy estimators

In addition to the MI, two other important quantities for information-theoretic analysis are the entropy and the conditional entropy defined in (2.86) and (2.87). Similar to the estimation of MI, one can either estimate entropy via simulated data or by assuming that the distribution of data is close to a well-known one. We introduce here a distribution-agnostic entropy estimator, called kpN estimator, developed in [11]. The entropy estimator will be used to calculated the conditional entropy as a measure of correlation among sub-carriers in Sec. 6.5. To understand the kpN estimator, we should first go through the principle of the classical entropy estimator.

The classical entropy estimator was first formulated in 1976 [98]. We adopt here a more modern formulation described in [11]. Recall the continuous entropy

$h(\mathbf{X})$ of a d-dimensional random variable $\mathbf{X} \in \mathbb{R}^d$ defined as

$$h(\mathbf{X}) = -\mathbb{E}_{\mathbf{X}}\left[\log_2 \ f_{\mathbf{X}}(\mathbf{X})\right] \tag{2.94}$$

Let $\mathbf{x}_i,\ i = 1, 2, ..., N$ be the N samples of the random variable X, a simple Monte Carlo estimate of the entropy is

$$\hat{h}(\mathbf{X}) = -\frac{1}{N}\sum_{i=1}^{N} \log f_{\mathbf{X}}(\mathbf{x}_i). \tag{2.95}$$

Since the $f_{\mathbf{X}}(\mathbf{x_i})$ is unknown, the idea is to estimate it through k nearest neighbors (KNN) of $\mathbf{x_i}$. Consider a d-dimensional ball region of radius ϵ, denoted as $\mathcal{B}(\epsilon, \mathbf{x}_i) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_i\| \leq \epsilon\}$, around a sample point $\mathbf{x}_i \in \mathbb{R}^d$. The probability $f_k(\epsilon)d\epsilon$ is the probability that exactly one point is in $\left\{\mathcal{B}(\epsilon + d\epsilon, \mathbf{x}_i) - \mathcal{B}(\epsilon, \mathbf{x}_i)\right\}$, exactly $k-1$ points are in $\mathcal{B}(\epsilon, \mathbf{x}_i)$, and exactly $N-k-1$ points are outside $\mathcal{B}(\epsilon + d\epsilon, \mathbf{x}_i)$

$$f_k(\epsilon)d\epsilon = \underbrace{\binom{N-1}{1}\frac{dF_i(\epsilon)}{d\epsilon}d\epsilon}_{\substack{\text{probability of exactly one point} \\ \text{in } \left\{\mathcal{B}(\epsilon + d\epsilon, \mathbf{x}_i) - \mathcal{B}(\epsilon, \mathbf{x}_i)\right\}.}} \times \underbrace{\binom{N-2}{k-1} F_i(\epsilon)^{k-1}(1 - F_i(\epsilon))^{N-2-(k-1)}}_{\text{probability of exactly } k-1 \text{ points in } \mathcal{B}(\epsilon, \mathbf{x}_i).}, \tag{2.96}$$

where $F_i(\epsilon)$ is the probability mass in $\mathcal{B}(\epsilon, \mathbf{x}_i)$ as

$$F_i(\epsilon) = \int_{\mathcal{B}(\epsilon, \mathbf{x}_i)} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \tag{2.97}$$

One can compute the expectation value of $\log F_i(\epsilon)$ [11, Eq. (5)]

$$\begin{aligned}
\mathbb{E}[\log F_i(\epsilon)] &= \int_0^{+\infty} f_k(\epsilon)d\epsilon \log F_i(\epsilon) \\
&= \binom{N-1}{1}\binom{N-2}{k-1}\int_0^1 F_i(\epsilon)^{k-1}(1 - F_i(\epsilon))^{N-k-1} \log F_i(\epsilon)dF_i(\epsilon) \\
&= \psi(k) - \psi(N)
\end{aligned} \tag{2.98}$$

where $\psi(x)$ is the digamma function that satisfies $\psi(x+1) = \psi(x) + 1/x$ and $\psi(1) = -0.5772156...$ [99, Eq. (4)]. The expectation is taken over the positions of all other $N-1$ points, with $\mathbf{x}_i$ fixed [99, Eq. (17)].

If the probability mass $F_i(\mathbf{x}_i)$ in $\mathcal{B}(\epsilon, \mathbf{x}_i)$ is approximated by a uniform distribu-

tion scaled by an arbitrary real factor η_i

$$F_i(\mathbf{x}_i) \approx \eta_i f_{\mathbf{X}}(\mathbf{x}_i), \tag{2.99}$$

we obtain the classical entropy estimator by substituting (2.99) into (2.95)

$$\hat{h}(\mathbf{X}) = -\frac{1}{N}\sum_{i=1}^{N} \log f_{\mathbf{X}}(\mathbf{x}_i) \tag{2.100}$$

$$= -\frac{1}{N}\sum_{i=1}^{N} \log F_i(\mathbf{x}_i) + \frac{1}{N}\sum_{i=1}^{N} \log \eta_i \tag{2.101}$$

$$= -\mathbb{E}[\log F_i(\epsilon)] + \frac{1}{N}\sum_{i=1}^{N} \log \eta_i \tag{2.102}$$

$$= \psi(N) - \psi(k) + \frac{1}{N}\sum_{i=1}^{N} \log \eta_i. \tag{2.103}$$

The kpN algorithm has the same principle as the above classical entropy estimator, however, makes a different approximation at the step of (2.99). The probability mass $F_i(\mathbf{x}_i)$ in $\mathcal{B}(\epsilon, \mathbf{x}_i)$ can be approximated by a Gaussian distribution with the mean and variance calculated by the samples in $\mathcal{B}(\epsilon, \mathbf{x}_i)$. Formally,

$$F_i(\mathbf{x}_i) \approx \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{g(\mathbf{x}_i)} G_i, \tag{2.104}$$

where

$$g(\mathbf{x}_i) = \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mu)^T \underline{\mathbf{S}}^{-1}(\mathbf{x}_i - \mu)\right) \tag{2.105}$$

$$G_i = \int_{\mathcal{B}(\epsilon, \mathbf{x}_i)} g(\mathbf{x}_i) d\mathbf{x} \tag{2.106}$$

$$\mu = \frac{1}{k}\sum_{\mathbf{x} \in \mathcal{B}(\epsilon, \mathbf{x}_i)} \mathbf{x} \tag{2.107}$$

$$\underline{\mathbf{S}} = \frac{1}{k}\sum_{\mathbf{x} \in \mathcal{B}(\epsilon, \mathbf{x}_i)} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T. \tag{2.108}$$

The one last piece of the kpN algorithm is the evaluation of the multi-dimensional Gaussian integral in (2.106). The seemingly trivial integral is the most difficult part of the kpN algorithm and it itself deserves another 27-page explanation [100]. The method to calculate the integral is called the expectation propagation multivariate Gaussian probability (EPMGP). For the details of the EPMGP, we refer interested

readers to [100] and its Matlab library online. The kpN algorithm is summarized in Al. 3 at the end of the chapter.

2.4.4 Interference channels

While the MI is instrumental for the development of point-to-point communication links, the multi-user information theory is more suitable for the analysis of the WDM-based optically routed network. In particular, the interference channel (IC) model, where multiple users share the same transmission medium and impose interference on each other, fits well to the non-linear effect we observe in the optical fibre channel. Through the study of simpler interference channels, we wish to learn strategies to combat the non-linear inter-channel interference in the optical fibre channel and also understand whether a monotonically increasing capacity in an ORN is possible at all. We start by reviewing the principle of a 2×2 IC [101, Chap. 21] depicted in Fig. 2.15.

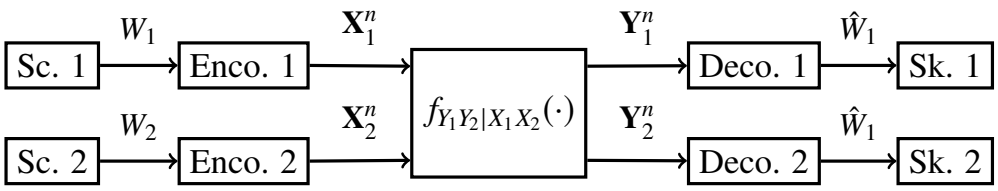

Figure 2.15: A two-transmitter IC. Sc. stands for source, Enco. for encoder, Deco. for decoder and Sk. for sink.

There are two statistically independent sources that produce symbols W_1 and W_2 with nR_1 and nR_2 bit, respectively. Encoder 1 maps w_1 to a sequence $x_1^n \in \mathcal{X}_1^n$, encoder 2 maps w_2 to a sequence $x_2^n \in \mathcal{X}_2^n$. The channel $f_{Y_1Y_2|X_1X_2}(\cdot)$ generates sequence $y_1^n, y_2^n \in \mathcal{Y}_1^n, \mathcal{Y}_2^n$. Decoders estimate $\hat{w}_1$ and $\hat{w}_2$ using y_1^n, y_2^n. We are interested in the rate pairs (R_1, R_2) for which one can make the error probability $P_e = Pr\left[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right]$ arbitrarily small. All achievable rate pairs (R_1, R_2) form the IC capacity region C_{IC}. The capacity region of a general IC is unknown. The channel model can be easily extended to K-users. To have a more concrete understanding of ICs, we introduce two classes of ICs, i.e., the AWGN IC and a type of deterministic ICs. The reasons of choosing these two types of ICs are twofold: 1) AWGN-IC is the simplest 2×2 IC we can analyse and simulate easily, 2) deterministic ICs capture the deterministic aspect of non-linear distortion in the optical fibre channel in an abstract way.

AWGN-IC

We write the real-valued channel model of an AWGN-IC in standard form as

$$\begin{aligned} Y_1 &= X_1 + \sqrt{a_{12}} X_2 + N_1, \\ Y_2 &= \sqrt{a_{21}} X_1 + X_2 + N_2, \end{aligned} \tag{2.109}$$

where $(N_1, N_2) \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$, $\rho \in [-1, 1]$ denotes the correlation coefficient between N_1 and N_1, and $\mathbb{E}[|X_k|^2] \leq P_k,\ k = 1, 2$.

There are three communication strategies for the AWGN-IC: a) treat interference as noise, b) use time-division multiplexing (TDM) or frequency-division multiplexing (FDM), c) decode the interference. We plot in Fig. 2.16 the capacity regions of the aforementioned strategies as shown in [101, Sec. 21.6.3]. Sato's outer bound on AWGN-IC [102] is also plotted for comparison. From Fig. 2.16 it can be observed that 1) Sato's capacity outer bound [102] on AWGN-IC is always smaller than the capacity region of two independent AWGN channels, with the difference increasing with the interference coefficients, 2) at different strength of interference (indicated by coefficients a_{12} and a_{21}), different strategy should be adopted. For example, while treating interference as noise is the best strategy among the three for $a_{12} = a_{21} = 0.1$, it is the worst for $a_{12} = a_{21} = 1.1$. In addition, the exact capacity region of an AWGN-IC is only known for strong interference where $a_{12} > 1$ and $a_{21} > 1$, where the optimal strategy is to decode the interference.

Deterministic ICs

At next, several deterministic ICs will be reviewed. We start from the general form of deterministic ICs, and tweak the channel model to make it closer to the WDM multi-user optical fibre network.

The deterministic IC in its most general form reads

$$\begin{aligned} Y_1 &= f_1(X_1, X_2), \\ Y_2 &= f_2(X_1, X_2), \end{aligned} \tag{2.110}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are two functions mapping X_1, X_2 to Y_1 and Y_2, and X_1, X_2 are subject to power constraint $\mathbb{E}[|X_k|^2] \leq P_k,\ k = 1, 2$. An interesting sub-class of deterministic ICs with known capacity region [103, Eq. (1)-(4)] is

$$\begin{aligned} Y_1 &= f_1(X_1 + g_2(X_2)), \\ Y_2 &= f_2(X_2 + g_1(X_1)), \end{aligned} \tag{2.111}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are invertible functions. Due to the invertibility, (2.111) are

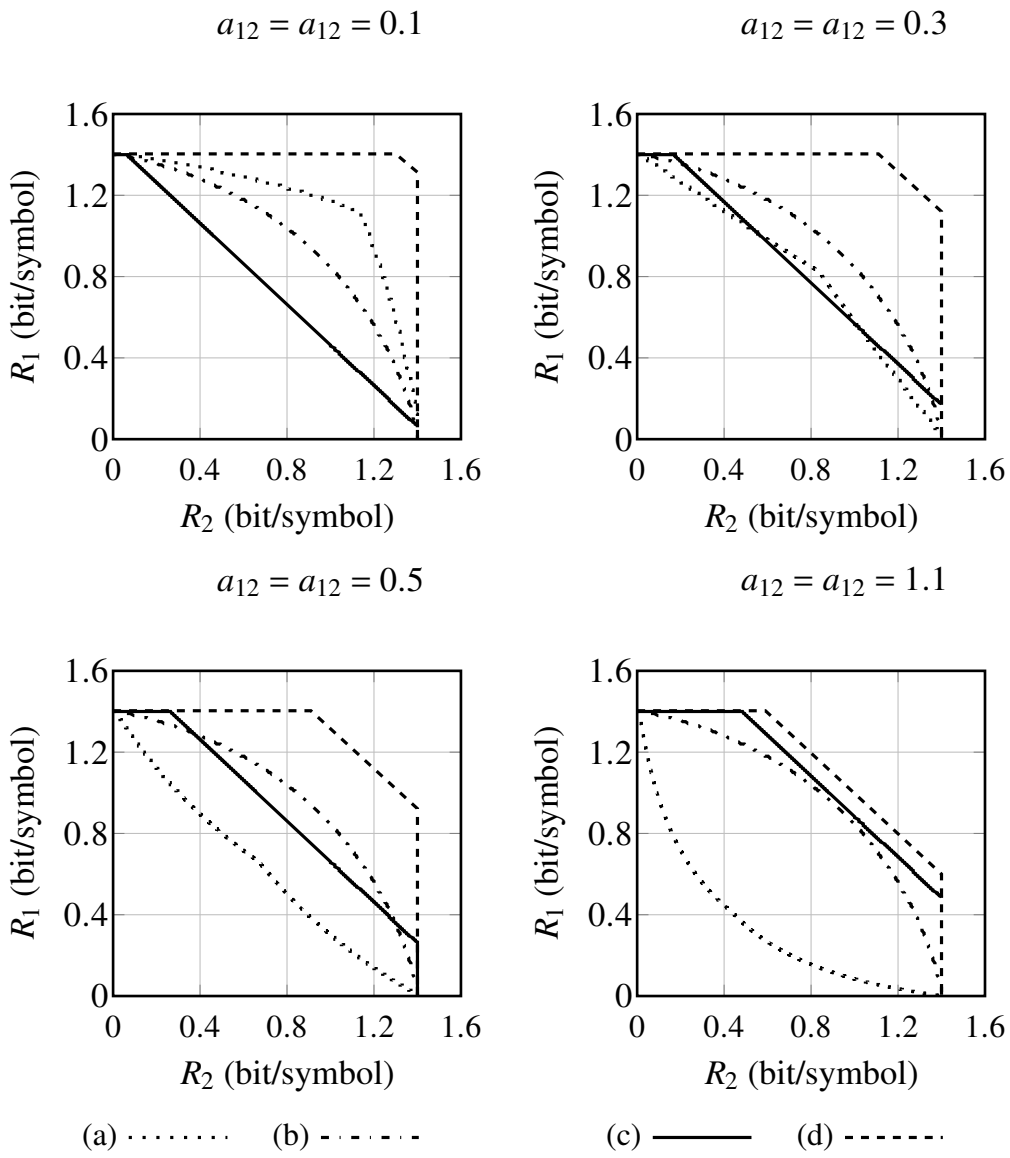

Figure 2.16: Rate regions for $P_1 = P_2 = 6$ and increasing $a_{12} = a_{21}$. The legends are for (a) treating interference as noise; (b) using TDM/FDM; (c) decoding interference; (d) Sato's outer bound.

said to be equivalent to

$$\begin{aligned} Y_1 &= X_1 + g_2(X_2), \\ Y_2 &= X_2 + g_1(X_1). \end{aligned} \tag{2.112}$$

The proved capacity region is expressed in terms of entropies, which makes it difficult to plot and provides little insight. We omit the result and the proof here and refer interested readers to [103].

Adding noise into (2.111), we have another slightly generalised IC

$$\begin{aligned} Y_1 &= f_1(X_1, g_2(X_2, N_1)), \\ Y_2 &= f_2(X_2, g_1(X_1, N_2)), \end{aligned} \tag{2.113}$$

where $f_1(X_1, \cdot)$ and $f_2(X_2, \cdot)$ are invertible functions, and $(N_1, N_2) \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$ are independent of X_1, X_2. Observe (2.113) closely, you can see that $f_1(\cdot)$ could represent the mapping between the input and output of an optical fibre channel and $g_2(\cdot)$ could correspond to the interference generating mechanism from adjacent channels/users with one noise source taken into account. In real channel, the two functions should work iteratively on their respective inputs, in a similar way as how the SSFM works in Sec. 2.2.5. We can only approximate the optical channel model with $f_1(\cdot)$ and $f_2(\cdot)$.

The above idea was further developed for K-user WDM optical fibre channel in the weak non-linear regime [104]. The Volterra series transfer function of single-mode fibres [105] is used to derive the corresponding $f_1(\cdot)$ and $f_2(\cdot)$ in the channel model (2.113). The obtained channel model for K users reads [104, Eq. (8)]

$$Y_i = X_i + \sum_{k=1}^{K} \sum_{\ell=1}^{K} \sum_{m=1}^{K} \xi_{k,\ell,m}^{(i)} X_k X_\ell^* X_m + N_i, \quad k = 1, 2, ..., K, \tag{2.114}$$

where X_i and Y_i are the complex-valued random variables at input and output of the optical fibre channel for ith user, N_i is the additive circularly symmetric complex Gaussian random variable with zero mean and variance σ^2, and $\xi_{k,\ell,m}^{(i)}$ is the interference coefficient that can be calculated from the Volterra series transfer function. Since we do not use $\xi_{k,\ell,m}^{(i)}$ to calculate any result here, we refer readers to [105, Appx. II] for the detailed derivation.

The conclusions from that work are twofold: 1) in the multi-user scenario, if the signal of all users are detected cooperatively for interference cancellatioin, the AWGN channel capacity is achievable in optical fibre channel, 2) if users detect their signal independently, the interference from other users cannot be cancelled and causes the channel capacity of the individual user to saturate at high signal powers. Admittedly, the model has its own limitations, for example, it only holds in the weak non-linear regime and it only includes one noise source. However, when we enter the highly non-linear regime, even if the inter-user interference is strong enough (such as $a_{12} > 1$ in (2.109)) to be beneficial to individual users, we would not be able to decode them because they originate from a complicated mixing process.

Once we establish that the WDM-based optical fibre network is a multi-user interference channel, we see that individual user detection can by no means achieve AWGN capacity. However, in a real optical fibre network, since we do not have all channels available at one location, a joint detection scheme is not possible. As a logical alternative, we should seek other strategies that are useful for general ICs, such as signal-space orthogonalization and interference alignment [106].

Non-linear frequency-division multiplexing (NFDM) is a way of signal-space orthogonalization. Despite that the theory of NFDM is based on the lossless NLSE, NFDM is a logical step for combating the non-linearity in optical fibre and will be closely examined in Chap. 4, 5 and 6.

2.4.5 Models and capacity of fibre channels

The capacity of the fibre channel is a haunting problem in optical communications. The complications of the problem include: a) the NLSE does not have an explicit closed-form solution, b) varying signal dimensions due to the existence of spectral broadening after propagation [107], which makes multi-sample detection reasonable and attractive [108], c) the optical fibre channel is in general a channel with memory [109], hence the MI between input and output has to be optimized across multiple dimensions, d) channels with in-line additive noises are in general not well studied in information theory.

Several simplified models for optical fibre channels have been proposed to facilitate the derivation of the channel capacity: the split-step Fourier model [110, 111] for the point-to-point optical fibre channel, the perturbative model [29, 104] for the optical fibre interference channel, and the phase noise model [91, 112] for both. They are summarised below.

The phase noise model uses uncorrelated (white) phase noise to model the XPM in the optical fibre interference channel or the SPM in the point-to-point optical fibre channel. The continuous-time channel model with input $X(t)$ and output $Y(t)$ can be written as

$$Y(t) = X(t)e^{j\Theta(t)} + N(t) \tag{2.115}$$

where $N(t)$ is a complex-valued AWGN process with variance $2\sigma_n^2$, and $\Theta(t)$ is the wrapped Gaussian process [112, Appx. A] with zero mean and variance σ^2 that models the phase noise process. The autocorrelation function $R_\Theta(\tau)$ of the

multiplicative noise $e^{j\Theta(t)}$ is [91, Eq. (2)]

$$R_\Theta(\tau) = E[e^{j\Theta(t)}e^{-j\Theta(t)}] = \begin{cases} 1, & \tau = 0 \\ e^{-\sigma^2}, & \tau \neq 0 \end{cases} \tag{2.116}$$

$$= e^{-\sigma^2} + \lim_{B\to\infty}(1 - e^{-\sigma^2}) \cdot \mathrm{sinc}(B\tau). \tag{2.117}$$

The Wiener-Khinchin theorem states that, if $x(t)$ is a wide-sense stationary process such that its autocorrelation function can be defined as $r_{xx}(\tau) = E[x(t)x^*(t-\tau)]$, then its power spectral density function $S_X(f)$ is the Fourier transform of its autocorrelation function $r_{xx}(\tau)$. Therefore, the PSD $S_Y(f)$ of the received signal $Y(t)$ is

$$\begin{aligned} S_Y(f) &= \mathcal{F}\{R_Y(\tau)\} = S_X(f) \otimes S_\Theta + S_W(f) \\ &= e^{-\sigma^2} S_X(f) + \lim_{B\to\infty} S_X(f) \otimes (1 - e^{-\sigma^2}) \frac{\mathrm{rect}(f/B)}{B} + S_W(f), \end{aligned} \tag{2.118}$$

where $\mathcal{F}\{\cdot\}$ denotes the Fourier transform and $\otimes$ is the convolution operator. (2.118) shows that the PSD of $S_X(f)$ experiences a spectral loss of $1 - e^{-\sigma^2}$, $e^{-\sigma^2} < 1$. The lost power spreads to an infinite bandwidth as suggested by the convolution operation, which agrees with the spectral broadening observed in the optical fibre [113]. Consequently, the phase noise channel is considered to be equivalent to a discrete-time AWGN channel with energy loss

$$Y = \mu_\Theta X + W, \tag{2.119}$$

so the effect of the spectral loss is included as SNR loss with the coefficient $\mu_\Theta = e^{-\sigma^2}$. In the limit of $\mu_\Theta \to 0$, no reliable communication is possible. If we make $\sigma^2 = c\mathcal{P}_s^2$, where $\mathcal{P}_s$ denotes the average signal power, the phase noise channel behaves in a similar way that the point-to-point optical fibre channel does [112, Fig. 8], but it fails to capture other effects in the optical fibre interference channel such as XPM.

The perturbative model is based on some form of approximated solutions of NLSE. For instance, the well-known GN model is built on the first-order perturbation solution of NLSE [29]. The root mean squared error (RMSE) between first-order perturbation solution and true solution can be measured to indicate at which power the perturbation theory fails. The interference from other channels is implicitly treated as an AWGN whose variance is proportional to the total received signal power $\mathcal{P}^3(z)$. The received SNR can be predicted by

$$\text{SNR} = \frac{\mathcal{P}(0)}{\sigma_N^2 + \sigma_{\text{NLI}}^2}, \tag{2.120}$$

where the σ_N and σ_{NLI} are the vriances of ASE noise and the non-linear interference, respectively. When a non-linear compensation scheme is applied at the receiver, σ_{NLI}^2 will be reduced accordingly. As a further development, an enhanced Gaussian noise (EGN) model was also proposed to correct the overestimation of non-linear interference along the link due to the non-Gaussianity of the input signals [30, and references therein]. Overall, the GN and EGN models have been validated to be able to make accurate predictions, in a very efficient way, of system performance (SNR or maximum reach) for dispersion-unmanaged coherent systems.

Another example of a perturbation model would be the application of Volterra series solution of NLSE [104] as explained in (2.114).

The split-step Fourier model proves that the capacity of AWGN channel also serves as an upper bound for a cascade of non-linear and noisy channels, which corresponds to an optical fibre link with distributed Raman amplification [110]. The authors of [110] made use of the fact that non-linear step in the NLSE does not change the differential entropy of a signal. This result settled down the debate on whether non-linearity can bring the information rate in the optical fibre channel beyond the AWGN capacity.

The above-mentioned models and capacity bounds are listed in the order of their closeness to the realistic fibre channel and are plotted for a specific link in Fig. 2.17. The phase noise channel is far from emulating the optical fibre channels, but in the case of the point-to-point link, it successfully captures the SPM effects and achieves a good agreement with simulation results regarding the achievable rates. The split-step Fourier model includes all the major impairments in the fibre, i.e., dispersion, non-linearity, and in-line noise. The model has the advantage of being the first entirely non-linear model. However, the obtained upper bound on the capacity is monotonically increasing against launch power and deviates gradually from most existing lower-bounds that saturate at the optimal launch power, leaving a large unknown territory between the upper and lower bound as marked in gray colour in Fig. 2.17. The perturbation models have been validated to be very accurate in predicting system performance, but it restricted itself to the existing transmission scheme.

For further development of models or capacity bounds, situations are slightly different for the optical fibre interference and point-to-point channels. In the case of the fibre interference channel, it has almost become a consensus that the AIR of each user has a maximum regardless of the transmission scheme. In case of

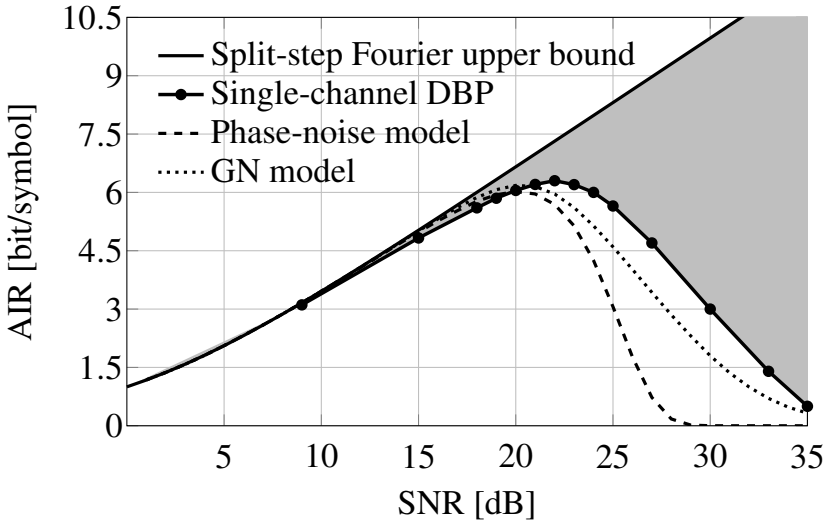

Figure 2.17: The AIR of the above models in a 4000 km optical fibre link.

the fibre point-to-point channel, it becomes very tough to rigorously prove tight upper- and lower-bound within 1 bit/s on the channel capacity without damaging the sufficient representation of the realistic channel. The cost of "mathematical correctness" needs to be balanced against the significance of results.

Algorithm 1: Standard RLS algorithm.

Data: $\mathbf{u}_i$: signals on both polarisations

Result: $\mathbf{h}_x$, $\mathbf{h}_y$: filter weights

Initialization;

$\mathbf{h}_{x/y}$ all zeros except a central spike, $\underline{\mathbf{P}}_{x/y} = \delta \underline{\mathbf{I}}_{2N}$, where δ is a small positive number, $\underline{\mathbf{I}}_{2N}$ the identity matrix of size 2N.

while $\epsilon > 0.001$ **do**

Gain vector computation

$$\begin{aligned}
\mathbf{z}_x(n) &= \mathbf{u}_i(n)\mathbf{u}_i^H(n)\mathbf{h}_x, \\
\mathbf{z}_y(n) &= \mathbf{u}_i(n)\mathbf{u}_i^H(n)\mathbf{h}_y, \\
\mathbf{k}_x &= \frac{\underline{\mathbf{P}}_x \mathbf{z}_x(n)}{\lambda + \mathbf{z}_x^H(n)\underline{\mathbf{P}}_x \mathbf{z}_x(n)}, \\
\mathbf{k}_y &= \frac{\underline{\mathbf{P}}_y \mathbf{z}_y(n)}{\lambda + \mathbf{z}_y^H(n)\underline{\mathbf{P}}_y \mathbf{z}_y(n)}.
\end{aligned}$$

Error estimation

$$\begin{aligned}
\epsilon_x(n) &= R_2 - \mathbf{h}_x^H \mathbf{z}_x(n), \\
\epsilon_y(n) &= R_2 - \mathbf{h}_y^H \mathbf{z}_y(n).
\end{aligned}$$

Tap-weights updates

$$\begin{aligned}
\mathbf{h}_x &\leftarrow \mathbf{h}_x + \mathbf{k}_x \epsilon_x^*(n) \\
\mathbf{h}_y &\leftarrow \mathbf{h}_y + \mathbf{k}_y \epsilon_y^*(n)
\end{aligned}$$

Correlation matrix updates

$$\begin{aligned}
\underline{\mathbf{P}}_x &\leftarrow (\underline{\mathbf{P}}_x - \mathbf{k}_x \mathbf{z}_x^H \underline{\mathbf{P}}_x)/\lambda \\
\underline{\mathbf{P}}_y &\leftarrow (\underline{\mathbf{P}}_y - \mathbf{k}_y \mathbf{z}_y^H \underline{\mathbf{P}}_y)/\lambda
\end{aligned}$$

Algorithm 2: The Arimoto-Blahut algorithm

Initialization:

$$f_X^{(0)}(x) \rightarrow \text{uniform distribution}$$

$$f_{X|Y}^{(0)}(x|y) = \frac{f_{Y|X}(y|x) f_X^{(0)}(x)}{\sum_X f_{Y|X}(y|x') f_X^{(0)}(x')}$$

$$\epsilon = \text{a reasonable small number}$$

while $\left\| f_X^{(k)}(x) - f_X^{(k-1)}(x) \right\| > \epsilon$ **do**

$$f_X^{(k)}(x) = \frac{\exp\left(\sum_Y f_{Y|X}(y|x) \log_2 \left(f_{X|Y}^{(k-1)}(x|y) \right) \right)}{\sum_X \exp\left(\sum_Y f_{Y|X}(y|x') \log_2 \left(f_{X|Y}^{(k-1)}(x'|y) \right) \right)}$$

$$f_{X|Y}^{(k)}(x|y) = \frac{f_{Y|X}(y|x) f_X^{(k)}(x)}{\sum_X f_{Y|X}(y|x') f_X^{(k)}(x')}$$

end

Return:

$$C = \sum_{X,Y} f_X^{(k)}(x) f_{Y|X}(y|x) \log_2 \frac{f_{X|Y}^{(k)}(x|y)}{f_X^{(k)}(x)}$$

Algorithm 3: The kpN algorithm.

Data:

- $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, ..., N$: the samples
- k: the number of nearest neighbors for calculating $\mathcal{B}(\epsilon, \mathbf{x}_i)$
- p: the number of nearest neighbors for calculating the local Gaussian approximation ($p > k$)

Result: $\hat{H}(\mathbf{x})$: the kpN entropy estimate

initialization;

$i = 1$;

while $i < N$ **do**

  $\{\mathbf{x}_i\}^p \leftarrow$ set of p-nearest neighbors of $\mathbf{x}_i$;

  $i \leftarrow i + 1$

$\hat{H}(\mathbf{x}) = \psi(N) - \psi(k)$;

$i = 1$;

while i<N **do**

  $\epsilon_i \leftarrow$ distance to the k-th nearest neightbor of $\mathbf{x}_i$;

  $\mathcal{B}(\epsilon, \mathbf{x}_i) \leftarrow \mathbf{x}_i \pm \epsilon_i \mathbb{E}$, $\mathbb{E}$ being the canonical basis;

  $\mu_i \leftarrow$ mean of $\{\mathbf{x}_i\}^p$;

  $\mathbf{S}_i \leftarrow$covariance of $\{\mathbf{x}_i\}^p$;

  $G_i \leftarrow$ Gaussian integral through EMPGP of μ_i and $\mathbf{S}_i$;

  $g(\mathbf{x}_i) \leftarrow$ equation (2.106);

  $\hat{H}(\mathbf{x}) \leftarrow \hat{H}(\mathbf{x}) + N^{-1}[\log(G_i) - \log(g(\mathbf{x}_i))]$;

3

Spectral broadening of Gaussian processes in optical fibre channel

3.1 Related work

The information-theoretic capacity of the optical fibre channel has been an unsolved problem in the optical communication society for a long time. An upper bound on the capacity of point-to-point optical fibre channel was proved to be the same as the capacity of the AWGN channel [110]. Significant efforts have been made to search for a non-saturating lower bound [37, 114–116]. However, as one can see from Fig. 2.17, the upper and lower bounds on the capacity are loose at the high power range. Among the many problems encountered in the quest for the unknown territory, a subtle one is how to define the transmission bandwidth and in particular, the spectral broadening along propagation at all range of powers.

In this chapter, the problem of upper bound on the optical fibre channel capacity is approached by studying and characterising the spectral evolution of a Gaussian process in optical fibre in a semi-analytical way. The KZ model, explained in Sec. 2.2.3, is used to calculate the PSD at the end of a link. The KZ model originates from the theory of WWT [59, pp. 63-82]. The WWT framework is of particular interest for Raman fibre laser modelling [117, 118] because such lasers often operate at high signal power that induces strong non-linear wave interactions.

Two published works [56, 119] both provide detailed derivation and explanation for the application of WWT in NLSE, with the former in the context of fibre laser (single-span fibre) and the latter focusing on the long-haul optical fibre communication (multi-span fibre). As shown in Fig. 2.6, the KZ model can estimate the PSD at moderate signal power and link distance much better than the GN model does. However, as with any perturbation-based solution, the model breaks down as the strength of accumulated non-linear effect increases through either increase of signal power or link distance. Fig. 3.1 shows the breakdown of the KZ model at the signal power that is twice of the one in Fig. 2.6. The irregular oscillation around the one sample frequency, in this case ±16 GHz, is the typical sign of the breakdown. It is also worth noticing that the KZ model breaks down at much lower power if the RRC-pulses have smaller roll-off factor.

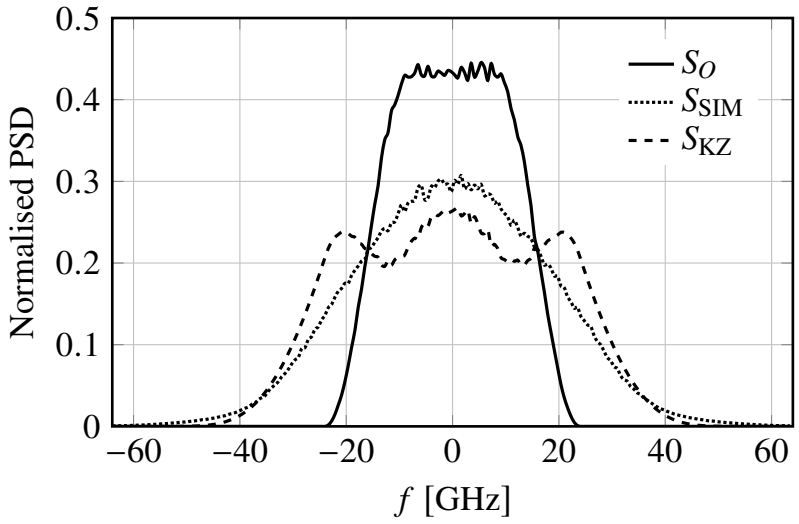

Figure 3.1: PSDs from simulations scaled by the maximum of the GN PSD with same link parameters and doubled signal power of Fig. 2.6.

There are, in general, two ways to increase the range of validity of a perturbative model, i.e., increase the order of the perturbation solution or solve the perturbative model iteratively for several times. We found that the latter approach is simpler and more effective. Therefore, at high power regime, the PSD at receiver is calculated by iteratively solving KZ model at a shorter distance.

We will use the in-band energy ratio ρ and 99%-energy bandwidth expansion factor η to characterise the spectral broadening. We show numerically that the in-band energy ratio is related to signal power P, link distance L, and initial signal bandwidth W. Furthermore, the in-band energy ratio can be approximately described by a monotonically decreasing function $\rho \approx f_\rho(\frac{P\sqrt{L}}{W^{1.33}})$, where $f_\rho(x) = 1/(1+\log(1+\xi x))$ and ξ is a fitting parameter. The result implies that: (i)

Given a fixed processing bandwidth, increasing signal power of a Gaussian process causes more severe spectral broadening, limiting the achievable SE, (*ii*) if the signal power and bandwidth scale up at the same time and the signal power increases at a faster rate than the bandwidth does, the spectral broadening can be negligible. Hence, an ever-increasing SE might become possible.

3.2 The optical fibre model

The optical fibre channel of concern is a multi-span point-to-point dispersion unmanaged fibre link with equally-spaced EDFAs, with signal only present in one polarisation. The single-span wave propagation is described by the NLSE (2.13). For convenience we rewrite it here

$$\frac{\partial A}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{\alpha}{2}A + j\gamma|A|^2 A, \tag{3.1}$$

where $A(t,z)$ is the complex envelope of the signal as a function of time t and distance z along the optical fibre, β_2, γ, and α are the dispersion, non-linear, and loss coefficients of the optical fibre; see Tab. 3.1. Lumped amplification is performed at the end of each span with the power gain $g = \exp(\alpha L_{\text{sp}})$, where L_{sp} is the span length and N_{sp} is the number of spans. In order to see the spectral evolution clearly, we have turned off the ASE noise introduced by EDFAs.

Table 3.1: Parameters of the optical fibre communication system.

ν	193.44 THz	centre carrier frequency
α	0.046 km^{-1}	fibre loss
γ	1.27 $(\text{W.km})^{-1}$	non-linearity coefficient
β_2	-21.5×10^{-27} s^2/m	dispersion coefficient
W	32/64/128 GHz	optical bandwidth
R_o	4	oversampling rate
r	0.01	RRC roll-off factor
L_{sp}	80 km	span length
N_{sp}	20/40/80	number of spans
L	$L_{\text{sp}}N_{\text{sp}}$	total length

It is assumed the input signal $A(t,0)$, $0 < t < T_0$, is a band-limited complex Gaussian (cyclo) stationary process, defined by its Fourier series

$$A(t,0) = \sum_{k=-\infty}^{\infty} \tilde{A}_k \exp\left(j2\pi\frac{k}{T_0}t\right), \quad 0 < t < T_0. \tag{3.2}$$

The Fourier coefficients $\tilde{A}_k$ are independent symmetrically-complex Gaussian ran-

dom variables with zero mean and the variance S_k. For a band-limited signal of bandwidth W, we have $S_k = 0$ for $|k| > \lfloor WT_0 \rfloor$. The PSD of $A(t)$ is

$$S_q(f) = S_k \delta(f - \frac{k}{T_0}). \tag{3.3}$$

Note that any desired PSD can be obtained through a suitable linear filter. Such an input signal satisfies the necessary conditions of the KZ model [56] that (i) $\tilde{A}_k$ are uncorrelated and, (ii) $A(t,z)$ has quasi-Gaussian distribution for all z (under sufficiently weak non-linearity framework). We further assume that the Fourier coefficients stay uncorrelated during the transmission. This assumption should be justifiable by the fact that the FWM in (2.21) is among many frequency components, and according to central limit theorem, its effect is close to a normal distribution.

Although the Gaussian process is a restrictive assumption, it has some practical relevance. For example, in current coherent optical fibre systems, signals generated by probabilistic amplitude shaping are very close to Gaussian process. Such signals are of great interest for the data rate flexibility, high SE as well as increasing the transmission distance. Moreover, under sufficiently weak non-linearity [29, 120], the distribution of transmitted symbols converges anyhow to a quasi-Gaussian distribution after a long enough (non-linear length) propagation distance.

We wish to estimate the PSD of NLSE using the KZ model, and adopted the iterative method to enlarge the range of validity of the KZ model.

3.3 The iterative Kolmogorov-Zakharov model

The iterative KZ model is formulated in this section. Let Δz denote an integer multiple of span length L_{sp}. We define $S_k[i]$ as the PSD of $A(t, z = i\Delta z)$ after propagation of $i\Delta z / L_{\text{sp}}$ spans. At the transmitter, $S_k[0] = S_k(0)$. The output PSD $S_k(L)$ can be iteratively approximated as

$$S_k[0] \xrightarrow[\Delta z]{\text{KZ}} S_k[1] \cdots\cdots S_k[N-1] \xrightarrow[\Delta z]{\text{KZ}} S_k[N] \tag{3.4}$$

$$S_k[i+1] \approx S_k[i] + 2\gamma^2 \sum_{(l,m,n)\in V(k)} |\tilde{H}_{\ell mnk}(\Delta z)|^2 T_{\ell mnk}[i] \delta_{\ell mnk}. \tag{3.5}$$

(3.5) is the first order perturbation approximation of the non-linear kinetic equation described earlier in (2.45). The accuracy of approximation depends on how small Δz is chosen and also on the signal power. For relatively small powers, one single iteration ($\Delta z = L_{\text{sp}} N_{\text{sp}}$) is quite accurate. For a higher range of powers, more iterations are required with smaller Δz as illustrated in Fig. 3.2. The number of iterations, which varies with signal power and bandwidth, were empirically obtained in

Fig. 3.2. "Empirically" means that for a certain input PSD, one applies KZ model with no or a few iterations first. If the irregular oscillation at one sample frequency is observed in the output PSD, one simply increases the number of iterations until a smooth output PSD is obtained. Note that $\Delta z = L_{\text{sp}}$ is the smallest Δz.

3.4 Characterisation of spectral broadening

Two figures of merit were chosen to characterise the spectral broadening: (i) ρ: the in-band energy ratio, the fraction of the energy preserved in the initial bandwidth W. (ii) η: the bandwidth expansion factor. The bandwidth is defined as the smallest frequency band holding more than 99% of energy. The precise definition of signal bandwidth has never been without controversy [121]. The 99%-energy bandwidth criteria is adopted from [107].

We first compare the analytic estimation of (3.4) with split-step Fourier simulations. The simulation parameters are outlined in Tab. 3.1. The ASE noise was switched off in the simulation. Fig. 3.2 and 3.3 show the variation of ρ and η in terms of input launch power for different input signal bandwidth W and for different number of spans. We observe that the analytic predictions (dashed lines) are quite consistent with the simulation results (dots). For not very large values of launch power ($\rho > 0.9$), ρ can be empirically approximated by,

$$\rho(W, L, \mathcal{P}) \approx \frac{1}{1 + \ln\left(1 + \zeta \frac{\mathcal{P}^2 L}{W^{2.65}}\right)}, \tag{3.6}$$

where $L = L_{\text{sp}} N_{\text{sp}}$ is the total distance, ζ is a constant, depending on the link parameters and the span length, $\mathcal{P}$ is the launch power in linear-scale (not in dBm). We observe in Fig. 3.2 and 3.3 that the above approximation fits closely to the simulation for $\mathcal{P} < 10$ dBm. For larger values of $\mathcal{P}$, we observe that the scaling behaviour of $\rho(W, L, \mathcal{P})$ and $\eta(W, L, \mathcal{P})$ tends to become a function of $\frac{\mathcal{P}\sqrt{L}}{W^2}$. It means that

$$\rho \approx f_\rho\left(\frac{\mathcal{P}\sqrt{L}}{W^2}\right) = \frac{1}{1 + \ln\left(1 + \zeta \frac{\mathcal{P}^2 L}{W^4}\right)}, \quad \eta \approx f_\eta\left(\frac{\mathcal{P}\sqrt{L}}{W^2}\right). \tag{3.7}$$

Fig. 3.4 also shows two examples of spectral broadening, calculated by KZ-model, of 32 GHz signals after 1600 km propagation.

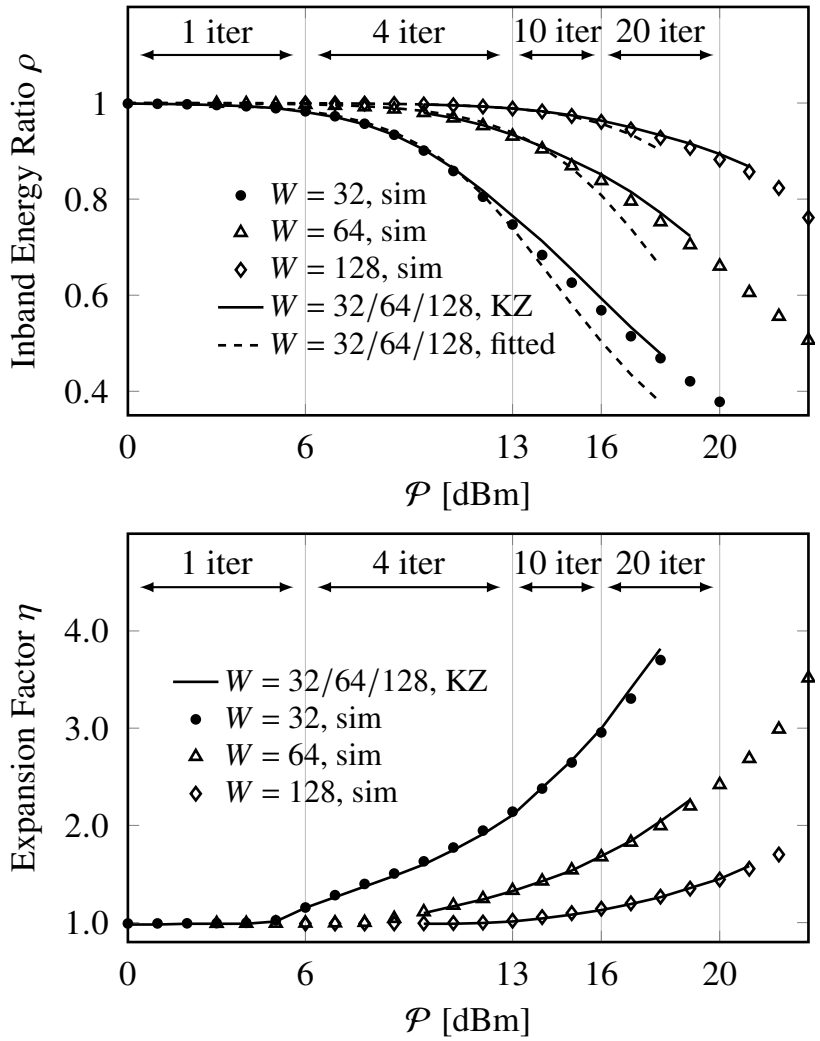

Figure 3.2: Characterisation of spectral broadening in a 20-span transmission link with different initial bandwidths of 32/64/128 GHz. The number of iterations holds for 32 GHz bandwidth. Simulations are with 16-QAM 0.01-RRC-pulse signals.

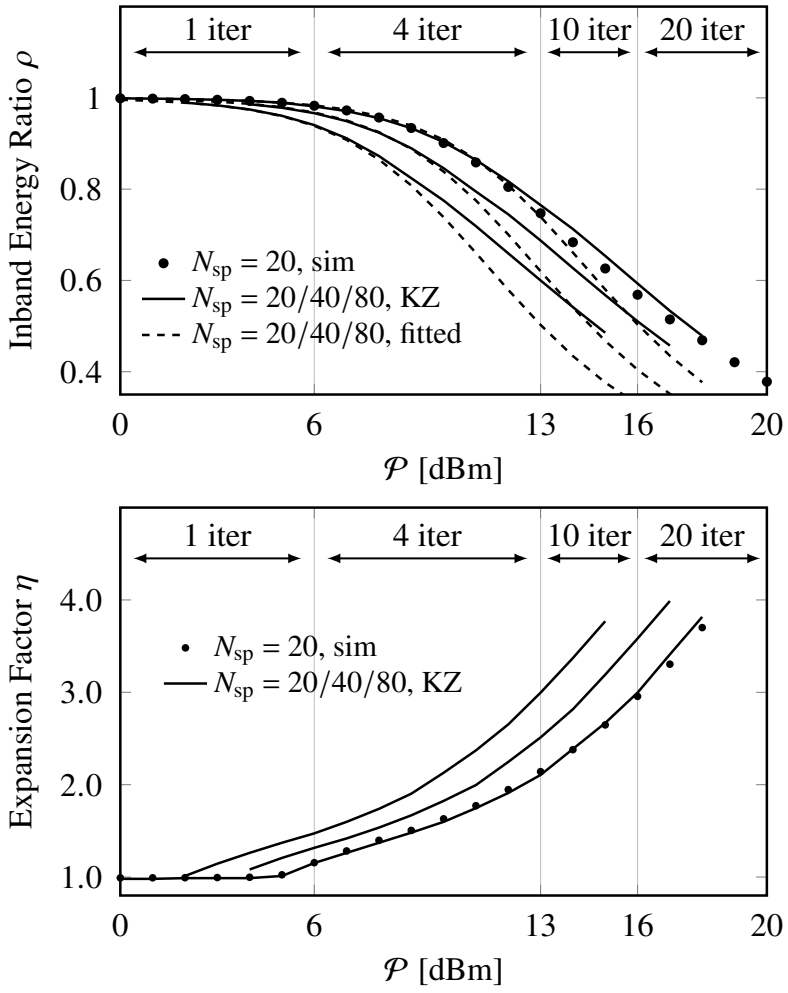

Figure 3.3: Characterisation of spectral broadening of signals with 32 GHz initial bandwidth at different transmission distances (20/40/80 spans).

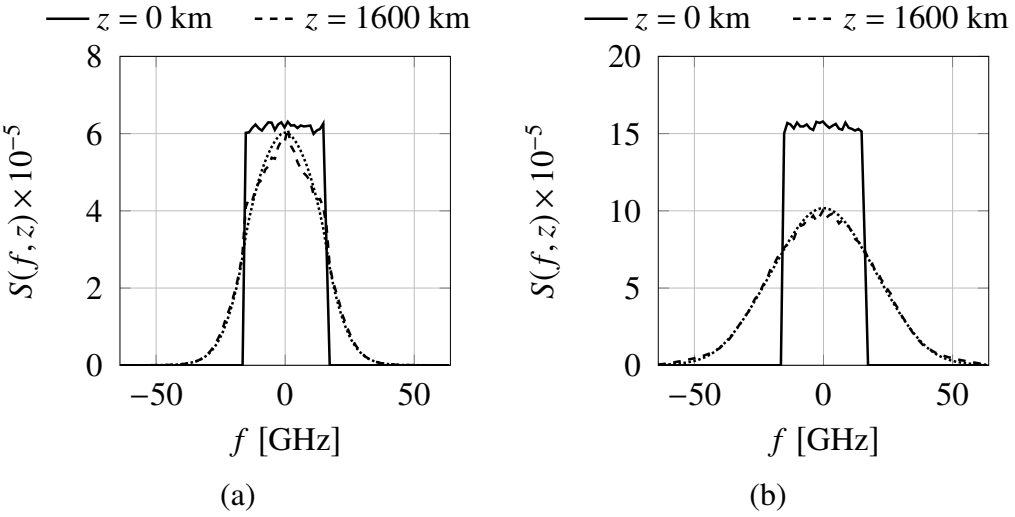

Figure 3.4: Spectra of a 32 GHz signal at (a) 12 and (b) 16 dBm. Solid, dashed and densely dotted lines are input, output, and KZ model signal spectra, respectively.

3.5 Discussion on the spectral efficiency

It is shown recently [110] that the SE of the channel model (3.1), with ideal distributed amplification, is upper-bounded by the AWGN capacity

$$\mathrm{SE} \leq \log\left(1 + \frac{\mathcal{P}}{N_{\mathrm{ase}} W_{\max}}\right) \tag{3.8}$$

where N_{ase} is the noise PSD and $W_{\max}$ is the maximum signal bandwidth. A recent result [107] shows that, for dispersion-less fibre, fixed initial bandwidth and large power, the bandwidth scales at least as a constant times $\sqrt{\mathcal{P}}$. For the optical fibre channel with dispersion, we show a similar scaling law in (3.7) for Gaussian process.

Let us assume that, for EDFA-amplified links, the upper-bound (3.8) is still valid, and we can look at how SEs behave for different types of receivers. Receivers with different limitations on its bandwidth are considered here. For receivers with fixed-bandwidth (W) and variable-bandwidth ($f_\eta(\frac{\mathcal{P}\sqrt{L}}{W^2})W$), the SE should be bounded by

$$\mathrm{SE} \leq \max\left\{\log\left(1 + \frac{1}{f_\eta(\frac{\mathcal{P}\sqrt{L}}{W^2})}\frac{\mathcal{P}}{N_{\mathrm{ase}}W}\right),\ \log\left(1 + f_\rho(\frac{\mathcal{P}\sqrt{L}}{W^2})\frac{\mathcal{P}}{N_{\mathrm{ase}}W}\right)\right\}. \tag{3.9}$$

It can be seen from Fig. 3.2 that $1/f_\eta(\frac{\mathcal{P}\sqrt{L}}{W^2})$ is smaller than $f_\rho(\frac{\mathcal{P}\sqrt{L}}{W^2})$ over the simulated power range. Furthermore, for a given signal power $\mathcal{P}_0$ and bandwidth W_0, if the signal power $\mathcal{P}_0$ is multiplied by a factor θ while the signal bandwidth W_0 is multiplied by a factor $\sqrt{\theta}$, the ratio $\frac{\mathcal{P}}{W^2} = \frac{\theta\mathcal{P}_0}{(\sqrt{\theta}W_0)^2} = \frac{\mathcal{P}_0}{W_0^2}$ remains unchanged. The overall change to the term $f_\rho(\frac{\mathcal{P}\sqrt{L}}{W^2})\frac{\mathcal{P}}{N_{\mathrm{ase}}W}$ inside the upper bound on SE is then $\sqrt{\theta}$. This indicates that if $\mathcal{P}$ and W scale up at the same time and $\mathcal{P}$ increases at a faster rate than W does, the spectral broadening effect stays the same while SE increases boundlessly. It is worth noting that to increase the SE in this way is very slow and costly. For example, for a given $\mathcal{P}_0$ and W_0, if one wishes to have a four time increase on the term $f_\rho(\frac{\mathcal{P}\sqrt{L}}{W^2})\frac{\mathcal{P}}{N_{\mathrm{ase}}W}$ without changing the spectral broadening effect, the signal power needs to be increased by 16 times and bandwidth by 4 times.

For receivers with intermediate bandwidth, i.e., $W < W_{\mathrm{Rx}} < f_\eta(\frac{\mathcal{P}\sqrt{L}}{W^2})W$, it is beneficial to be able to receive the spread-out signal power and at the same time not take in too much out-of-band noise. It is expected and observed that receiving extra spectrum leads to an increased SE [113] up to a point where too much noise starts to level off the benefit. However, we have not managed to bound the trade-off.

Note that our scaling laws are obtained for the Gaussian process and based on

the result from iterative KZ-model without any ASE noise. Therefore, the scaling law can be different for an arbitrary input process or for the highly non-linear regime in which the KZ model is not valid any more.

3.6 Summary

In this chapter, the spectral evolution of Gaussian process with iterative KZ model solution for a large range of input powers has been analysed. The spectral boradening phenomenon is one of the many obstables we face in our quest into the gray area in Fig. 2.17.

From the analysis presented in this chapter, one can reach the conclusions that 1) for a fixed Rx bandwidth, the SE of Gaussian process is limited by the effect of spectral broadening, 2) an ever-increasing SE might be possible if the signal bandwidth and power are increased at the same time and the signal power increases at a faster rate than bandwidth does.

The following chapters focus on a non-linear transmission scheme that has the potential to achieve a monotonically-increasing SE in ORN.

4

Nonlinear Fourier Transform

In the previous chapter, research results have shown that Gaussian process suffers from significant spectral broadening in the non-linear regime, and thus, is not practical to achieve a monotonically increasing SE. In addition, spectral broadening is not the only non-linear effect that limits the network throughput. It is also believed that the linear signal multiplexing scheme (WDM) in ORN is not compatible with the non-linear nature of optical fibre channel [79].

WDM is one of the key technologies that enable high capacity optical fibre networks, which allocates different frequency bands to different users and makes efficient use of the available fibre bandwidth. However, driven by the increased traffic demand, WDM systems are now approaching their theoretical limits set by fibre non-linearity as shown in Fig. 1.7. The non-linear interactions among WDM channels pose a limitation to the AIR in WDM optical networks [28].

WDM systems modulate information in the Fourier basis corresponding to different frequencies or wavelengths. The basis elements could be modulated independently at the transmitter, but they couple together in the non-linear fibre channel. Consequently, the optical fibre channel should be viewed as a multi-user interference channel (IC). IC was briefly introduced in Sec. 2.4.4. Several distinct strategies exist for communication over an IC: (i) treating interference as noise, (ii) orthogonalisation and (iii) interference alignment. An extensive body of work exists presenting the AIRs of WDM optical networks, which are sometimes referred to as the

"non-linear Shannon limit" [5,28,122–124]. However, these rates are lower bounds on the capacity of the user-of-interest (UOI), implicitly using strategy (i). Other strategies for WDM channels, such as (ii) and (iii), exist as well, which predict different AIRs.

Non-linear frequency-division multiplexing (NFDM), a recently proposed [77] signal multiplexing scheme based on the non-linear Fourier transform (NFT), is a promising approach to overcome the "capacity crunch" problem in WDM optical systems [79], and could potentially achieve a monotonically increasing SE in ORN. It is often argued that the idea of NFDM should be traced back to the concept of eigenvalue communication [125] back in 1993. NFT, also known as the inverse scattering transform (IST) [126], is an established mathematical tool to solve the initial boundary problem of NLSE. A key property of NFT is that in the non-linear Fourier domain, the non-linear optical fibre channel is linearised. In the non-linear Fourier domain, one can potentially assign non-linear frequency bands to different users, hence the name NFDM, so that users can share the transmission medium without interference.

While WDM treats interference as noise, NFDM is an approach to channel orthogonalisation. NFDM is our candidate transmission scheme to potentially achieve monotonically increasing AIR in an ORN.

In a NFDM system, information is encoded on non-linear frequencies, which is defined by the NFT of a time-domain signal. The non-linear frequencies consist of real-valued frequencies and complex frequencies (called eigenvalues). They have the advantage of being independent components in the NLSE for single polarisation (SP) or Manakov equation for dual polarisation (DP). Interference-free or weak-interference transmission seems possible in a network environment as conjectured in [77]. NFDM suffers significantly less from inter-channel interference (ICI) than the conventional linear multiplexing schemes as shown in [115, 127] utilising the so-called continuous spectrum that contains real-valued frequencies.

The research described in the remaining chapters (Chap. 4, 5 and 6) will be focused on the investigation of NFT-based transmission schemes. This chapter is dedicated to the principle of NFT with minimum mathematics, the numerical methods to calculate NFT and INFT, and the transceiver schemes based on NFT-INFT pairs.

4.1 Review of single-polarisation NFT

The development of NFT is attributed to the extraordinary work of mathematicians in 1960s-70s who found the exact solution of certain non-linear evolution equations. Different from linear equations, whose solutions are often superposition of

special solutions (or eigenfunctions), e.g., exponential functions, non-linear evolution equations, in general, do not have solutions that can be formulated out of regular solutions. In 1960s Kruskal and Zabusky [128] first discovered the soliton solution for the Korteweg-de Vries (KdV) equation and explained its existence and the underlying principle. Their result paved the way for Lax's general theory [129] that the wave evolution in the integrable[1] non-linear equation can be associated with linear operators so that the eigenvalues of the linear operator are invariant functionals (also called integrals) of the non-linear equation. For optical fibre communications, we can use the framework of NFT in the integrable (lossless) NLSE

$$\frac{\partial A}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + j\gamma|A|^2 A. \tag{4.1}$$

The integrable model deviates from our channel model (2.13) by one attenuation term. This deviation will be addressed later in Sec. 4.5. In the soliton theory, it is customary to normalise (4.1) by the following change of variables

$$q = \frac{A}{A_n}, \qquad \ell = \frac{z}{L_n}, \qquad \tau = \frac{t}{T_n}, \tag{4.2}$$

$$A_n = \sqrt{2/(\gamma L)}, \quad L_n = \operatorname{sgn}(\beta_2)L, \quad T_n = \sqrt{|\beta_2|L/2}, \tag{4.3}$$

where β_2, γ, and L are defined in Tab. 3.1. The normalised integrable NLSE reads

$$j\frac{\partial q(\tau,\ell)}{\partial \ell} = \frac{\partial^2 q(\tau,\ell)}{\partial \tau^2} - 2\operatorname{sgn}(\beta_2)|q(\tau,\ell)|^2 q(\tau,\ell), \tag{4.4}$$

where the function $\operatorname{sgn}(\cdot)$ denotes the sign function. Alternatively, one can use another set of normalisation factors [67, Eq. (5.2.1)]

$$q = \frac{NA}{P_0}, \qquad \ell = \frac{z}{L_D}, \qquad \tau = \frac{t}{T_0}, \qquad N^2 = \frac{\gamma P_0 T_0^2}{|\beta_2|}, \tag{4.5}$$

where P_0 is the peak power, T_0 is the width of the incident pulse, $L_D = T_0^2/|\beta_2|$ is the dispersion length. The normalised NLSE becomes

$$j\frac{\partial q(\tau,\ell)}{\partial \ell} - \operatorname{sgn}(\beta_2)\frac{1}{2}\frac{\partial^2 q(\tau,\ell)}{\partial \tau^2} + |q(\tau,\ell)|^2 q(\tau,\ell) = 0. \tag{4.6}$$

[1]The concept of 'integrability' is difficult to define within the current mathematical scope [130, Intro.]. We can, however, use the three identifier of integrable systems proposed by Nigel Hitchin [131]: the existence of many conserved quantities, the presence of algebraic geometry, and the ability to give explicit solutions.

We choose the first set of normalisation factors because they depend only on the link parameters, not on the characterisation of the incident pulse.

According to Lax's theory, for integrable evolution equations such as the NLSE, one can find a unique operator $L(q(\ell,\tau))$ whose spectrum does not change along ℓ. The spectrum of an operator is similar to the spectrum of a matrix (the set of discrete eigenvalues), apart from the fact that an operator is an infinite-dimensional matrix and its spectrum has both discrete and continuous eigenvalues. The unique operator L for the NLSE is

$$L(\ell,\tau) = j\begin{pmatrix} \frac{\partial}{\partial\tau} & -q(\ell,\tau) \\ \operatorname{sgn}(\beta_2)q^*(\ell,\tau) & -\frac{\partial}{\partial\tau} \end{pmatrix}. \tag{4.7}$$

We wish to calculate the invariant eigenvalues of the operator L. The eigenvalues of the operator L are defined via the eigenvalue equation

$$L\mathbf{v} = \lambda\mathbf{v} \tag{4.8}$$

where v is the eigenvector. We expand the (4.8) and reorganise it

$$\begin{cases} j\frac{\partial}{\partial\tau}v_1 - jq\cdot v_2 = \lambda v_1 \\ j\operatorname{sgn}(\beta_2)q^*v_2 - j\frac{\partial}{\partial\tau}v_2 = \lambda v_2 \end{cases} \rightarrow \quad \begin{cases} \frac{\partial}{\partial\tau}v_1 = -j\lambda v_1 + qv_2 \\ \frac{\partial}{\partial\tau}v_2 = \operatorname{sgn}(\beta_2)q^*v_1 + j\lambda v_2 \end{cases}. \tag{4.9}$$

Writing the above equation in the matrix form, we obtain

$$\frac{\partial\mathbf{v}(\ell,\lambda,\tau)}{\partial\tau} = P\mathbf{v} = \begin{pmatrix} -j\lambda & q \\ \operatorname{sgn}(\beta_2)q^* & j\lambda \end{pmatrix}\mathbf{v}, \quad \lambda\in\mathbb{C}. \tag{4.10}$$

(4.10) is the well-known Zakharov-Shabat (ZS) system [132, Eq. (8)]. The operator P, together with another operator M, will form another complete representation of the NLSE through the following relation

$$P_z - M_t + PM - MP = 0, \tag{4.11}$$

$$M = \begin{pmatrix} 2j\lambda^2 + j\operatorname{sgn}(\beta_2)|q|^2 & -2\lambda q - j\frac{\partial q}{\partial t} \\ \operatorname{sgn}(\beta_2)\left(-2\lambda q^* + j\frac{\partial q^*}{\partial t}\right) & -2j\lambda^2 - j\operatorname{sgn}(\beta_2)|q|^2 \end{pmatrix}. \tag{4.12}$$

One can easily verify that (4.11) reduces to the NLSE by substitution.

To proceed, we also need to make some assumptions on the signal $q(\tau,\ell)$: 1) $q(\tau)\in L^1(\mathbb{R})$, i.e., $\|q(\tau)\|_1 < \infty$, 2) $q(\tau)\to 0$ as $|t|\to\infty$. Through the study of the ZS system, we can find the spectrum of the operator L and the spectral amplitude

(or non-linear coefficient) defined on every spectral point. From the assumptions on the signal $q(\tau)$, we can obtain the boundary condition for the ZS system

$$\mathbf{v}(\tau,\lambda) = \begin{pmatrix}1\\0\end{pmatrix} e^{-j\lambda\tau}, \quad \tau \to -\infty. \tag{4.13}$$

The non-linear coefficients $\{a(\lambda),\ b(\lambda)\},\ \lambda \in \mathbb{C}$ can be found by solving the ZS system to $\tau \to +\infty$ with the boundary condition (4.13) at $\tau \to -\infty$

$$a(\lambda) = e^{+j\lambda\tau} v_1(\tau,\lambda), \quad \tau \to +\infty,\ \ \lambda \in \mathbb{C}, \tag{4.14}$$

$$b(\lambda) = e^{-j\lambda\tau} v_2(\tau,\lambda), \quad \tau \to +\infty,\ \ \lambda \in \mathbb{C}. \tag{4.15}$$

In general, the operator L has both discrete and continuous spectrum denoted as $\lambda_m \in \mathbb{C}^+$ and $\lambda \in \mathbb{R}$, respectively. The discrete eigenvalues of the operator L can be found via

$$\Omega = \{\lambda_m | a(\lambda_m) = 0,\ \lambda_m \in \mathbb{C}^+\}. \tag{4.16}$$

We can now define NFT as a mapping from a time-domain signal $q(\tau)$ to its non-linear spectrum and spectral amplitudes

$$\mathrm{NFT}(q(\tau)) = \begin{cases} \Omega = \{\lambda_m | a(\lambda_m) = 0,\ \lambda_m \in \mathbb{C}^+\}, \\ \big\{b(\lambda_m), a'(\lambda_m)\big\}, \quad \lambda_m \in \Omega, \\ \big\{b(\lambda), a(\lambda)\big\}, \quad \lambda \in \mathbb{R}, \end{cases} \tag{4.17}$$

where $a'(\lambda) = \frac{\partial a(\lambda)}{\partial \lambda}$. (4.17) can also be reduced to just ratios of the a- and b-coefficients [77, 126, 133]. They are referred to as the continuous and discrete spectral amplitudes, respectively

$$\mathrm{NFT}(q(\tau)) \begin{cases} \Omega = \{\lambda_m | a(\lambda_m) = 0,\ \lambda_m \in \mathbb{C}^+\}, \\ \frac{b(\lambda_m)}{a'(\lambda_m)}, \quad \lambda_m \in \Omega, \quad \text{discrete spectral amplitude } \ q_d(\lambda_m), \\ \frac{b(\lambda)}{a(\lambda)}, \quad \lambda \in \mathbb{R}, \quad \text{continuous spectral amplitude } \ q_c(\lambda). \end{cases} \tag{4.18}$$

Note that (4.17) and (4.18) are completely equivalent representations of the NFT. The most important property of the NFT is that if the evolution of $q(\tau,\ell)$ is governed by the NLSE, the non-linear spectrum and coefficients evolve in a simple linear

manner

$$\begin{cases} \lambda_m(\ell) = \lambda_m(0), & \\ b(\ell, \lambda_m) = b(0, \lambda_m) e^{4j\lambda_m^2 \ell}, & \lambda_m \in \Omega, \\ a'(\ell, \lambda_m) = a'(0, \lambda_m), & \lambda_m \in \Omega, \\ b(\ell, \lambda) = b(0, \lambda) e^{4j\lambda^2 \ell}, & \lambda \in \mathbb{R}, \\ a(\ell, \lambda) = a(0, \lambda), & \lambda \in \mathbb{R}. \end{cases} \tag{4.19}$$

Another more interesting representation of NFT is solely based on $b(\lambda)$. It exploits the uni-modularity condition [134, Eq. (A.3)] and the fact that $a(\lambda)$ is an analytic function [135, Lemma 2.1]. For an analytic function, its real and imaginary parts are Hilbert transforms of one another

$$a(\lambda) = \big|a(\lambda)\big| \exp\Big(j\mathcal{H}\big(\log\big|a(\lambda)\big|\big)\Big). \tag{4.20}$$

Moreover, the uni-modularity condition states

$$\big|a(\lambda)\big|^2 - \mathrm{sgn}(\beta_2)\big|b(\lambda)\big|^2 = 1, \quad \text{for } \lambda \in \mathbb{R}. \tag{4.21}$$

Based on (4.20) and (4.21), for the special case of our interest, i.e., no discrete spectrum, $a(\lambda)$ can be expressed in terms of $b(\lambda)$

$$a(\lambda) = \sqrt{1 + \mathrm{sgn}(\beta_2)\big|b(\lambda)\big|^2} \exp\left(j\mathcal{H}\left(\log\sqrt{1 + \mathrm{sgn}(\beta_2)\big|b(\lambda)\big|^2}\right)\right), \quad \lambda \in \mathbb{R}. \tag{4.22}$$

Note that the above expression is not always true but holds for finite-extent signals where $b(\lambda)$ is also an analytic function in $\mathbb{C}^+$. In this thesis, only finite-extent signals are considered. Similarly, $a(\lambda)$ can also be expressed in terms of $q_c(\lambda)$

$$a(\lambda) = 1/\sqrt{1 - \mathrm{sgn}(\beta_2)\big|q_c(\lambda)\big|^2} \exp\left(j\mathcal{H}\left(-\log\sqrt{1 - \mathrm{sgn}(\beta_2)\big|q_c(\lambda)\big|^2}\right)\right), \quad \lambda \in \mathbb{R}. \tag{4.23}$$

The general expression that includes the discrete spectrum can be found in [126, Eq. (1.6.19)] [136, Eq. (6.23)]. It is, however, outside the scope of the thesis. It is also worth noting that if the complex-valued signal $q(t)$ fulfills the condition $\int_{-\infty}^{\infty} |q(t)|dt < \pi/2$, then the NFT of $q(t)$ has no discrete eigenvalues [137, Eq. (4.9)].

We now turn to the definition of NFT in dual-polarisation case.

4.2 Review of dual-polarisation NFT

The DP signal evolution in the optical fibre channel is governed by the Manakov equation as explained in (2.54). The DP-NFT is developed from the integrable Manakov equation [134] written below

$$\frac{\partial \mathbf{A}}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{A}}{\partial t^2} + j\frac{8}{9}\gamma|\mathbf{A}|^2\mathbf{A}. \tag{4.24}$$

It is also convenient to normalise (4.24) with the factors

$$A_n = \sqrt{2\left/\left(\frac{8}{9}\gamma L\right)\right.}, \;\; L_n = \mathrm{sgn}(\beta_2)L, \;\; T_n = \sqrt{(|\beta_2|L)/2}. \tag{4.25}$$

With the normalised variables $\mathbf{q}(\ell,\tau) = \mathbf{A}(z,t)/A_n$, $\ell = z/L_n$ and $\tau = t/T_n$, (4.24) can be reduced to

$$j\frac{\partial \mathbf{q}(\ell,\tau)}{\partial \ell} = \frac{\partial^2 \mathbf{q}(\ell,\tau)}{\partial \tau^2} - 2\,\mathrm{sgn}(\beta_2)\|\mathbf{q}(\ell,\tau)\|^2\mathbf{q}(\ell,\tau) \tag{4.26}$$

where $\mathbf{q}(\ell,\tau) = [q_1(\ell,\tau)\; q_2(\ell,\tau)]$. The ZS system for DP signals is

$$\frac{\partial \mathbf{v}(\tau,\lambda)}{\partial \tau} = \begin{pmatrix} -j\lambda & q_1 & q_2 \\ \mathrm{sgn}(\beta_2)q_1^* & j\lambda & 0 \\ \mathrm{sgn}(\beta_2)q_2^* & 0 & j\lambda \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \;\; \lambda \in \mathbb{C}, \tag{4.27}$$

with the boundary condition

$$\mathbf{v}(\tau,\lambda) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} e^{-j\lambda\tau}, \;\; \tau \to -\infty. \tag{4.28}$$

We can then define the DP non-linear coefficients $a(\lambda)$, $b_1(\lambda)$ and $b_2(\lambda)$ as

$$a(\lambda) = e^{+j\lambda\tau}v_1(\tau,\lambda), \;\; \tau \to +\infty, \tag{4.29}$$

$$b_1(\lambda) = e^{-j\lambda\tau}v_2(\tau,\lambda), \;\; \tau \to +\infty, \tag{4.30}$$

$$b_2(\lambda) = e^{-j\lambda\tau}v_3(\tau,\lambda), \;\; \tau \to +\infty. \tag{4.31}$$

The evolution of the non-linear spectrum and coefficients follows

$$\begin{cases} \lambda_m(\ell) = \lambda_m(0), & \\ b_1(\ell,\lambda_m) = b_1(0,\lambda_m)e^{4j\lambda_m^2\ell}, & \lambda_m \in \Omega, \\ b_2(\ell,\lambda_m) = b_2(0,\lambda_m)e^{4j\lambda_m^2\ell}, & \lambda_m \in \Omega, \\ a'(\ell,\lambda_m) = a'(0,\lambda_m), & \lambda_m \in \Omega, \\ b_1(\ell,\lambda) = b_1(0,\lambda)e^{4j\lambda^2\ell}, & \lambda \in \mathbb{R}, \\ b_2(\ell,\lambda) = b_2(0,\lambda)e^{4j\lambda^2\ell}, & \lambda \in \mathbb{R}, \\ a(\ell,\lambda) = a(0,\lambda), & \lambda \in \mathbb{R}. \end{cases} \tag{4.32}$$

The uni-modularity for DP signals is

$$\left|a(\lambda)\right|^2 - \mathrm{sgn}(\beta_2)\left\|\mathbf{b}(\lambda)\right\|^2 = 1, \quad \text{for } \lambda \in \mathbb{R} \tag{4.33}$$

where $\mathbf{b}(\lambda) = [b_1(\lambda)\ b_2(\lambda)]$. In the absence of the discrete spectrum, $a(\lambda)$ can be similarly expressed in terms of $\mathbf{b}(\lambda)$ or $\mathbf{q_c}(\lambda)$ by simply replacing $\left|b(\lambda)\right|$ with $\left\|\mathbf{b}(\lambda)\right\|$ in (4.22) and $\left|q_c(\lambda)\right|$ with $\left\|\mathbf{q_c}(\lambda)\right\|$ in (4.23)

$$a(\lambda) = \sqrt{1+\mathrm{sgn}(\beta_2)\left\|\mathbf{b}(\lambda)\right\|^2}\exp\left(j\mathcal{H}\left(\log\sqrt{1+\mathrm{sgn}(\beta_2)\left\|\mathbf{b}(\lambda)\right\|^2}\right)\right), \quad \lambda \in \mathbb{R}, \tag{4.34}$$

$$a(\lambda) = 1/\sqrt{1-\mathrm{sgn}(\beta_2)\left\|\mathbf{q_c}(\lambda)\right\|^2}\exp\left(j\mathcal{H}\left(-\log\sqrt{1-\mathrm{sgn}(\beta_2)\left\|\mathbf{q_c}(\lambda)\right\|^2}\right)\right), \quad \lambda \in \mathbb{R}. \tag{4.35}$$

It is also worth mentioning the Parseval identity in the non-linear frequency domain when there is no discrete spectrum [126, Eq. (1.7.8)]

$$\begin{aligned} \int \left\|\mathbf{q}(\tau)\right\|^2 d\tau &= \frac{1}{\pi}\int \log\left(1-\mathrm{sgn}(\beta_2)\left\|\mathbf{q_c}(\lambda)\right\|^2\right)d\lambda \\ &= -\frac{1}{\pi}\int \log\left(1+\mathrm{sgn}(\beta_2)\left\|\mathbf{b}(\lambda)\right\|^2\right)d\lambda. \end{aligned} \tag{4.36}$$

We can use the Parseval identity to verify our NFT and INFT algorithm, and check whether any discrete eigenvalue emerges during noisy wave propagation.

So far, we have clarified the relation between various non-linear coefficients $a(\lambda)$, $\mathbf{b}(\lambda)$ and $\mathbf{q_c}(\lambda)$. One remaining puzzle is to numerically solve the ZS system that relates the non-linear coefficients to the time-domain signal. In the next section, we describe the numerical procedure of solving ZS system that are used in later simulations and experiments.

4.3 Numerical methods for the ZS system

In this section, we shall look closely at the ZS system in SP case and describe the procedure of numerically solving it in the forward (NFT) and backward (INFT) directions. The NFT-INFT in DP case will be a straightforward extension of the SP-NFT-INFT and, thus will simply be listed afterwards. Suppressing the distance dependence, we write down again the ZS system in SP case

$$\frac{\partial \mathbf{v}(\lambda,\tau)}{\partial \tau} = \begin{pmatrix} -j\lambda & q(\tau) \\ \mathrm{sgn}(\beta_2)q^*(\tau) & j\lambda \end{pmatrix} \mathbf{v}, \quad \lambda \in \mathbb{C}. \tag{4.37}$$

NFT is a procedure that calculates the solution of $\mathbf{v}$ at $\tau \to +\infty$ in the ZS system from the signal $q(\tau)$ and a boundary condition of $\mathbf{v}$ at $\tau \to -\infty$. In principle, any numerical method for differential equations should be able to perform NFT, at various degree of accuracy. In contrast, INFT calculates the solution of $\mathbf{v}$ at $\tau \to -\infty$ from the boundary condition at $\tau \to +\infty$. At the same time, $q(\tau)$ is extracted at each discrete time step from $\mathbf{v}(\tau)$. Note that in practice, $\tau \to \pm\infty$ means some large τ where the signal $q(\tau)$ decays to zeros. In this thesis, we only used the continuous spectrum of NFT for information transmission, the numerical method introduced here will only involve the mapping from continuous spectrum to the time domain signal.

The NFT algorithm

Let us construct the signal $q(\tau)$ as the following

$$q(\tau) = \begin{cases} q(\tau), & T_0 \le \tau < T_1, \\ 0, & \text{elsewhere.} \end{cases} \tag{4.38}$$

We discretise the signal $q(\tau)$ into N samples by the time step $\Delta T = (T_1 - T_0)/N$. The discretised signal $q[k] = q(T_0 + k\Delta T), \quad k = 0, 1, 2, ..., N-1$. Similarly, $t_k = T_0 + k\Delta T, \ \ k = 0, 1, 2, ..., N-1$. Using the Ablowitz-Ladik scheme [78, Eq. (17)], the ZS system can be discretised to

$$\mathbf{v}[k+1,\lambda] = c_k \hat{P}[k]\mathbf{v}[k,\lambda] = c_k \begin{pmatrix} s^{1/2} & Q[k] \\ \mathrm{sgn}(\beta_2)Q^*[k] & s^{-1/2} \end{pmatrix} \mathbf{v}[k,\lambda], \tag{4.39}$$

$$\mathbf{v}[0,\lambda] = \begin{pmatrix} 1 \\ 0 \end{pmatrix} s^{T_0/(2\Delta T)}, \ \ 0 \le k \le N-1, \tag{4.40}$$

where $Q[k] = q[k]\Delta T$, $s = e^{-2j\lambda\Delta T}$ and $c_k = 1/\sqrt{|\det \hat{P}[k]|}$. Introducing a change of

variables

$$\hat{\mathbf{v}}[k,\lambda] = \begin{pmatrix} s^{-\frac{k}{2}-\frac{T_0}{2\Delta T}} & 0 \\ 0 & s^{-\frac{k}{2}-\frac{T_0}{2\Delta T}+\frac{1}{2}} \end{pmatrix} \mathbf{v}[k,\lambda], \tag{4.41}$$

we obtain a corresponding ZS system

$$\hat{\mathbf{v}}[k+1,\lambda] = c_k \begin{pmatrix} 1 & Q[k]s^{-1} \\ \mathrm{sgn}(\beta_2)Q^*[k] & s^{-1} \end{pmatrix} \hat{\mathbf{v}}[k,\lambda], \tag{4.42}$$

$$\hat{\mathbf{v}}[0,\lambda] = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ 0 \le k \le N-1. \tag{4.43}$$

Looking closely at (4.42), we observe that if $\mathbf{v}[k,\lambda]$ can be expressed as a polynomial series of s, e.g.,

$$\hat{v}_1[\lambda] = \sum_{m=0}^{m=\infty} \tilde{v}_{1m} s^{-m}, \quad \hat{v}_2[\lambda] = \sum_{m=0}^{m=\infty} \tilde{v}_{2m} s^{-m}, \tag{4.44}$$

the updating step will become a shift of the coefficients $\tilde{v}_{1m},\ \tilde{v}_{2m}$ due to the multiplication of s^{-1}. It turns out that the polynomial coefficients $\tilde{v}_{1m},\ \tilde{v}_{2m}$ are similar to the discrete Fourier coefficient of $\hat{\mathbf{v}}[k,\lambda]$ [138, Eq. (28)]

$$\tilde{v}_{1m} = \frac{1}{\Lambda_1 - \Lambda_0} \int_{\Lambda_0}^{\Lambda_1} \hat{v}_1[\lambda] s^{-m} d\lambda = \frac{1}{\Lambda_1 - \Lambda_0} \int_{\Lambda_0}^{\Lambda_1} \hat{v}_1[\lambda] e^{-j2m\Delta T\lambda} d\lambda, \tag{4.45}$$

$$\tilde{v}_{2m} = \frac{1}{\Lambda_1 - \Lambda_0} \int_{\Lambda_0}^{\Lambda_1} \hat{v}_2[\lambda] s^{-m} d\lambda = \frac{1}{\Lambda_1 - \Lambda_0} \int_{\Lambda_0}^{\Lambda_1} \hat{v}_2[\lambda] e^{-j2m\Delta T\lambda} d\lambda. \tag{4.46}$$

To calculate $\tilde{v}_{1m}$ and $\tilde{v}_{2m}$, we first discretise λ on the interval $[\Lambda_0,\ \Lambda_1]$ with $\Lambda_1 = -\Lambda_0 = \pi/(2\Delta T)$ so that $\lambda[\ell] = \Lambda_0 + \ell\Delta\Lambda,\ \ell = 0, 1, 2, .., M-1$. Let $\tilde{\mathbf{v}}_1$ denote $[\tilde{v}_{1,0},\ \tilde{v}_{11}, ..., \tilde{v}_{1(M-1)}]$ and $\hat{\mathbf{v}}_1$ denote $\left[\hat{v}_1[\cdot,0],\ \hat{v}_1[\cdot,1], ..., \hat{v}_1[\cdot,M-1]\right]$. (4.45) can be implemented as

$$\tilde{\mathbf{v}}_1 = \frac{1}{M}\mathbf{e} \circ \mathrm{DFT}\left(\hat{\mathbf{v}}_1\right), \tag{4.47}$$

where $\circ$ is the element-wise product and

$$\mathbf{e} = \left[1, e^{-j2\pi\frac{\Lambda_0}{\Lambda_1-\Lambda_0}}, ..., e^{-j2\pi\frac{\Lambda_0}{\Lambda_1-\Lambda_0}(M-1)}\right], \tag{4.48}$$

$$= \left[(-1)^0,\ (-1)^1, ..., (-1)^{M-1}\right]. \tag{4.49}$$

Consequently, the iteration (4.42) in the transformed domain is written as

$$\begin{pmatrix} \tilde{v}_{1m}[k+1] \\ \tilde{v}_{2m}[k+1] \end{pmatrix} = c_k \begin{pmatrix} 1 & Q[k]T_m^1 \\ \mathrm{sgn}(\beta_2)Q^*[k] & T_m^1 \end{pmatrix} \begin{pmatrix} \tilde{v}_{1m}[k] \\ \tilde{v}_{2m}[k] \end{pmatrix}, \tag{4.50}$$

where T_m^n is the n-step shift operator acting on the index m, i.e., $T_m^n f[k,m] = f[k,m-n]$, where

$$f[k,m] = \begin{cases} 0, & m < 0 \text{ and } m > M-1, \\ f[k,m], & \text{others.} \end{cases} \tag{4.51}$$

After N iterations, we can calculate the non-linear coefficients according to (4.14) at finite time T_1

$$a(\lambda) = s^{-\frac{N}{2} - \frac{T_0}{2\Delta T}} v_1[N,\lambda] = \hat{v}_1[N,\lambda], \tag{4.52}$$

$$b(\lambda) = s^{\frac{N}{2} + \frac{T_0}{2\Delta T}} v_2[N,\lambda] = s^{N + \frac{T_0}{\Delta T} - \frac{1}{2}} \hat{v}_2. \tag{4.53}$$

The INFT algorithm

In the absence of the discrete spectrum, the INFT maps the continuous spectrum and its non-linear coefficients to a time-domain signal. Formally, INFT : $q_c(\lambda),\ \lambda \in \mathbb{R} \mapsto q(\tau)$ or INFT : $b(\lambda),\ \lambda \in \mathbb{R} \mapsto q(\tau)$. The INFT algorithm requires we solve the ZS system backwards from $+\infty$ to $-\infty$ and at each iteration step, we extract $Q[k]$ from the eigenvectors $[\tilde{\mathbf{v}}_1\ \tilde{\mathbf{v}}_2]^T$.

Given the non-linear spectrum and coefficients $q_c(\lambda),\ \lambda \in \mathbb{R}$ or $b(\lambda),\ \lambda \in \mathbb{R}$, we can obtain the discretised eigenvectors $[\tilde{\mathbf{v}}_1\ \tilde{\mathbf{v}}_2]^T$ via (4.23) or (4.22), (4.52) and (4.47). The backward iteration can be derived from (4.50)

$$\begin{pmatrix} \tilde{v}_{1m}[k] \\ \tilde{v}_{2m}[k] \end{pmatrix} = c_k \begin{pmatrix} 1 & -Q[k] \\ -\mathrm{sgn}(\beta_2)Q^*[k]T_m^{-1} & T_m^{-1} \end{pmatrix} \begin{pmatrix} \tilde{v}_{1m}[k+1] \\ \tilde{v}_{2m}[k+1] \end{pmatrix}. \tag{4.54}$$

At each iteration, we can recover the signal $Q[k]$ according to [138, Eq. (34)]

$$Q^*[k] = \mathrm{sgn}(\beta_2)\tilde{v}_{20}[k+1]. \tag{4.55}$$

The above relation can be seen by induction from (4.42)

$$\begin{aligned}
&\begin{pmatrix}\hat{v}_1[0,\lambda]\\ \hat{v}_2[0,\lambda]\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}, \quad \begin{pmatrix}\hat{v}_1[1,\lambda]\\ \hat{v}_2[1,\lambda]\end{pmatrix} = c_0 \begin{pmatrix}1\\ \mathrm{sgn}(\beta_2)Q_0^*\end{pmatrix},\\
&\begin{pmatrix}\hat{v}_1[2,\lambda]\\ \hat{v}_2[2,\lambda]\end{pmatrix} = c_0c_1 \begin{pmatrix}1+\mathrm{sgn}(\beta_2)Q_0^*Q_1 s^{-1}\\ \mathrm{sgn}(\beta_2)Q_1^* + \mathrm{sgn}(\beta_2)Q_0^* s^{-1}\end{pmatrix},\\
&\begin{pmatrix}\hat{v}_1[3,\lambda]\\ \hat{v}_2[3,\lambda]\end{pmatrix} = c_0c_1c_2 \begin{pmatrix}1+\mathrm{sgn}(\beta_2)(Q_0^*Q_1+Q_1^*Q_2)s^{-1}+\mathrm{sgn}(\beta_2)Q_0^*Q_2 s^{-2}\\ \mathrm{sgn}(\beta_2)Q_2^* + \left(Q_0^*Q_1Q_2^* + \mathrm{sgn}(\beta_2)Q_1^*\right)s^{-1} + \mathrm{sgn}(\beta_2)Q_0^* s^{-2}\end{pmatrix},\\
&\quad\vdots
\end{aligned} \tag{4.56}$$

where Q_k is the short note for $Q[k]$. Recall that $\tilde{v}_{1m}$ is just the coefficient of the s^{-m} term in the polynomial expression of $\hat{v}_1$, we can obtain (4.55).

DP-(I)NFT

The numerical methods for DP-(I)NFT are straightforward extensions of the SP ones. To guarantee reproducibility, we write down the corresponding ZS system

$$\hat{\mathbf{v}}[k+1,\lambda] = c_k \begin{pmatrix}1 & Q_1[k]s^{-1} & Q_2[k]s^{-1}\\ \mathrm{sgn}(\beta_2)Q_1^*[k] & s^{-1} & 0\\ \mathrm{sgn}(\beta_2)Q_2^*[k] & 0 & s^{-1}\end{pmatrix}\hat{\mathbf{v}}[k,\lambda], \tag{4.57}$$

$$\hat{\mathbf{v}}[0,\lambda] = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad 0 \le k \le N-1. \tag{4.58}$$

If one follows the same steps of the SP-(I)NFT algorithm, it should be easy enough to implement the DP-(I)NFT algorithm.

4.4 Numerical examples of INFT-NFT algorithms

In this section, some toy examples are demonstrated in the single polarisation case to provide a preliminary validation of the aforementioned algorithms with a focus on the continuous spectrum. Let $q(\tau,\ell)$ be the fundamental soliton as introduced in [67, Eq. 5.2.15]

$$q(\tau,\ell) = A\,\mathrm{sech}(A\tau)\exp\left(jA^2\ell/2\right). \tag{4.59}$$

In the case of $\mathrm{sgn}(\beta_2) = -1$, the continuous spectrum of the fundamental soliton calculated by the NFT algorithm should be negligible, while in the case of $\mathrm{sgn}(\beta_2) = 1$ and finite-extent signals, the consecutive operation of NFT and INFT algorithm should be able to recover the time-domain pulse with reasonable accuracy, i.e.,

INFT{NFT{$q(\tau)$}}$\approx q(\tau)$. These are shown to be true in Fig. 4.1 with $A = 2$ and $\ell = 0$. The energy of the fundamental soliton with $A = 2$ is $\int_{-\infty}^{\infty} |q(\tau)|^2 d\tau = 4$. The energy of the corresponding continuous spectrum calculated by the NFT algorithm in the case of $\text{sgn}(\beta_2) = -1$ (Fig. 4.1(b)) is $1/\pi \int_{-\infty}^{\infty} \log\big(1 + |q_c(\lambda)|^2\big) d\lambda \approx 2 \times 10^{-5}$, which is fairly close to zero when compared with the total pulse energy, and can be regarded as being purely from numerical errors. In the case of $\text{sgn}(\beta_2) = 1$, the discrepancy of the original and recovered fundamental solitons shown in Fig. 4.1(c) is $\int_{-\infty}^{\infty} |q(\tau) - q_r(\tau)|^2 d\tau / \int_{-\infty}^{\infty} |q(\tau)|^2 d\tau \approx 4 \times 10^{-7}$ where $q_r(\tau)$ denotes the recovered pulse. The energy of the corresponding continuous spectrum (Fig. 4.1(d)) calculated by the NFT algorithm in the case of $\text{sgn}(\beta_2) = 1$ is 4.01.

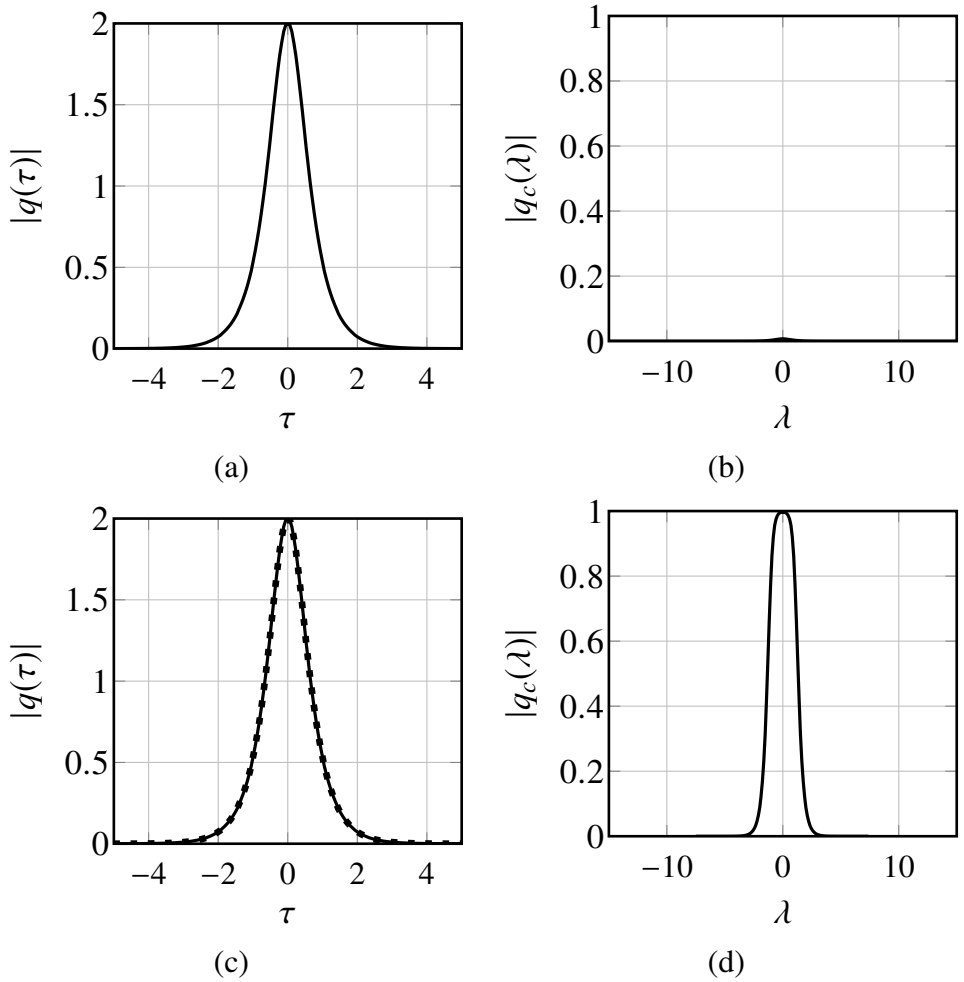

Figure 4.1: (a) The fundamental soliton $q(\tau) = 2\,\text{sech}(2\tau)$. (b) The continuous spectrum of the fundamental soliton calculated by the NFT algorithm in the case of $\text{sgn}(\beta_2) = -1$. (c) The recovered (dotted) and original (solid) fundamental soliton in the case of $\text{sgn}(\beta_2) = 1$. (d) The continuous spectrum of the fundamental soliton calculated by the NFT algorithm in the case of $\text{sgn}(\beta_2) = 1$.

4.5 Approximating the integrable model

One major criticism about the NFT-based system is that it is based on the integrable (lossless) NLSE in (4.1), while most of the optical fibre links are lossy and amplified by EDFAs. This deviation will introduce a systematic error into any NFT-based system operating in EDFA-amplified links. In this section, the integrable path-average (PA) model with a modified non-linear coefficient γ_a is introduced. The field solution of such a model approximates the field solution in lossy fibre links that has an actual non-linear coefficient γ. The technique is commonly used in the soliton communication systems in optical fibre [139].

Consider an optical fibre link with periodic lumped amplification. SP signal propagation in one span is described by the NLSE in (2.13)

$$\frac{\partial A}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} - \frac{\alpha}{2}A + j\gamma|A|^2 A. \tag{4.60}$$

Introducing a change of variable

$$Q(t,z) = e^{\frac{\alpha}{2}z}A(t,z), \tag{4.61}$$

(4.60) is transformed to

$$\frac{\partial Q}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 Q}{\partial t^2} + j\gamma e^{-\alpha z}|Q|^2 Q, \tag{4.62}$$

By approximating $\gamma e^{-\alpha z}$ with its average γ_a in the interval $0 < z < L_{\text{sp}}$, we obtain the integrable PA model

$$\frac{\partial A_{\text{pa}}}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A_{\text{pa}}}{\partial t^2} + j\gamma_a|A_{\text{pa}}|^2 A_{\text{pa}}, \tag{4.63}$$

where

$$\gamma_a = \frac{\gamma}{\alpha L_{\text{pa}}}(1 - e^{-\alpha L_{\text{sp}}}). \tag{4.64}$$

If $A_{\text{pa}}(t, z = 0) = A(t, z = 0)$, then $A_{\text{pa}}(t,z)$ approximates $A(t,z)$ with small error after each EDFA. The normalisation factors will be changed to

$$A_n = \sqrt{2/(\gamma_a L)}, \;\; L_n = \text{sgn}(\beta_2)L, \;\; T_n = \sqrt{(|\beta_2|L)/2}. \tag{4.65}$$

The same principle also applies to the Manakov equation in (2.54), we state its PA

version without derivations

$$\frac{\partial \mathbf{A}_{\mathrm{pa}}}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 \mathbf{A}_{\mathrm{pa}}}{\partial t^2} + j\frac{8}{9}\gamma_a |\mathbf{A}_{\mathrm{pa}}|^2 \mathbf{A}_{\mathrm{pa}}. \tag{4.66}$$

The normalisation factors become

$$A_n = \sqrt{2\Big/\left(\frac{8}{9}\gamma_a L\right)}, \quad L_n = \mathrm{sgn}(\beta_2)L, \quad T_n = \sqrt{(|\beta_2|L)/2}. \tag{4.67}$$

4.6 Summary

In this chapter, NFT and INFT in both SP and DP cases for integrable evolution systems have been introduced. The numerical methods to calculate all of them were also introduced in detail. Furthermore, to adapt the NFT framework to the non-integrable realistic optical fibre system, the path-average model that approximates the integrable evolution model was also considered.

In Chapter 5 and 6, these numerical methods will be used to simulate NFT-based transmission schemes.

5

Per-sample capacity estimation of a NFDM optical network

In the previous chapter, the building blocks of an NFT-based transmission system, i.e., NFT-INFT pairs and PA model, were introduced in detail. With all these tools, we are now in the position to review NFDM-related work and simulate an NFDM-based optical fibre network.

5.1 Related work

The non-linear Fourier spectrum consists of a continuous component and a discrete component. Information transmission using the continuous spectrum is studied in [79,140–146], while discrete spectrum modulation is studied in [79,147–149]. Recent experimental demonstrations of data transmission based on the NFT include joint discrete and continuous spectrum modulation [143]. A record data rate of 32 Gbit/s was demonstrated in [150] using the continuous spectrum of NFT and the 32-QAM modulation format. A peak-SNR gain of 1.3 dB was achieved over a comparable orthogonal frequency-division multiplexing (OFDM) system; for more comparison, see also [151–157].

The aforementioned work have mostly applied the NFT to point-to-point links, often as a non-linear compensation scheme. However, the main potential of NFDM can only be realised in network environments. Here, users' signals would be mul-

tiplexed in the non-linear Fourier spectrum in disjoint intervals and, in the absence of noise, propagate independently in the network. Crucially, the signals of the UOI will not be distorted by the co-propagating signals. As a result, the deterministic inter-symbol and inter-user interference are simultaneously zero for all users of a network.

Our research in NFDM initially demonstrated proof-of-concepts, showing how this scheme works in point-to-point channels. However, to date, the AIRs of the NFDM signals were at best the same as the AIRs of the WDM signals in these initial demonstrations [79]. Advances in numerical methods have made it possible to multiplex signals in the non-linear Fourier domain and explore the NFT at high powers. The AIRs of NFDM and WDM were compared for the first time recently in [115] for fibre in the defocusing regime ($\mathrm{sgn}(\beta_2) = 1$). It was shown that NFDM achieves data rates higher than WDM rates, subject to the same power and bandwidth constraints, in illustrative simulations of 5 users and 1 symbol per channel use.

The goal of this chapter is twofold. First, we summarise the AIRs of NFDM, compared to WDM, for a given input power and bandwidth in an ideal integrable model in the focusing regime ($\mathrm{sgn}(\beta_2) = -1$). For this model, it is shown that the NFDM AIR is greater than the WDM AIR in a representative system with five users and one symbol per channel use. The first part is a summary of the recent conference work on NFDM [40]. Second, we study the impact of the following perturbations on NFDM: fibre loss, periodic amplification, third-order dispersion and polarisation effects. One of the main limitations of NFDM is that it is based on ideal integrable models of the optical fibre, for example, with ideal distributed Raman amplification. Therefore, it is important to see how NFDM performs in realistic systems with non-idealities and perturbations. It is shown that using a path-averaged loss model, the impact of the attenuation (with periodic EDFA amplification) on NFDM is noticeable. It is also shown that the WDM AIR with joint dual-polarisation back-propagation and perturbations compensation is higher than the NFDM AIR with two independent single-polarisation demodulations and without perturbations compensation, in a representative system with five users and one symbol per channel use. We also study the dependency of the AIRs with the number of symbols. As the number of symbols is increased, the AIRs of both WDM and NFDM are decreased. The results in this chapter clarify to what extent NFDM applies to realistic systems.

The chapter is organised as follows. In Sec. 5.2, we review the optical fibre channel model and the network environment. Sec. 5.3 briefly summarise the NFDM theory. The achievable rates of NFDM and WDM are computed and compared in Sec. 5.4, in a multi-user system with five users and one symbol per channel use.

In Sec. 5.5, NFDM is studied for non-integrable models with non-idealities and perturbations. After comments on the complexity of the NFT in Sec. 5.6, the chapter is concluded in Sec. 5.7.

5.2 Optical fibre networks

Signal propagation in single-mode single-polarisation optical fibre with ideal distributed Raman amplification can be modelled by the stochastic NLSE. For convenience, the equation in the normalized form is rewritten here

$$j\frac{\partial q}{\partial z} = \frac{\partial^2 q}{\partial t^2} - 2\,\mathrm{sgn}(\beta_2)|q|^2 q + n(t,z), \tag{5.1}$$

where $q(t,z)$ is the complex envelope of the signal as a function of time t and distance z along the fibre, $n(t,z)$ is white GN with PSD σ_0^2. The signal and noise are band-limited to the same bandwidth B for all $0 \le z \le 1$. The reader is referred to Sec. 4.1 for further details about the model, such as the normalization procedure.

The research described in this chapter considers a *network environment*. This is an optical network with the following set of assumptions: (a) there are multiple transceiver pairs, (b) there are add-drop multiplexers (ADMs) in the network; the locations and the number of ADMs are unknown, (c) each transceiver pair does not have information of the incoming and outgoing signals in the path that connects them. In a network environment, the signal of the UOI co-propagates with the signals of the other transceiver pairs in the network and, in the case of WDM, is subject to inter-user interference [79, Sec. II. B.3.]. A typical network environment is depicted in Fig. 1.7(a). In simulations, the network environment is emulated by multi-channel point-to-point transmissions with the Rx only having access to the signal in the channel of interest.

5.3 Review of NFDM

We briefly review basic NFDM theory from [40]. The theory in this section applies to both focusing ($\mathrm{sgn}(\beta_2) = -1$) and defocusing ($\mathrm{sgn}(\beta_2) = 1$) regimes. The simulations in the subsequent sections are presented only for the focusing regime.

Let $q_c(\lambda, z) : \mathbb{R} \times \mathbb{R}^+ \mapsto \mathbb{D}$ be the non-linear Fourier transform of $q(t,z)$ as a function of the non-linear frequency λ and distance z, where, in the focusing regime, $\mathbb{D}$ is the complex plane $\mathbb{C}$ and, in the defocusing regime, is the unit disk $\mathbb{T} = \{q_c \in \mathbb{C} : |q_c| < 1\}$.

In the defocusing regime $|q_c(\lambda, z)| < 1$. We thus introduce the following trans-

formation to map functions with co-domain $\mathbb{D}$ to functions with co-domain $\mathbb{C}$:

$$U(\lambda, z) = \Big(-\operatorname{sgn}(\beta_2) \log\Big(1 - \operatorname{sgn}(\beta_2)|q_c(\lambda, z)|^2\Big)\Big)^{\frac{1}{2}} e^{j \angle q_c(\lambda, z)}, \tag{5.2}$$

where $\angle q_c$ is the phase of q_c. This eliminates the unit peak power constraint in the defocusing regime. We use the transform (5.2) in both focusing and defocusing regimes. In the defocusing regime, the transform is necessary so that the signal space is a vector space. In the focusing regime it is not necessary, but it makes signal energies in the time and non-linear Fourier domain the same, which is useful.

We define a *generalized time* τ that is related via the Fourier transform to the non-linear frequency $\lambda/2\pi$. Let $u(\tau, z) \leftrightarrow U(\lambda, z)$ be a Fourier transform pair, i.e.,

$$u(\tau, z) = \sqrt{2}\mathcal{F}^{-1}\{U(\lambda, z)\}. \tag{5.3}$$

The factor $\sqrt{2}$ is introduced so that the energy (computed using (4.36)) of $u(\tau, z)$, $U(\lambda, z)$, $q_c(\lambda, z)$ and $q(t, z)$ are the same; see also [138, Eq. 28].

5.3.1 NFDM transmitter

Consider a multi-user system with N_u users, each having N_s symbols, with total linear bandwidth B and total average power $\mathcal{P}$. Modulation in NFDM begins by forming

$$u(\tau, 0) = \sqrt{2} \sum_{k=-\frac{N_u}{2}}^{\frac{N_u}{2}-1} \sum_{\ell=-\frac{N_s}{2}}^{\frac{N_s}{2}-1} s_\ell^k \phi(\tau - \ell T_0) e^{j2\pi k W_0 \tau}, \tag{5.4}$$

where $\phi(\tau)$ is chosen to be a RRC function in this work with unit energy, bandwidth W_0 and the roll-off factor r, $T_0 = 1/W_0$, $\{s_\ell^k\}_\ell$ are the symbols of user k at time instance ℓ chosen from a multi-ring constellation Ξ; see Fig. 2.14.

Next, $U(\lambda, 0)$ is computed according to (5.3) and subsequently $q_c(\lambda, 0)$ is obtained as

$$q_c(\lambda, 0) = \Big(\operatorname{sgn}(\beta_2) - \operatorname{sgn}(\beta_2) e^{-\operatorname{sgn}(\beta_2)|U(\lambda,0)|^2}\Big)^{\frac{1}{2}} e^{j \angle U(\lambda, 0)}.$$

Finally, we have the signal for transmission

$$q(t, 0) = \mathrm{INFT}\{q_c(\lambda, 0)\}.$$

5.3.2 Reducing the peak-to-average-power-ratio (PAPR)

The NFT signals often have high peak-to-average-power-ratio (PAPR) in the time domain, requiring a large number of samples to accurately represent. We use a simple pre-equalisation to reduce the PAPR through expanding the signal over a wider support.

The idea is similar to the split DBP in [158]. Recall that the NFT spectrum evolves according to the all-pass-like filter (5.5). We multiply the NFT spectrum by $e^{4j\lambda^2 l}$ at the transmitter for a suitable l, to broaden the signal in the time domain. A similar idea is proposed by Tavakkolnia et al. [10]. The two approaches are related and are compared schematically in Fig. 5.1. In the simulations presented in this work, we used our proposed PAPR reduction approach shown in Fig. 5.1, with very small l.

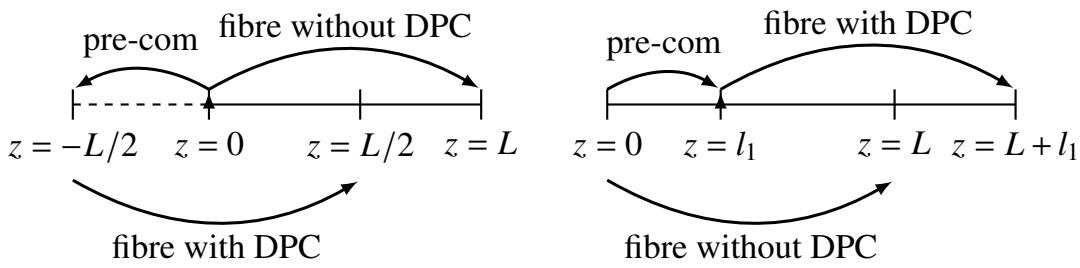

Figure 5.1: Pre-equalisation methods used (left) in [10] and (right) in Sec. 5.4.

5.3.3 Channel filter

An important property of the NFT, crucial to the communication problem, is that, in the absence of noise, $q_c(\lambda, z)$ propagates in distance according to an all-pass-like filter

$$q_c(\lambda, z) = e^{4j\,\mathrm{sgn}(\beta_2)\lambda^2 z} q_c(\lambda, 0). \tag{5.5}$$

It can be seen that signal propagation in the non-linear Fourier domain is governed by simple phase shift. Importantly, the propagation of different non-linear frequency components is independent of one another, which is the reason that multiplexing in the non-linear Fourier domain is of interest. The input-output relation (5.5) is similar to the linear dispersion relation in the frequency domain for the NLSE without the non-linear term.

5.3.4 NFDM receiver

At the receiver, forward NFT is applied to obtain $q_c(\lambda, L) = \mathrm{NFT}\{q(t, L)\}$. Then the channel filter (5.5) is applied to calculated the transmitted non-linear spectrum. Subsequently, $U(\lambda, L)$ and $u(\tau, L)$ can be computed using (5.2) and (5.3). The received symbols are then obtained by matched filtering

$$\hat{s}_\ell^k = \int_\infty^{-\infty} u(\tau, L)\phi^*(\tau - \ell T_0)e^{-j2\pi k W_0 \tau} d\tau.$$

Table 5.1: Fibre and system parameters

n_{sp}	1.1	spontaneous emission factor
h	$6.626\times10^{-34} J\cdot s$	Planck's constant
ν	193.44 THz	center carrier frequency
α	0.046 km^{-1}	fibre loss
γ	1.27 $(\text{W.km})^{-1}$	non-linearity parameter
β_2	-21.5×10^{-27} s^2/m	dispersion coefficient
L	2000 km	transmission distance
W	100 GHz	total linear bandwidth
R_o	3	oversampling rate
W_u	20 GHz	per-user linear bandwidth
N_u	5	number of users
N	2^{14}	number of samples per frame
r	0.25	roll-off factor of the raised-cosine

5.4 Comparison of the achievable rates of WDM and NFDM

In this section, we compare the achievable rates of NFDM and WDM under the same bandwidth and power constraints in the focusing regime of optical fibre channel with ideal distributed Raman amplifier. The material in this section is a summary of the conference work [40].

5.4.1 AIRs of WDM and NFDM

First, a simple noiseless multi-user simulation for one signal without noise is presented to show the origin of the NFDM gain. We consider a 5-user NFDM and 5-user WDM system with the same overall linear bandwidth of 100 GHz and total signal power $P = E/T$, where E and T are the energy and time duration of the (entire) multiplexed signal $q(t,0)$. Time duration and bandwidth are defined as intervals that contain 99% of the signal energy. RRC pulses with roll off factors of 25% and standard single-mode fibre with ideal distributed Raman amplification are considered. In each channel use, only one symbol per user was transmitted. The fibre parameters in simulation are shown in Tab. 5.1. With this set of parameters, the average symbol duration is around 6.8 ns. The data rate in bit/s can be obtained by simply deviding the AIRs in bit/symbol with the average symbol duration.

Fig. 5.2(a) shows a WDM signal at the transmitter (Tx) and Rx with total power of 8 dBm (at Tx) in the absence of noise. The WDM receiver first filters the UOI and then applies DBP to reverse the intra-channel interactions. The relative error in the frequency domain is defined as $e = \|\tilde{q}(f,z)) - \tilde{q}(f,0)\|/\|\tilde{q}(f,0)\|$, where $\tilde{q}(f,t) = \mathcal{F}(q(t,z))$ after equalisation. The mismatch error for the UOI in WDM is

95%. The resulting mismatch comes from the non-linear interactions that cannot be compensated for in network environments using linear multiplexing. In contrast, NFDM users do not interfere with one another as shown in Fig. 5.2(b). The NFDM signals can be recovered almost perfectly with a relative error $e = 10^{-3} - 10^{-9}$. The Fourier spectra of signals in both systems are shown in Fig. 5.2(c). It can be seen that the amount of the spectral broadening in WDM and NFDM is similar.

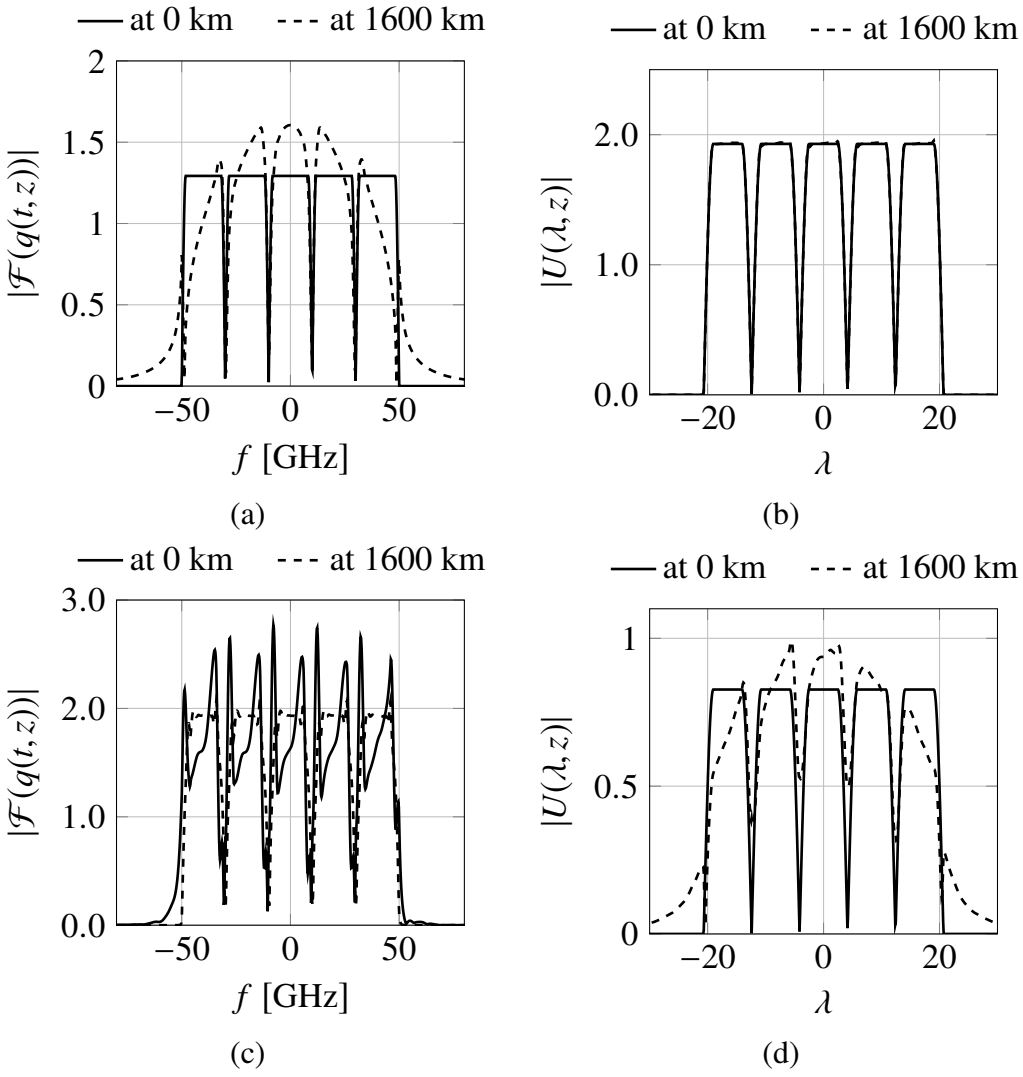

Figure 5.2: For the integrable model: (a) interference in WDM; (b) lack of interference in NFDM; (c) Fourier spectra at distances with maximum input-output bandwidths. For the lossy model with periodic amplification: (d) interference in NFDM. Noise is set to zero in these figures.

We now turn to the main simulations that estimate AIRs of UOI in both WDM and NFDM. The average mutual information between the input output matrix of

symbols (using WDM or NFDM) is

$$\frac{1}{N_s} I(\{s_\ell^0\}; \{\hat{s}_\ell^0\}) \tag{5.6}$$

where $\{s_\ell^0\}$ and $\{\hat{s}_\ell^0\}$ are the transmitted and received symbols at all time indices for user 0. For channels with memory, i.e., s_ℓ is not statistically independent of $s_{\ell-k}$ where $k = 1, 2, 3\ldots$, the evaluation of (5.6) is not computationally feasible. We chose to evaluate another quantity

$$I(s_0^0; \hat{s}_0^0) \tag{5.7}$$

that to large extent closely follows the quantity $\frac{1}{N_s} I(\{s_\ell^0\}; \{\hat{s}_\ell^0\})$. The term $I(s_0^0; \hat{s}_0^0)$ is the per-sample AIR that could be interpreted as the MI of an optical fibre channel that is only affected by additive noise and inter-channel interference. The following simulations aim to gather statistics to estimate $I(s_0^0; \hat{s}_0^0)$.

In simulation, transmitted symbols s_ℓ^k are chosen from the points on the positive real axis of a uniformly-spaced multi-ring constellation Ξ, introduced in Fig. 2.14, in the U domain (leading to a geometrically-spaced constellation in the q_c domain). The constellation Ξ consists of at most 64 rings and 128 phase points on each ring (13 bit per symbol). The reduction of signal space from an entire ring constellation to one of its radii is a common technique to reduce the simulation time by utilising the rotational invariance (Sec. 2.4.2) of both the channel and Ξ.

Noise is introduced in a distributed manner along the fibre. The noise bandwidth is set to be the maximum of the input-output signal bandwidth corresponding to the highest energy signal. The PSD σ_0^2 of the noise arising from distributed Raman amplification is the constant $n_{sp} h_p \nu \alpha$ for all z and in-band frequencies, with parameters in Tab. 5.1.

The number of samples per channel use in time, linear and non-linear spectrum is 2^{14}. The conditional probability density function (PDF) $p(\hat{s}_0^0 | s_0^0)$ of the central channel $s_0^0 \mapsto \hat{s}_0^0$ is estimated using histogram (Sec. 2.4.2) based on 4500 noise realisations. The mutual information is then maximized using the Arimoto-Blahut algorithm (Al. 2).

The AIRs are plotted in Fig. 5.3 for the focusing regime. It can be seen that the AIR of WDM reaches its maximum of 10 bit/symbol at a power of -4.5 dBm. In contrast, the AIR of NFDM reaches a higher maximum of around 11.5 bit/symbol at 1 dBm. The saturation of the NFDM AIR is attributed to neglecting the stochastic memory and the numerical error in the NFT algorithm at high powers, and the potentially non-optimal constellation; see Sec. 5.4.2 for the explanation.

The received constellations of WDM and NFDM are compared in Fig. 5.4.

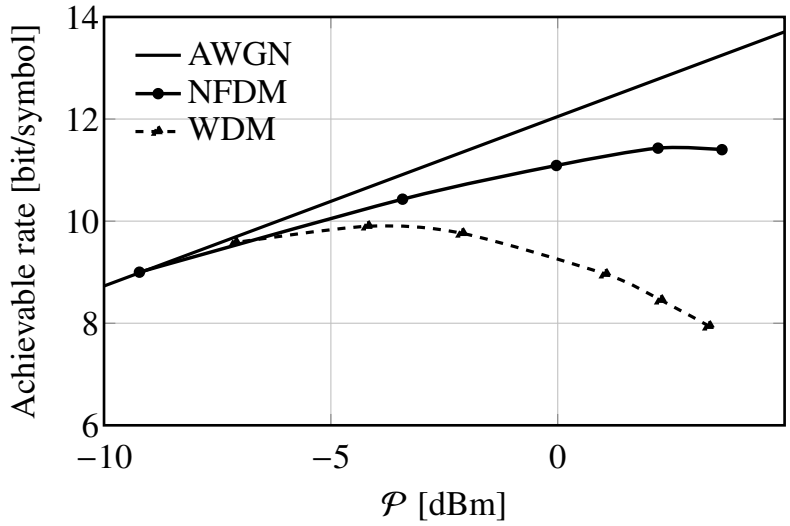

Figure 5.3: WDM and NFDM AIRs in the integrable model and focusing regime.

On one hand, it can be seen that the WDM constellation suffers from significant phase distortion. There is also a phase rotation, due to cross-phase modulation, whose average is approximately $\gamma L\mathcal{P}$ [5, Sec. X.C]. On the other hand, the NFDM received constellation is similar to that in an AWGN channel [159]: the noise variances (i.e., noise 'clouds') are small and approximately equal. The NFDM scheme is theoretically limited only by signal-noise interactions.

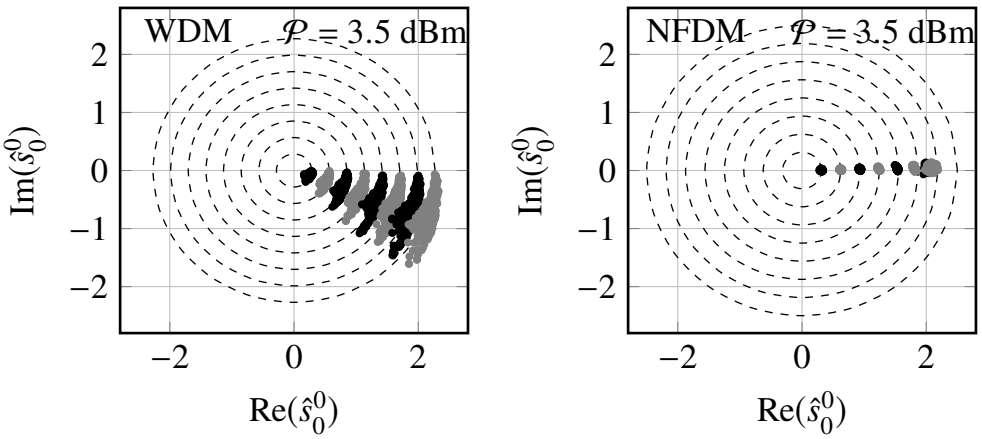

Figure 5.4: Constellation in (left) WDM and (right) NFDM in the focusing regime.

The AIRs of NFDM and WDM in the defocusing regime [138, Fig. 6] are qualitatively similar to the AIRs in the focusing regime.

We close this subsection by emphasizing that, the result that the NFDM-AIR is more than WDM-AIR in Fig. 5.3 is obtained for one symbol per channel use, namely $N_s = 1$, and for an integrable model with single polarisation. The AIRs

change with N_s, as shown later in Sec. 5.5.3 for NFDM and WDM with $N_s = 15$. As a result, the above conclusion remains to be verified for parameters not considered here, particularly when $N_s > 15$. More NFDM system simulations with $N_s > 15$ will be described in Chap. 6.

5.4.2 Stochastic memory

In the aforementioned simulations, only one symbol was transmitted in each channel use in each user. On the Rx side, symbol energy could be dispersed into the adjacent symbols, leading to a loss of information rate. Therefore, we consider the single-input multi-output channel $s_0^0 \mapsto \{\hat{s}_\ell^0\}_\ell$, where $\{\hat{s}_\ell^0\}_\ell$ denotes output symbols for the UOI after equalisation. Using the chain rule for mutual information, we find

$$I\big(s_0^0; \{\hat{s}_\ell^0\}_\ell\big) = I\big(s_0^0; \hat{s}_0^0\big) + \underbrace{I\big(s_0^0; \{\hat{s}_\ell^0\}_{\ell\neq 0} \big| \hat{s}_0^0\big)}_{\text{stochastic memory}}. \tag{5.8}$$

The stochastic memory refers to a component of the memory (signal-noise interactions) that is a function of noise; namely, it vanishes when noise is set to zero. This is the second term in (5.8) and is determined by signal-noise interactions. The stochastic memory is signal-dependent, growing with the input power, and not equalised by digital signal processing at the receiver in this thesis. A recnetly proposed neural-network-based technique [160] has shown some promising results in combating the stochastic memory in the integrable optical fibre channel. It would be interesting to see the technique tested in more realistic settings.

The stochastic memory causes the symbol energy to flow from s_0^0 to $\hat{s}_\ell^0$, $\ell \neq 0$. As $|s_0^0|$ is increased, the stochastic memory grows because the energy in $\{\hat{s}_\ell^0\}_{\ell\neq 0}$ is increased, leading to a smaller $|\hat{s}_0^0|$; see the outer ring in Fig. 5.4 (right) and Fig. 5.5 (right). This amounts to a reduced SNR at the receiver for the channel $s_0^0 \mapsto \hat{s}_0^0$. Since only the first term in (5.8) is used for computing the AIR in this work, a loss of SNR translates to a loss of AIR. Therefore, the stochastic memory, if not accounted for, incurs a penalty on the NFDM AIR at high powers.

The stochastic memory was examined in detail in a simulation with 5 users and (total linear) bandwidth of roughly 100 GHz. We compare the extent of the stochastic memory in the fibre channel in the focusing regime with that in an AWGN channel whose total noise power equals the total noise power in the fibre channel. It is shown in Fig. 5.5 that the received symbols $|\hat{s}_0^0|$ in NFDM cluster in a cloud with a mean smaller than the transmit symbol $|s_0^0|$. This effect is the SNR (or energy) loss due to the signal-noise interaction. The signal-noise interaction in the focusing regime appears to be stronger than that in the defocusing regime.

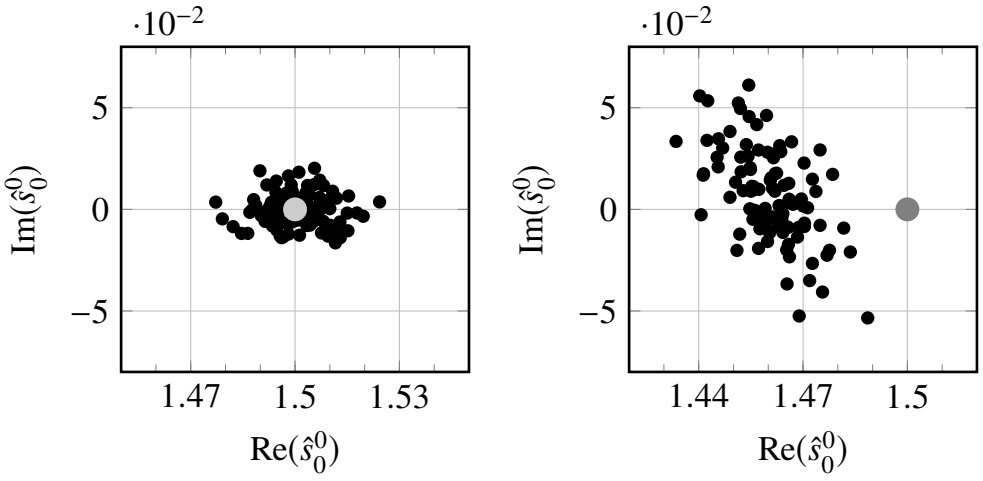

Figure 5.5: Transmitted and received (grey and black markers) symbols in (left) AWGN and (right) NFDM, showing the SNR loss due to the stochastic memory.

The stochastic memory can be addressed by considering the single-input multiple-output channel $s_0^0 \mapsto \{\hat{s}_\ell^0\}_\ell$, instead of the single-input single-output channel $s_0^0 \mapsto \hat{s}_0^0$. Such a stochastic mapping is recently modelled in [161] for ideal distributed amplification and high SNR regime. Developing a maximum likelihood sequence detection based on this model at the Rx can potentially stop, or at least delay, the rolling off of the AIR of the NFDM in Fig. 5.3 at the maximum power $\mathcal{P} = 2.7$ dBm.

Note that the AIRs of different DoFs are not the same. Different users suffer from the stochastic memory to different extent. For example, the central user in WDM suffers more from non-linear impairments.

5.5 Impact of perturbations and non-idealities

Research on NFDM has so far presented proof-of-concepts, demonstrating NFDM in ideal or simplified models. There remain many non-idealities and practical constraints, some of which are studied in this section. Specifically, we study the impacts of the loss, periodic amplification, third-order dispersion, and polarisation-mode dispersion (PMD).

5.5.1 Loss with periodic amplification

In the framework of NFT, it is often assumed that the fibre loss can be perfectly compensated using distributed Raman amplification. However, many optical systems are based on lumped amplification using EDFAs. For EDFA-amplified channels, signal propagation is governed by the NLSE explained in (4.60). Lumped amplification is performed at the end of each span of length $L_{\mathrm{sp}} = L/N_{\mathrm{sp}}$, where N_{sp}

is the number of spans. After this, ASE noise is added to the signal, which is usually modelled as a white complex circular symmetric Gaussian stochastic process, with the well-known PSD for EDFA; see [162, Sec. 3.1.3].

To put it in the framework of NFT, we use the PA model described in (4.63). The equation is rewritten here for convenience

$$\frac{\partial A_{\mathrm{pa}}}{\partial z} = -j\frac{\beta_2}{2}\frac{\partial^2 A_{\mathrm{pa}}}{\partial t^2} + j\gamma_a |A_{\mathrm{pa}}|^2 A_{\mathrm{pa}}, \tag{5.9}$$

$$\gamma_a = \frac{\gamma}{\alpha L_{\mathrm{sp}}}(1 - e^{-\alpha L_{\mathrm{sp}}}). \tag{5.10}$$

Using parameters in Tab. 5.1 and 5.2, we have $\gamma_a = 0.3364\ \mathrm{W}^{-1}\cdot\mathrm{km}^{-1}$. Clearly, (5.9) can be normalized according to the change of variable with the scale factor $A_n = \sqrt{2/(\gamma_a L)} = 37.16\times 10^{-4}$ W. The NFT-signals are generated using the PA-model and transmitted in the EDFA-amplified link.

Table 5.2: Parameters of the periodic amplification model

L_{sp}	80 km	span length
N_{sp}	20	number of spans
N_F	4.5 dB	noise figure of EDFAs
L	1600 km	total length

The error between the exact model and the path-averaged model grows linearly with distance and power [163]. It is not clear how an NFDM system designed using the averaged model works in realistic EDFA-based systems represented by the original model. For example, although there is no interference in the ideal model in Fig. 5.2 (b), loss introduces interference as illustrated in Fig. 5.2 (d). To examine the AIRs, an optical link including loss and periodic amplification was simulated using the parameters found in Tab. 5.2 and the same method described in Sec. 5.4.1. The WDM and NFDM received constellations with 4000 noise realisations are displayed in Fig. 5.6. It is observed that NFDM is subject to a more phase distortion compared with NFDM in the ideal integrable model (compare Fig. 5.6(a) with Fig. 5.4 (b)). Nevertheless, the sizes of the noise clouds in NFDM are still smaller than those in WDM. This improvement then translates to a higher AIR for NFDM.

The AIRs of NFDM and WDM in the focusing regime with loss and periodic amplification are calculated with Arimoto-Blahut algorithm (Al. 2) and shown in Fig. 5.7. Although the AIR of NFDM also rolls off at high powers (potentially because the path-averaged model breaks down at high power), it reaches a higher maximum than the maximum WDM AIR. We conclude that NFDM offers a marginal peak-AIR gain of 0.41 bit/symbol, shown in Fig. 5.7, compared with WDM when

considering an optical link with loss and lumped amplification.

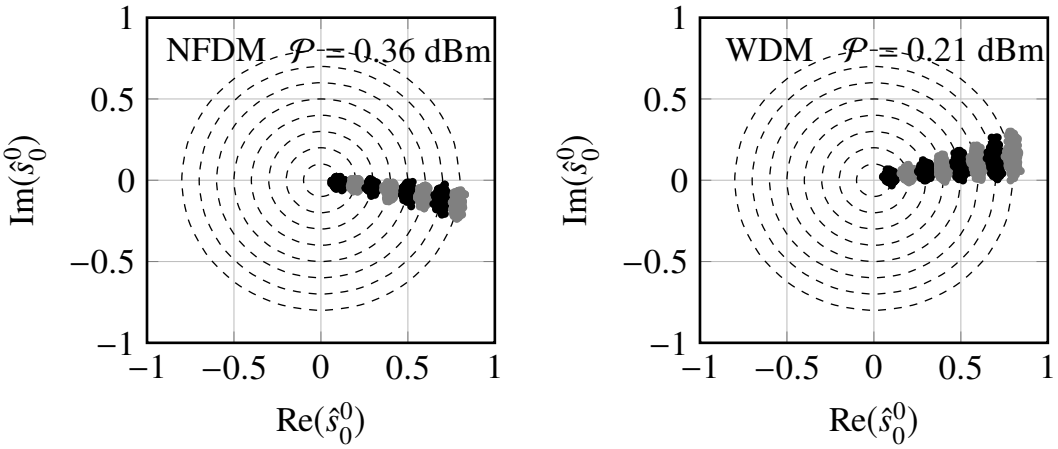

Figure 5.6: Received constellations of (left) NFDM and (right) WDM systems in the focusing regime for the non-integrable model.

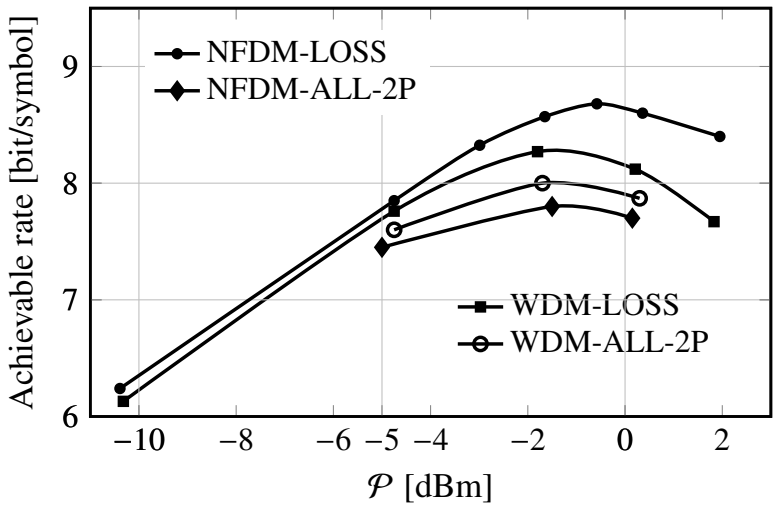

Figure 5.7: AIRs of WDM and NFDM in non-integrable models. The curves denoted by LOSS correspond to the lossy model with periodic amplification (no PMD). The points denoted by ALL include all perturbations (loss, PMD and third-order dispersion).

5.5.2 Polarisation effects and higher-order dispersion

Two signals modulated in two polarisations of light may travel at different speeds along the fibre because of the randomly varying fibre birefringence. This leads to a different arrival time of the signals on two polarisations that is known as PMD; see also Sec. 2.2.4. There is also third-order dispersion, which can cause temporal broadening if the signal bandwidth is large. The impacts of the PMD and higher-order dispersion on WDM have been well studied and understood. It is shown

Table 5.3: polarisation and dispersion parameters

β_3	$0.06\ \mathrm{ps}^3 \cdot \mathrm{m}^{-1}$	third-order dispersion coefficient
D_p	$0.1\ \mathrm{ps}/\sqrt{\mathrm{km}}$	PMD parameter
ℓ_c	1 km	section length ℓ_c

in [164] that the performance improvement in applying DBP to WDM saturates due to PMD as the number of back-propagating channels is increased. On the other hand, the impact of these effects on NFDM has not been investigated to date.

The propagation of two signals on two polarisations of light can be simulated by the CNLSE as explained in (2.47). We write it here for convenience

$$\begin{aligned}\frac{\partial \mathbf{A}}{\partial z} =&(-\frac{\alpha}{2}-\beta_1\frac{\partial}{\partial t}-j\frac{\beta_2}{2}\frac{\partial^2}{\partial t^2}+\frac{\beta_3}{6}\frac{\partial^3}{\partial t^3})\mathbf{A}+\\ &j\gamma\Big[|\mathbf{A}|^2\mathbf{I}-\frac{1}{3}(\mathbf{A}^H\sigma_3\mathbf{A}\sigma_3)\Big]\mathbf{A}\\ =&(\hat{\mathbf{D}}+\hat{\mathbf{N}})\mathbf{A}.\end{aligned} \tag{5.11}$$

The average amount of pulse broadening due to DGD is approximately $D_p\sqrt{L}$, where L is the (total) fibre length and D_p is the PMD parameter [67, 165]. NFDM simulations were carried out in such a channel with parameters in Tab. 5.3 using the same method described in Sec. 5.4.1. Note that at this stage, the NFT-signals on two polarisations are generated individually using SP-INFT, instead of DP-INFT. This choice can be justified by the reduced-complexity and undegraded performance of such a scheme explained in [166]. Nevertheless, optical fibre communication systems based on DP-NFT will be investigated in Chap. 6.

We assume that the receiver has the channel state information. Under this assumption, all random rotations and random DGDs during propagation are compensated at the receiver in both schemes. In commercial systems, this assumption is fulfilled by the polarisation demultiplexing algorithm such as the constant-modulus algorithm (CMA).

In WDM, the UOI is filtered in the frequency domain in both A_1 and A_2 and the resulting vector is back-propagated according to (2.54). This includes the inter-polarisation non-linearity mitigation (jointly across x and y) and the third-order dispersion compensation. In NFDM, the UOI is also filtered in the non-linear frequency domain and applies SP-NFT to the signal on each polarisation. Note that while the effects of β_2 and γ are equalised in the non-linear Fourier domain via the channel filter (5.5), the third-order dispersion and non-linear interactions between the two polarisations act as distortions in NFDM. Fig. 5.8 shows the received constellations for WDM and NFDM with two polarisations. The cor-

responding AIRs are shown in Fig. 5.7. There is a performance degradation in both multiplexing schemes compared to the single polarisation case. The AIR of the WDM and NFDM are nearly equal when all perturbations are included.

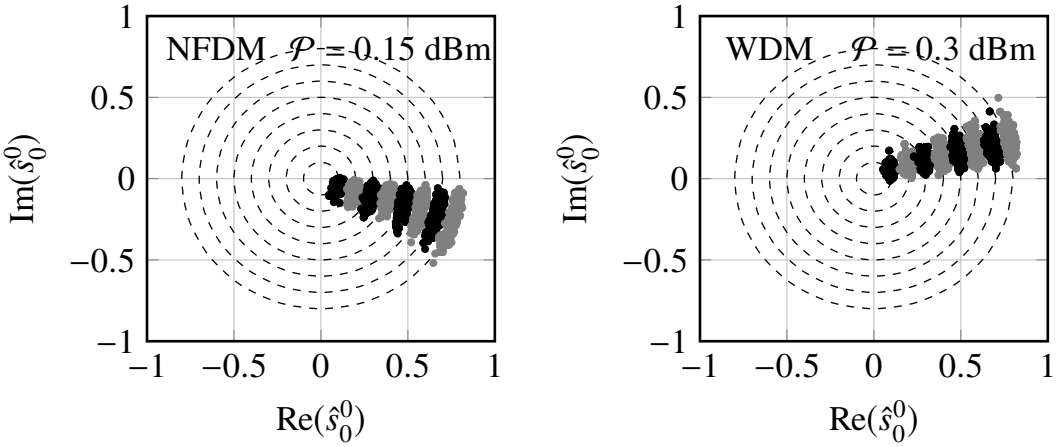

Figure 5.8: Received constellations of (left) NFDM and (right) WDM in the focusing regime with PMD and third-order dispersion, in dual polarisation transmission.

We conclude that the AIR of DP transmission of two SP-INFT-signals and SP-NFT detection subject to third-order dispersion and loss is worse the WDM AIR with joint DP-DBP, in a system with $N_u = 5$ and $N_s = 1$. However, it has been shown recently that if joint (de)-modulation is performed across two polarisations in NFDM using the NFT of the two-dimensional signals, the NFDM outperforms WDM in terms of Q-factor [133]. This technique will be explored in the next chapter.

The algorithm used in this work is a simple discrete layer-peeling method described in Sec. 4.3. More accurate algorithms exist that allow operation at higher powers and reduce the contribution of the numerical error to the AIR, which is desirable at high powers. These better algorithms, such as that in [167], likely improve NFDM and its AIR presented in this work.

5.5.3 Spectral efficiency

In the simulations thus far, we considered one symbol per user and calculated the data rates in bit/symbol. In this section, we estimate the SE in bit/s/Hz by transmitting multiple symbols per channel use.

Numerical computation of the SE is difficult because the number of symbols N_s should be sufficiently large so that the ratio r_g between the guard-time and the information block is small. With small r_g, the SE of WDM can be equivalently computed by considering only the main lobe of the pulse shape (e.g., a RRC function) as the symbol duration. However, with $N_s = 1$, we have effectively very large

r_g, it is not possible to only consider the main lobe of pulses in both WDM and NFDM because there is no clear definition of the main lobe of such signals. Instead, we measure the time duration that contains 99% energy of the pulses. This would result in a loss in SE. As N_s increases, the loss in SE is expected to diminish, providing that NFT with large N_s does not introduce too much numerical error. Note that we do not consider signals generated by periodic NFT [157] here.

The data rates shown in Fig. 5.7 were obtained for the same time duration and bandwidth in WDM and NFDM. As a result, the gain in SE in bit/s/Hz (ratio of the SE of the NFDM and WDM) should be the same as the gain in AIR in bit/DoF. Nevertheless, its is still instructive to see the absolute numbers for the approximate SE as well. To do so we consider a system with $N_u = 5$ users, $N_s = 15$ and parameters in Tab. 5.2. The pre-compensation technique in [10] is used in NFDM to reduce the guard time. Its effect on the waveform is shown in Fig. 5.9. The split-DBP was applied in WDM simulations so that the signals in the two systems have approximately the same temporal broadening.

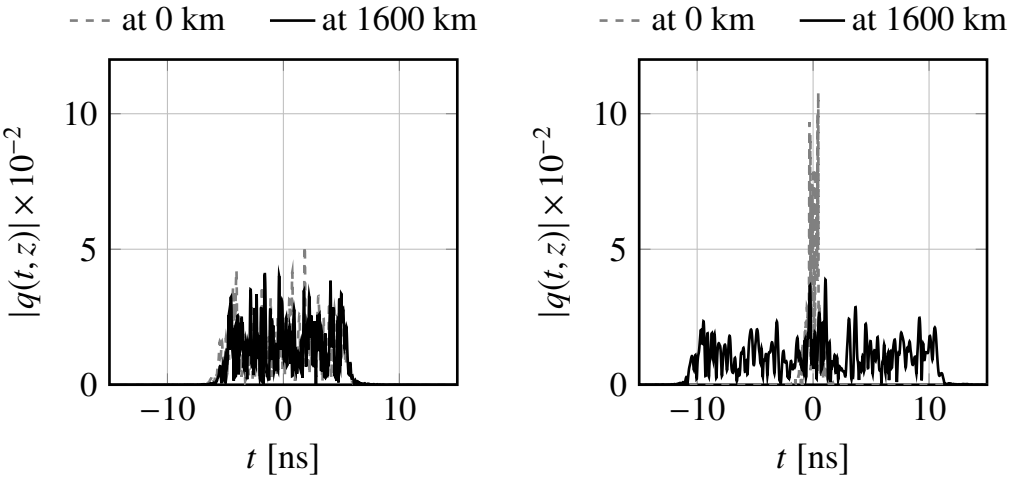

Figure 5.9: NFDM signals (left) with and (right) without pre-compensation.

In the multiple-symbol transmission, symbols at different time instances are treated as independent and identically distributed random variables for simplicity. Fig. 5.10 shows the received symbols in the NFDM and WDM system. It can be seen that the size of clouds is bigger for $N_s = 15$ compared to $N_s = 1$. This effect is attributed to the growth of the signal-noise interaction with the number of DoFs. The AIRs can be again estimated from the statistics of many noise realisations. The SEs are computed as the ratio between AIRs and the time-bandwidth product. The SEs of the NFDM and WDM system are shown in Fig. 5.11. The WDM SE is a little greater than NFDM SE. We attribute this to the fact that the NFT algorithm,

as well as potentially the path-averaged loss model, become less accurate as N_s is increased. This conjecture will be observed in numerical simulations as shown in Sec. 6.6.

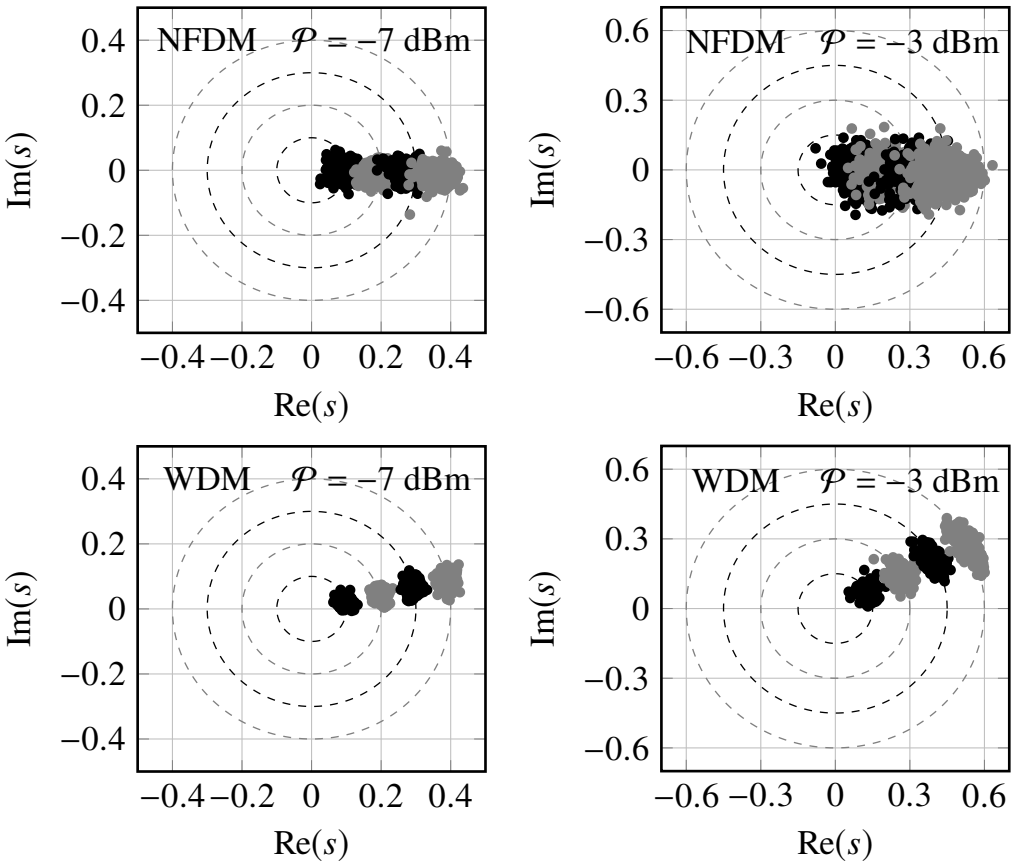

Figure 5.10: Received symbols in NFDM and WDM with $N_s = 15$ at different powers.

For clarity, we also compare time duration T, bandwidth W and guard-time T_g requirement of NFDM and WDM in Fig. 5.12. It can be seen that, for a fixed T, W and P at the input, T and W at the output, and thus T_g, are approximately the same in both schemes. The maximum input output time duration and bandwidth is about the same for all values of powers, around 11 ns and 96 GHz.

A SE of 2 bit/s/Hz was reported in [168], however the distance and bandwidth there are half of the values in this work. Doubling the distance will double the guard-time, and also introduce more noise. Although theoretically linear and non-linear multiplexing should require approximately the same guard-time, in practice r_g is lower for WDM because transmission is possible for larger N_s. This is, however, due to the lower complexity advantage of WDM, and not a fundamental performance difference. The SE may be maximised by generating compact signals using the periodic INFT [157] or the standard INFT with a modified modulation [169].

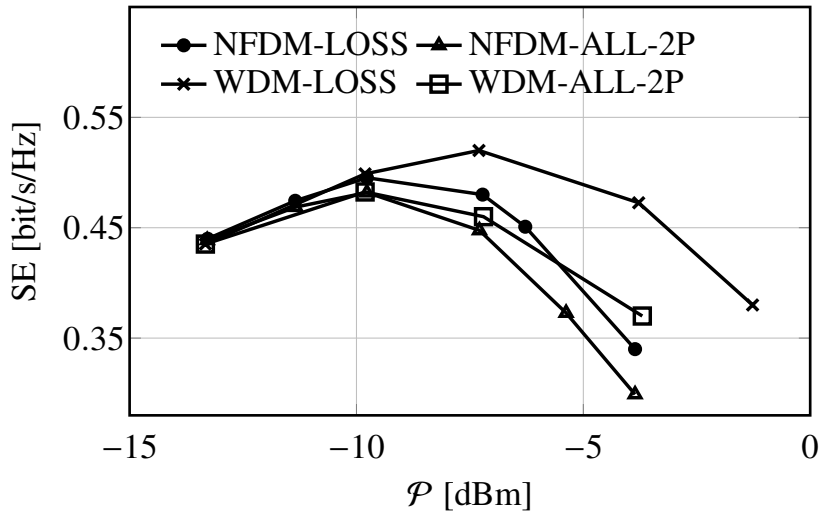

Figure 5.11: The SE of NFDM and WDM with $N_s = 15$ and guard-time. For dual-polarisation transmissions in both systems, the SE and power is measured per polarisation.

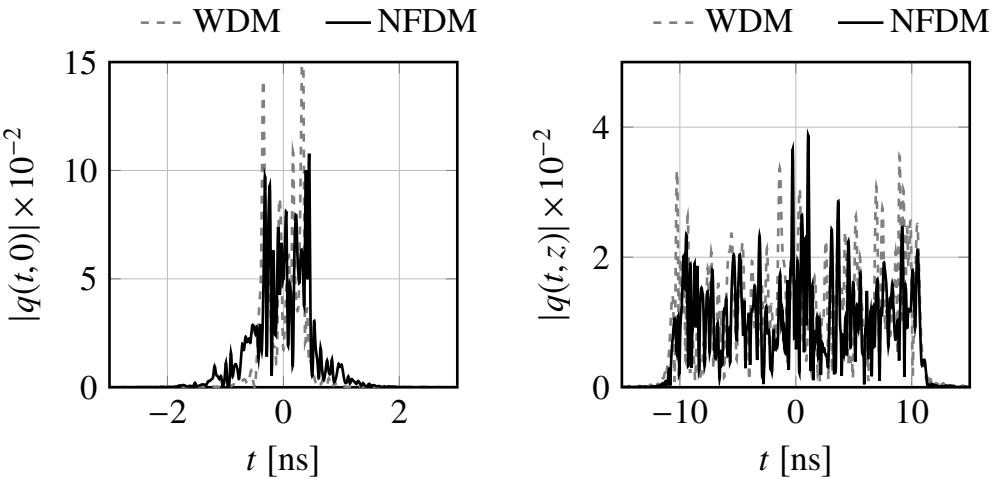

Figure 5.12: Comparison of the time duration at the (left) input and (right) output. Here $N_u = 5$ and $N_s = 15$. The time duration at input is $T_{\text{in}} \approx 2.1$ ns, while $T_{\text{out}} \approx 22.3$ ns, which can be reduced to 11 ns as shown in Fig. 5.9. The temporal expansion ratio for $N_s = 15$ is 10.

5.6 Comments on complexity

The computational complexity of the NFT with continuous spectrum in terms of the number of samples N in time is $O[N \log^2 N]$ [170]. Lima et al. investigated the required number of samples in the non-linear frequency per processing frame (resolution spectrum) in terms of the average power in the defocusing regime [167]. The processing frame is defined as the pulse duration at the Rx, which is typically longer than pulse duration at the Tx. It is argued that a prohibitively fine spectral

resolution is needed at high powers for NFDM to be viable in practice. The algorithm used in [167] is the discrete layer-peeling (DLP) method used in this work, but together with applying the Gelfand-Levitan-Marchenko (GLM) integral equations in each small step in time. It is known that such refinements of the DLP reduce the numerical error, albeit with increased complexity.

To reduce complexity, it is desirable to minimize pulse broadening and the inter-frame guard times. We apply the pre-equalisation proposed by Tavakkolnia et al. (with $l_1 = L/2$). This causes the signal to contract in the time domain as it evolves in $0 \leq z \leq L/2$, and subsequently broaden as it evolves in $L/2 \leq z \leq L$. Consequently, the signal durations at the Tx and Rx are similar.

Tab. 5.4 compares the number of samples N in the non-linear frequency domain and N_s in NFDM and WDM for one user with an oversampling rate of 3; see (5.4). The NFT and INFT are implemented by the basic forward and inverse DLP algorithms in the focusing regime, whose details can be found in [171]. The complexity of the DBP algorithm using the SSFM is evidently $O[MN \log N]$, where M is the number of segments in distance. In our simulation $M = 1600$, corresponding to a step size of 1 km, which is 20-30 times smaller than what is usually used. The complexity of the two algorithms was compared in terms of the total number of floating-point operations (FLOPs), which may depend on details of the implementations. The error E_s is the symbol error defined as $\|\{s_\ell^k\} - \{\hat{s}_\ell^k\}\|/\|\{s_\ell^k\}\|$, where $\{s_\ell^k\}$ and $\{\hat{s}_\ell^k\}$ are defined in Sec. 5.3.

The NFT algorithms typically work up to a certain power before numerical error becomes significant. Tab. 5.5 shows the maximum power that can be reached using the basic forward and inverse DLP with pre-equalisation.

It can be concluded that the complexity of the NFT with the DLP algorithm is substantially higher than the complexity of the fast Fourier transform (FFT), and rapidly grows with N_s. This seems to be consistent with the findings of [167] too. The main difference is that in [167], the complexity and the maximum power of INFT algorithms are studied in normal dispersion fibre where the reflection coefficient is pushed exponentially close to unity with the growth of signal powers.

Note that NFDM currently requires successive NFT and INFT operations at ADMs digitally in the electrical domain (thus requiring optical-to-electrical conversion). In WDM, these operations can be done easily using optical signal processing (without expensive optical-to-electrical conversion). As a result, NFDM is currently impractical. Implementing NFDM using optical signal processing is an interesting and important research direction.

Table 5.4: Complexity of NFDM and WDM

	$\mathcal{P}$[dBm]	N_s	N	E_s	FLOP
NFDM	0.33	2^7	2^{17}	1.13%	2.23×10^{11}
WDM	0.33	2^7	2^8	0.24%	6.93×10^6
NFDM	-0.9	2^6	2^{14}	1.16%	3.49×10^9
WDM	-0.5	2^6	2^7	0.24%	3.46×10^6

Table 5.5: Maximum power reached by the basic DLP

$\mathcal{P}$[dBm]	N_s	N	E_s
0.33	2^7	2^{17}	1.13%
-0.9	2^6	2^{14}	1%

5.7 Summary

In this chapter, the AIRs of the WDM and NFDM were compared. It is shown that: (i) the NFDM AIR is greater than the WDM AIR in an ideal integrable model with five users and one symbol per user; (ii) the AIR of the independent SP-NFDM transmission and detection (not using joint NFT) subject to third-order dispersion and per-span loss is worse the WDM AIR with joint DP back-propagation, in a system with five users and one symbol per user.

Simulations of five users and multiple symbols per channel use were also carried out. The NFDM AIR in bit/s has not surpassed the WDM AIR. This is attributed to the inefficient information modulation method used in the research described in this chapter. In the next chapter, research will be focused on investigating efficient modulation method to increase the AIR in NFT-based transmission systems in bit/s.

6

Dual-polarisation NFDM transmission with b-modulation

In the previous chapter, it has been shown that in an illustrative simulation of optical fibre networks with 5 users and 1 symbol per user in each channel use, NFDM system has achieved a higher AIR than WDM system over the simulated power range. However, when the number of symbols N_s per channel use was increased in simulations, the estimated SE has not been encouraging. This has led to a reconsideration of the NFT-signal generation process. Therefore, the research in this chapter investigated an alternative approach to modulate information on the nonlinear coefficients. The new transmission scheme, with careful optimisation of relevant parameters, has achieved record-high data rates of NFT-based systems in both simulation and experiment.

6.1 Related work

In the last few years, there has been very active research on the NFDM transmission systems. For the simplicity, the concept of NFDM has been firstly considered for single polarisation transmissions [140]. Many research groups have put a lot of efforts in demonstrating NFDM transmission systems in experiments [150, 172, 173]. Many encouraging results have been reported, showing that data rates beyond 100 Gbit/s per channel can be achieved. To date, the highest data rate so far of

a single polarisation NFDM transmission is 150 Gbit/s gross data rate (SE of 2.3 bit/s/Hz) over 976 km [174], which was achieved by modulating simultaneously 222 non-linear spectral components (or non-linear modes). Further increasing the data rate requires increasing the number of modulated non-linear modes, which has been shown to be challenging due to the non-linear cross-talks among non-linear modes in the presence of ASE noise, fibre loss and implementation imperfections [127]. For a thorough and up-to-date review, the readers are referred to [154]. More recently, it was proposed to modulate information on b-coefficients, instead of q_c-coefficients as in Chap. 5, in order to generate time-limited SP-NFDM signals [169]. Subsequently, the experiment reported in [175, 176] showed improved SNR of b-modulation in an SP-NFDM system. The improvement was partially attributed to the compactness of time-domain signals generated by b-modulation, which results in less tail-truncation. Another effective way to increase the data rate of NFDM transmission is to extend the concept of NFDM to DP transmission. In recent year, some proof of concept DP-NFDM transmission systems, both in simulation [133, 177] and experiments [172] have been reported. They all modulated q_c-coefficients. However, high data rate DP-NFDM transmission have not been reported. Some recent results are summarised in Fig. 6.1 in terms of gross data rate. We also categorise some recent implementations of NFDM systems in Tab. 6.1 in terms of the type of modulation, SP or DP, and simulation or experiment.

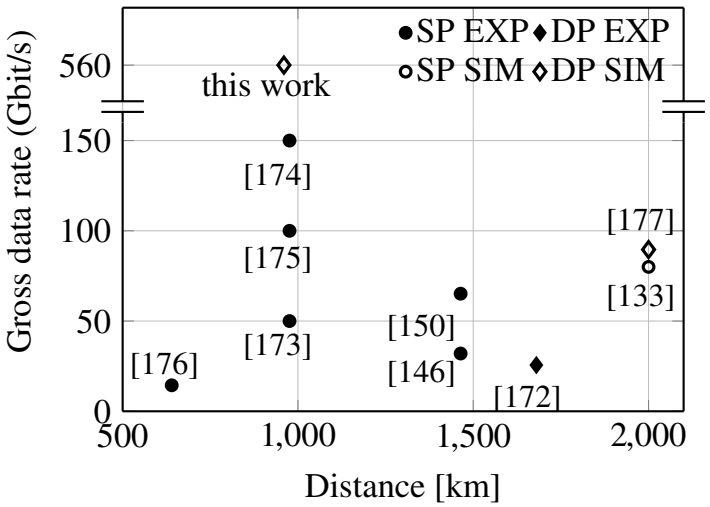

Figure 6.1: Gross data rate of recently implemented NFDM systems of different transmission distances in experiments and simulations.

Contributions of the research described in this chapter are threefold, i) b-modulation was modified for DP by a transformation built upon an existing one for SP in [115], and the scheme was compared with q_c-modulation in DP-NFDM

Table 6.1: Categorisation of recent results of NFDM systems.

		Continuous		Discrete		Both
		q_c	b	q_d	b_k	
Simulation	SP	[177]		[147]		[144]
	DP	[133]	[178]	[179]	[155]	
Experiment	SP	[146, 173, 174]	[175]	[148, 152]		[180]
	DP	[172]	[176]		[181]	[182]

systems, showing 1 dB Q-factor improvement. This can be observed similarly in [175, 176] for SP. To have a deeper insight about the underlying reason, we also studied the correlations of information sub-carriers in q_c-modulation and in b-modulation via the relevant information theoretic metrics of joint and individual entropies. A clear link has been observed between the Q-factor improvement and the weaker correlation of sub-carriers. ii) the b-modulated DP-NFDM system was systematically optimised for high data rate. As a result, a record net data rate of 400 Gbit/s with a SE of 7.2 bit/s/Hz was achieved in simulation over 12 spans of 80 km SSMF with EDFAs. iii) an b-modulated DP-NFDM system was demonstrated in experiment over an optical fibre link with same parameters, achieving a data rate of 220 Gbit/s with a SE of 4 bit/s/Hz.

The chapter is organised as follows: Sec. 6.2 describes the optical fibre channel in DP. Sec. 6.3 introduces the q_c- and b-modulated DP-NFDM systems. The optimisation of b-modulated system is carried out in Sec. 6.4. Sec. 6.5 analyses the correlation of sub-carriers in both q_c- and b-modulation. Some remarks on the drawbacks of the proposed system and their quantification are given in Sec. 6.6. The rest of the chapter discusses the b-modulated DP-NFDM experiment and its relevant DSP.

Notations

The continuous DP-signals are denoted as $\mathbf{A}(t,z) = [A_1(t,z)\ A_2(t,z)]$. The Euclidean norm is represented by $\|\cdot\|$. A discrete complex random vector of length N and its elements are both written in upper case letters, without and with subscripts, respectively, e.g., $\mathbf{X} = [X_1\ X_2\ \ldots\ X_N]^T$. A realisation of the complex random vector is denoted in lower case letter such as $\mathbf{x} = [x_1\ x_2\ \ldots\ x_N]^T$. In the case of two discrete random vectors, double subscripts p and k are used, where $p \in \{1,2\}$ is the vector index and k is the element index, e.g., X_{pk} denotes the k-th element of the p-th vector. Also note that, a continuous signal is always written with its dependency of other variable, e.g., $a(\lambda)$, whereas its discrete counterpart is often written in upper case with subscripts, e.g., A_k. At last, NFT and INFT operations are denoted as NFT$(\cdot)$ and INFT$(\cdot)$. $\mathcal{H}(\cdot)$ denotes the Hilbert transform.

6.2 The optical fibre channel model

The optical fibre channel model of concern is a multi-span point-to-point dual-polarisation dispersion unmanaged fibre link, which can be described by the Manakov equation (2.54) described earlier. For convenience, we rewrite it here

$$\frac{\partial \mathbf{A}}{\partial z} + \frac{\alpha}{2}\mathbf{A} + \frac{j\beta_2}{2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - j\frac{8}{9}\gamma \mathbf{A}\,\|\mathbf{A}\|^2 = 0 \tag{6.1}$$

where $\mathbf{A}(t,z) = [A_1(t,z)\ A_2(t,z)]$ is the complex envelope of the DP-signal as a function of time t and distance z along the fibre. Tab. 6.2 lists all parameters used in this work.

Table 6.2: Fibre and system simulation parameters

ν	193.44 THz	centre carrier frequency
α	0.2 dB km^{-1}	fibre loss
γ	1.3 (W $\cdot$ km)$^{-1}$	non-linearity parameter
β_2	-21.5×10^{-27} s^2/m	group velocity dispersion
W	56 GHz	linear bandwidth
R_o	8	oversampling rate
L_{sp}	80 km	span length
N_{sp}	12	number of spans
NF	5 dB	EDFA noise figure

6.3 q_c- and b-modulated DP-NFDM systems

The independent evolution of $\mathbf{b}(\lambda)$ for each non-linear frequency λ motivates to modulate the data independently over separate sub-carriers in the non-linear spectrum as it is done by the OFDM schemes in the classical linear channels. Therefore, we generate an NFDM symbol in the non-linear spectrum in a similar manner of an OFDM symbol in the (linear) Fourier spectrum. Here, we borrow the notations from [183] to explain the generation of NFDM symbols, either in b-domain or in q_c-domain.

Consider the input and output of the inverse discrete Fourier transform (IDFT). The input of IDFT is a complex vector $\mathbf{x} = [x_1\ x_2\ \ldots\ x_N]^T$ of length N. Each element x_k represents the data on the corresponding sub-carrier, chosen from a given constellation, e.g., 32-QAM in this work. The output of IDFT is a complex vector $\mathbf{y}$, representing the discrete time-domain samples. The IDFT is defined by

$$y_m = \frac{1}{N}\sum_{k=0}^{N-1} x_k \exp\left(\frac{j2\pi km}{N}\right), \quad \text{for } 0 \le m \le N-1. \tag{6.2}$$

Correspondingly, the DFT is

$$x_k = \sum_{m=0}^{N-1} y_m \exp\left(\frac{-j2\pi km}{N}\right), \text{ for } 0 \le k \le N-1. \tag{6.3}$$

Upon implementation, a suitable circular shift of vectors is assumed to ensure that the zero-frequency component of $\mathbf{y}$ is always at the centre of spectrum. We also define $f(\mathbf{x}, M)$ as a function that adds M zeros to a vector $\mathbf{x}$ of length N in the following way

$$f(\mathbf{x}, M) = [\underbrace{0 \ldots 0}_{M/2} \; x_1 \ldots x_N \; \underbrace{0 \ldots 0}_{M/2}], \tag{6.4}$$

which is used later to describe the action of up-sampling and adding guard interval.

The chromatic dispersion causes temporal broadening. Thus, some guard interval (GI) between NFDM symbols are necessary to avoid inter-symbol interference. The required interval T_G can be estimated by [174]

$$T_G \approx \pi W \beta_2 L_{\text{sp}} N_{\text{sp}} = 3.75 \text{ ns}, \tag{6.5}$$

which depends mainly on signal bandwidth W and transmission distance $L_{\text{sp}}N_{\text{sp}}$ that were both kept constant in this work (given in Tab. 6.2). Let N_C denote the number of sub-carriers in each polarisation. The pulse duration is $T_0 = N_C/W$. The total pulse duration is $T_0 + T_G$, which reduces the SE. We define $\eta = (T_0 + T_G)/T_0$, showing the SE loss due to the GI. Note that the GI insertion is done in the generalised time domain before the DP-INFT, as shown in Fig. 6.2(e).

Let $\mathbf{x}_p = [x_{p1} \; x_{p2} \; \ldots \; x_{pN_C}]^T$ denote the data modulated on N_C sub-carriers for each polarisation $p \in \{1, 2\}$. Before computing the discrete samples of the continuous spectrum, we describe the action of adding GI and up-sampling

$$\mathbf{d}_p = \text{IDFT}\{f(\mathbf{x}_p, N_C(R_o - 1))\}, \quad \text{(up-sampling)} \tag{6.6}$$

$$\mathbf{u}_p = \begin{cases} \text{DFT}\{f(\mathbf{d}_p, N_C R_o(\eta - 1))\}, \eta \ge 2, \\ \text{DFT}\{f(\mathbf{d}_p, N_C R_o)\}, 1 < \eta < 2, \end{cases} \quad \text{(adding GI)} \tag{6.7}$$

where R_o is the oversampling rate as in Tab. 6.2 and $N_C R_o(\eta - 1)$ is rounded up to its nearest even number. Next, the discrete samples of the continuous spectrum, in either $\mathbf{b}(\lambda)$ or $\mathbf{q_c}(\lambda)$, are computed from the resulted vectors $\mathbf{u}_p = [u_{p1} \; u_{p2} \; \ldots \; u_{pN}]^T$, $p \in \{1, 2\}$ as follows: for q_c-modulation schemes, the samples

of $\mathbf{q_c}(\lambda) = [q_{c1}(\lambda)\ q_{c2}(\lambda)]$ are obtained from $[\mathbf{u}_1\ \mathbf{u}_2]$ via the transformation

$$q_{c,pk} \quad = \sqrt{e^{|u_{pk}|^2} - 1} \cdot e^{j \arg\{u_{pk}\}},\ p \in \{1,2\}. \tag{6.8}$$

The above transformation, denoted here by Γ_c, is already explored for DP-NFDM systems in [133]. For b-modulation schemes, the samples of $\mathbf{b}(\lambda) = [b_1(\lambda)\ b_2(\lambda)]$ are obtained from $[\mathbf{u}_1\ \mathbf{u}_2]$ via a new transform

$$\Delta_k = \sqrt{1 - \exp\Big(-|u_{1k}|^2 - |u_{2k}|^2\Big)} \Big/ \sqrt{\Big(|u_{1k}|^2 + |u_{2k}|^2\Big)}, \tag{6.9}$$

$$b_{1k} = \Delta_k \cdot u_{1k}, \tag{6.10}$$

$$b_{2k} = \Delta_k \cdot u_{2k}, \tag{6.11}$$

$$a_k = \sqrt{1 - |b_{1k}|^2 - |b_{2k}|^2} e^{j\mathcal{H}\left(\frac{1}{2}\log\left(1 - |b_{1k}|^2 - |b_{2k}|^2\right)\right)}. \tag{6.12}$$

The above transform, denoted here by Γ_b, is the natural extension of the one introduced in [115] for SP-NFDM systems in the defocusing regime. Before taking INFT, the technique of PDC is applied on the resulted continuous spectrum, either $\mathbf{b}(\lambda)$ or $\mathbf{q_c}(\lambda)$, to minimise the required GI [10]. Finally, the time-domain NFDM signal of duration $T_0 + T_G$ is generated by applying DP-INFT algorithm. At the receiver, the DP-NFT and back-rotation equalisation are performed. Fig. 6.2 illustrates the transceiver digital signal processing structure and a realisation of signals at different stages of INFT-signal generation are illustrated. Note that, for $\eta < 2$, each signal generated by INFT has a duration of $2T_0$. Before transmission, the signal will be truncated symmetrically on both sides to a signal of duration $T_0 + T_G$ for transmission and also recovered to the duration $2T_0$ by adding zeros before NFT processing. This is due to the heuristic reason that higher spectral resolution is needed for NFT as N_C increases.

For any $\lambda \in \mathbb{R}$, the constraint $|b_1(\lambda)|^2 + |b_2(\lambda)|^2 < 1$ specifies a 4-dimensional unit sphere. The transform Γ_b facilitates the signal modulation on b-domain as it maps the entire $\mathbb{R}^4$ into this unit ball. Therefore, it allows using any arbitrary constellation format, e.g. QAM or equi-distance ring constellations without any concern on violating the constraint on b-coefficients. However, the transform Γ_b has a drawback. As shown in [169], modulating b-coefficients allows full control on the duration of NFDM signals in time-domain. Applying non-linear transforms like Γ_b causes losing this property to some extent. Nevertheless, this may not be practically a big issue. Although the generated NFDM signal may have some undesired tails, the tails are partially covered by the necessary long GI (specially after PDC) when the fibre length is long enough. Note that the longer GI is, the more ASE noise

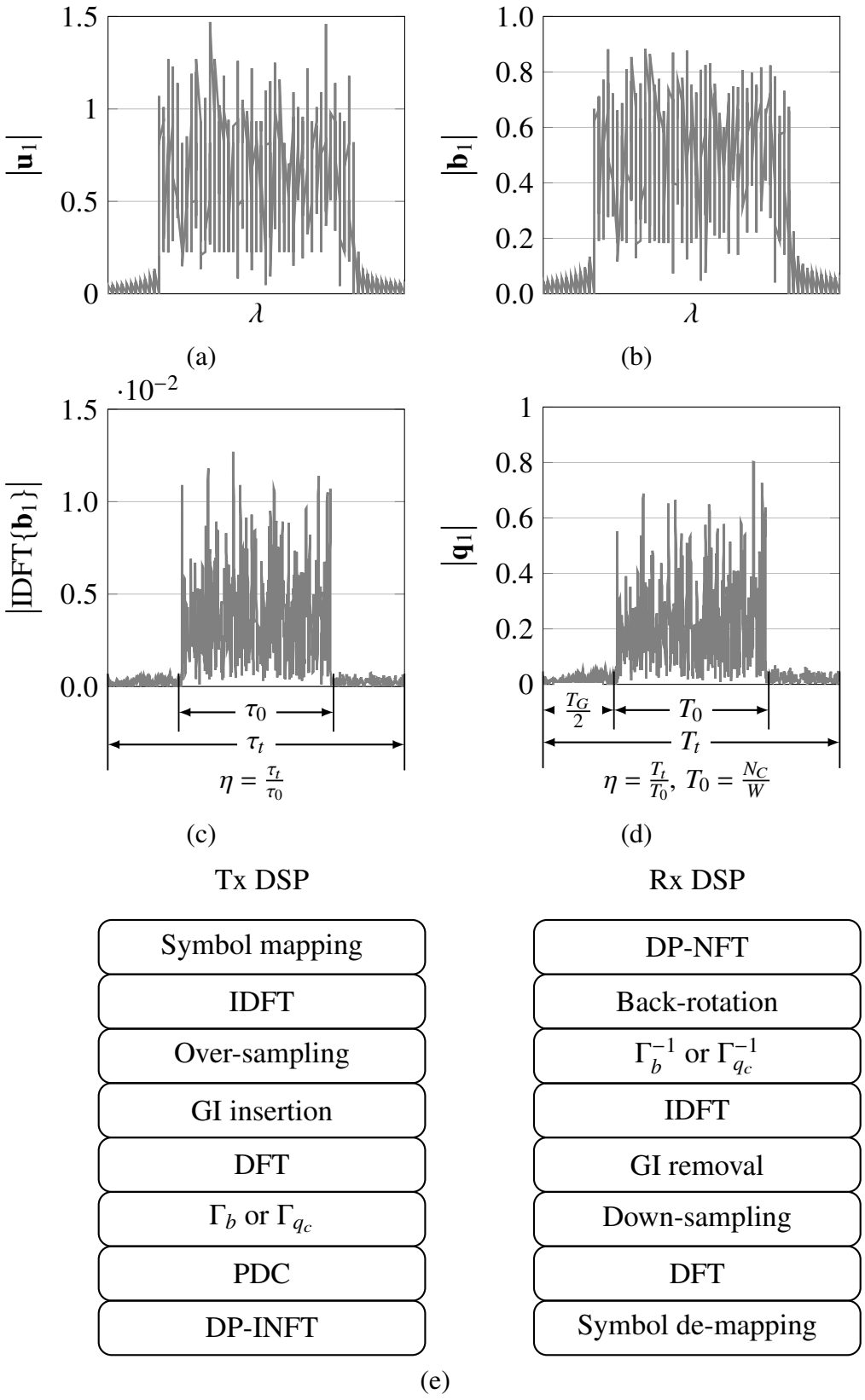

Figure 6.2: A realisation of signal at different stages of INFT-signal generation without pre-dispersion compensation. (a) OFDM spectrum as $\mathbf{u}_1$, (b) $\mathbf{b}_1$ spectrum when applying the transformation Γ_b on $\mathbf{u}$, (c) IDFT of $\mathbf{b}_1$ with tails due to the transformation Γ_b, (d) time-domain signal generated by INFT wihtout PDC. (e) Transceiver digital signal processing chains.

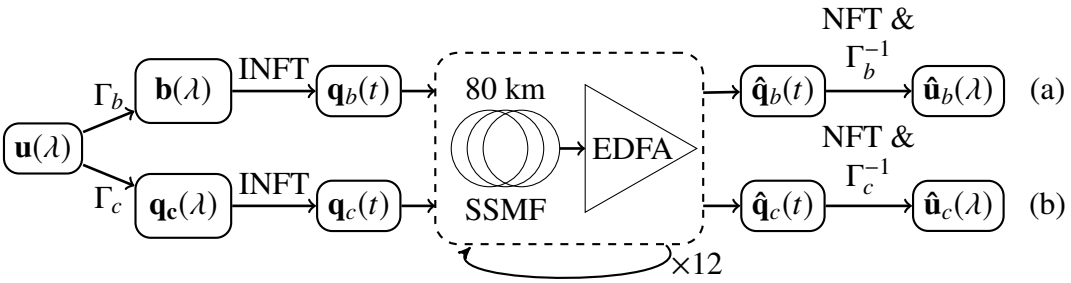

Figure 6.3: Simulation diagram of (a) b- or (b) q_c-modulated DP-NFDM systems. $\mathbf{q}_b(t)$ and $\mathbf{q}_c(t)$ are the time-domain signals of each system.

is taken into the NFT process, causing Q-factor degradation. The optimal GI that balances the extra ASE noise and the truncation of decaying tails is close to the estimated T_G in (6.5).

Directly modulating b-coefficients of SP-NFDM was investigated in [175, 176]. In [175], the clipping technique was applied to satisfy the constraint on $b(\lambda)$. Clipping, an non-invertible operation commonly used for the PAPR reduction in OFDM systems [183, Sec. V-A], inevitably introduces distortion in $b(\tau)$ and cannot lift the constraint completely. In [176], another alternative was proposed to guarantee the constraint on b-coefficients by optimising the carrier waveform and applying constellation shaping. The technique was applied for 9 sub-carriers. It is of practical interest how the design complexity and the achievable SE scale for large number of sub-carriers. Moreover, the above techniques were presented in single polarisation. It is of interest how the joint constraint on $b_1(\lambda)$ and $b_2(\lambda)$ in DP, which limits independent modulation of $b_1(\lambda)$ and $b_2(\lambda)$, may affect the application of both techniques.

We demonstrate now the advantage of b-modulation over q_c-modulation in a relatively ideal case having a large guard interval $\eta = 4$ ($N_C = 70$). The transmission of corresponding NFDM symbols were compared in two different simulated transmission scenarios: 1) the transmission over 12 spans of 80 km SSMF with EDFA amplification as illustrated in Fig. 6.3. It was simulated by the SSFM [184] using the parameters in Tab. 6.2. 2) the AWGN channel without fibre transmission (back-to-back). The total additive noise powers were the same in both scenarios. Fig. 6.4 shows the clear advantage of b-modulation over q_c-modulation in both scenarios. Extra Q-factor degradation in the fibre transmission is attributed to the approximation error of the path-averaged model. The power $\mathcal{P}$ reported in this chapter is always the signal power per polarisation.

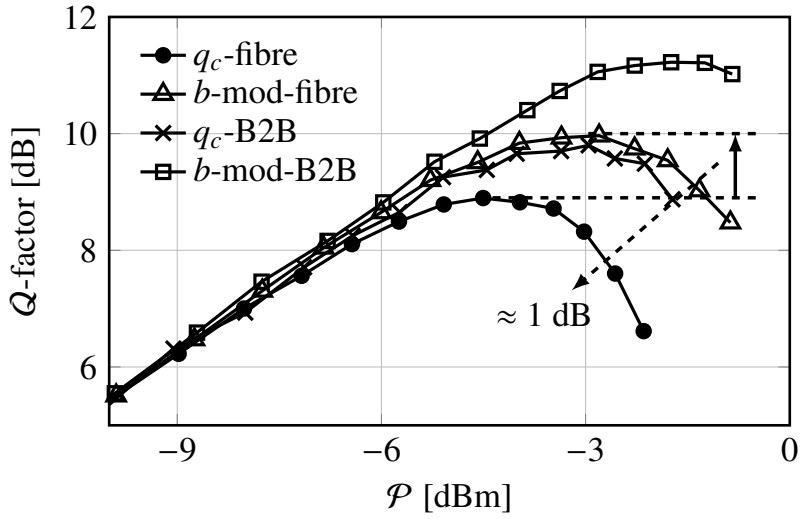

Figure 6.4: Q-factor vs. signal launch power in q_c- and b-modulated DP-NFDM systems with a large guard interval $\eta = 4$ in both AWGN case and fibre transmission. The total additive noise powers are the same.

6.4 Optimisation of b-modulated DP-NFDM transmission

Once the advantage of b-modulation is established, we optimise the b-modulated DP-NFDM system in terms of net data rate for the transmission scenario over $12\times$ 80 km SSMF with EDFAs, visualised in Fig. 6.3. The optimisation parameters are the ratio η and launch power per polarisation $\mathcal{P}$. The number of sub-carriers N_C and η are connected by

$$\eta = (T_0 + T_G)/T_0 = 1 + WT_G/N_C,$$

where $T_G = 3.75$ ns and W is fixed. For instance, it results in $N_C = 210$ for $\eta = 2$. For each pair of $(\eta, \mathcal{P})$, the transmission of the corresponding NFDM symbols was simulated according to the parameters in Tab. 6.2. For each NFDM symbol, the sub-carriers were detected individually and the performance was averaged over a large number of randomly generated NFDM symbols. To compute the SE, we estimate the MI (of individual sub-carrier detection) by assuming the channel conditional distribution as a Gaussian distribution [185, Sec. VI] [9]; see details in Sec. 2.4.2. This assumption underestimates the correct MI and serves as an achievable information rate. The SE is then computed by

$$\text{SE} = \frac{\text{MI} \cdot N_C}{(T_0 + T_G)W} = \text{MI}/\eta \text{ (bit/s/Hz/Pol.)}.$$

We expect that decreasing η will increase SE as well as the signal-noise interaction for a given launch power $\mathcal{P}$. For different values of η, the transmission performance metric Q-factor is plotted in Fig. 6.6(a) and SE in Fig. 6.6(b). They show respectively that Q-factor decreases monotonically as η decreases, while SE reaches its maximum at $\eta = 1.2$. Two received constellations at the power of the maximum Q-factor and SE are also plotted in Fig. 6.6(c) and 6.6(d). For $\eta = 1.2$ ($N_C = 1050$), and at power $\mathcal{P} = -8$ dBm per polarisation, we achieved a SE of 3.6 bit/s/Hz/pol, resulting 400 Gbit/s net data rate, while the gross data rate is 560 Gbit/s. For the sake of comparison, we repeat the simulation for q_c-modulated DP-NFDM systems and plotted the SE curves in Fig. 6.5.

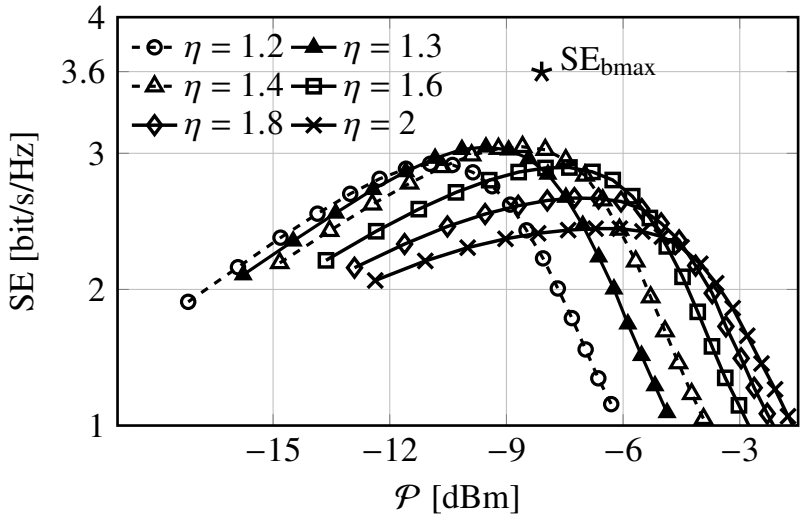

Figure 6.5: The impact of decreasing η on SE in q_c-modulated DP-NFDM system for different launch powers. $N_C = 210/(\eta - 1)$.

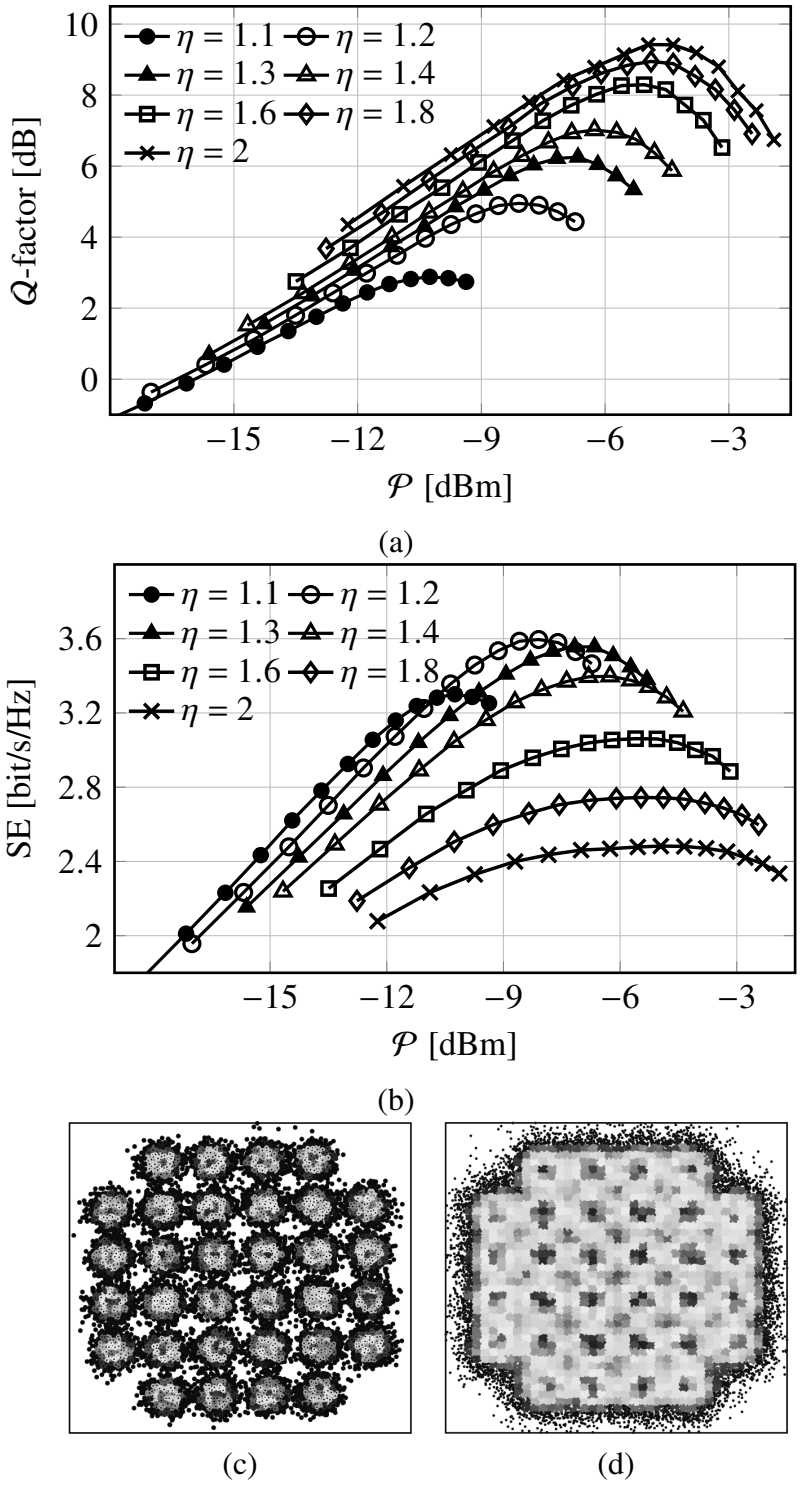

Figure 6.6: The impact of decreasing η on (a) Q-factor and (b) SE in b-modulated DP-NFDM system for different launch powers. $N_C = 210/(\eta - 1)$. The corresponding received constellation at the power of the maximum Q-factor (c) and SE (d).

6.5 Co-variance matrix of q_c- and b-modulated signals

In this section, we give some insights about why the b-modulation scheme outperforms the q_c-modulation scheme. Let us first justify verbally. As mentioned in Sec. 6.2, there is a one-to-one correspondence between $\mathbf{b}(\lambda)$ and $\mathbf{q}_c(\lambda)$ providing that $\mathbf{b}(\lambda)$ or $\mathbf{q}_c(\lambda)$ $\forall\lambda \in \mathbb{R}$ and all discrete eigenvalues $\{\lambda_m\}$ are known. Therefore, $\mathbf{b}(\lambda)$ $\forall\lambda \in \mathbb{R}$ carries the same amount of information as $\mathbf{q}_c(\lambda)$ if the correlation between spectral coefficients, in either $\mathbf{b}(\lambda)$ or $\mathbf{q_c}(\lambda)$ $\forall\lambda \in \mathbb{R}$, can be fully exploited and no discrete eigenvalue emerges[1]. The correlation could come from signal-noise interaction, inaccuracy of the INFT-NFT algorithms, and deviations from the ideal integrable channel model. However, our two DP-NFDM schemes have different performance for the following reasons:

(i) The sub-carriers are detected individually, where the correlation between sub-carriers is neglected. Strong correlations between sub-carriers cause then a large performance degradation.

(ii) The two DP-NFDM schemes use the same modulation technique, i.e. OFDM with 32-QAM modulated sub-carriers. Depending on the distribution of noisy received data, this particular modulation format can be more resilient against noise for one scheme than the other.

We focus here on the first item, the correlation between sub-carriers. We show that the sub-carriers of b-modulated DP-NFDM system have much less correlation than the ones of q_c-modulated DP-NFDM system. To quantify all correlation coefficients with a single meaningful scalar, we use the differential entropy. We compare two quantities: the joint entropy which takes the correlations into account and the individual entropy which neglects the correlations. The gap between these quantities reflects how much the sub-carriers are correlated and how much the net data rate will be decreased (to some extent) by neglecting the correlations.

For both DP-NFDM systems, we numerically compute the differential entropies. Let $\mathbf{X} = [X_1\ X_2\ ...\ X_{2N_C}]$ denote the randomly chosen 32-QAM data modulated on $2N_C$ sub-carriers (DP). Let $\mathbf{Y} = [Y_1\ Y_2\ ...\ Y_{2N_C}]$ denote the received noisy data of sub-carriers after the NFT processing. The joint entropy is defined as

$$h(\mathbf{Y}|\mathbf{X}) = \frac{1}{4N_C}\mathbb{E}_{\mathbf{x}}\left(h(\mathbf{Y}|\mathbf{X}=\mathbf{x})\right)$$

[1]The discrete eigenvalues may emerge because of non-ideal amplification or mixing with ASE noise. However, we did not observe any significant sign of their presence in our simulations.

where $h(\mathbf{Y}|\mathbf{X}=\mathbf{x})$ is the entropy of received data given $\mathbf{x}$, a realisation of inputs $\mathbf{X}$. $\mathbb{E}_{\mathbf{x}}$ denotes taking expectation over all realisations of $\mathbf{x}$. The individual entropy is defined as

$$h_{\Sigma}(\mathbf{Y}|\mathbf{X}) = \frac{1}{4N_C}\mathbb{E}_{\mathbf{x}}\left(\sum_{i=1}^{2N_C} h(Y_i|\mathbf{X}=\mathbf{x})\right)$$

where $h(\mathbf{Y_i}|\mathbf{X}=\mathbf{x})$ is the entropy of the individual received sub-carrier Y_i given the realisation $\mathbf{x}$. To compute the above quantities, we need the conditional probability density $f(\mathbf{Y}|\mathbf{X})$. It is, however, unknown yet for NFDM systems and difficult to estimate due to its large dimension $4N_C$. We approximate $f(\mathbf{Y}|\mathbf{X})$ by a complex multivariate Gaussian distribution. We justify our choice later in **Remark 1.** Define $\mathbf{W} = [\mathrm{Re}(\mathbf{Y}), \mathrm{Im}(\mathbf{Y})]$, then the corresponding probability density $f(\mathbf{W}|\mathbf{X})$ is approximated by a real-valued multivariate Gaussian distribution with covairance matrix $\mathbf{K}$. The element of $\mathbf{K}$ is defined as

$$k_{ij} = \mathbf{E}[(W_i - \mathbf{E}[W_i])(W_j - \mathbf{E}[W_j])].$$

In this case, the conditional entropies can be approximated as

$$h(\mathbf{Y}|\mathbf{X}=\mathbf{x}) \approx \frac{1}{8N_C}\log\left(\det\left(2\pi e\mathbf{K}\right)\right), \tag{6.13}$$

$$h_{\Sigma}(\mathbf{Y}|\mathbf{X}=\mathbf{x}) \approx \frac{1}{8N_C}\log\left(\prod_{i=1}^{2N_C} 2\pi e k_{ii}\right). \tag{6.14}$$

We estimated $\mathbf{K}$ for each realisation $\mathbf{x}$ from the simulated data. We simulated the fibre transmission scenario visualised in Fig. 6.3 for both DP-NFDM systems. We set $\eta = 2$, resulting in $N_C = 210$ sub-carriers. For each DP-NFDM system, we randomly generated 20 input realisations $\mathbf{x}$. For each realisation, the transmission of its corresponding DP-NFDM symbol was simulated 2^{14} times with different random noise realisations. The simulation is repeated for different launch powers $\mathcal{P}$. Accordingly, the above entropies are computed from the resulted in $\mathbf{K}$ for each input realisation $\mathbf{x}$. The entropies of each realisation $\mathbf{x}$ for both DP-NFDM systems at launch power $\mathcal{P} = -3.75$ dBm were calculated and shown in Fig. 6.7(a). It can be observed that the curves have small variations for different realisations $\mathbf{x}$. We finally estimated $h(\mathbf{Y}|\mathbf{X})$ and $h_{\Sigma}(\mathbf{Y}|\mathbf{X})$ by the empirical average of 20 input realisations.

The conditional entropies in terms of launch power for both DP-NFDM systems were calculated and the results are shown in Fig. 6.7(b). It can be observed that the gap $h(\mathbf{Y}|\mathbf{X}) - h_{\Sigma}(\mathbf{Y}|\mathbf{X})$ is smaller for the b-modulated scheme than the one

for the q_c-modulated scheme. It implies that the penalty of individual detection is smaller for the b-modulated scheme. For both schemes, the increase of the gap in terms of launch power implies that the sub-carriers becomes more correlated for larger $\mathcal{P}$. Moreover, $h(\mathbf{Y}|\mathbf{X})$ of the b-modulated scheme is not only smaller but also has a slower growth in terms of $\mathcal{P}$. It indicates that in comparison of both systems, the co-variance matrix $\mathbf{K}$ of the b-modulated scheme has, on average, smaller diagonal elements which grow also slower in $\mathcal{P}$. In plain language, the "effective noise" contaminating b-modulated scheme has smaller power and its power grows also slower with increasing $\mathcal{P}$ than the one of q_c-modulation. Let us emphasise that the constellation of both DP-NFDM systems are the same in $u(\lambda)$ domain (shown in Fig. 6.3) for a fair comparison in Fig. 6.7(b).

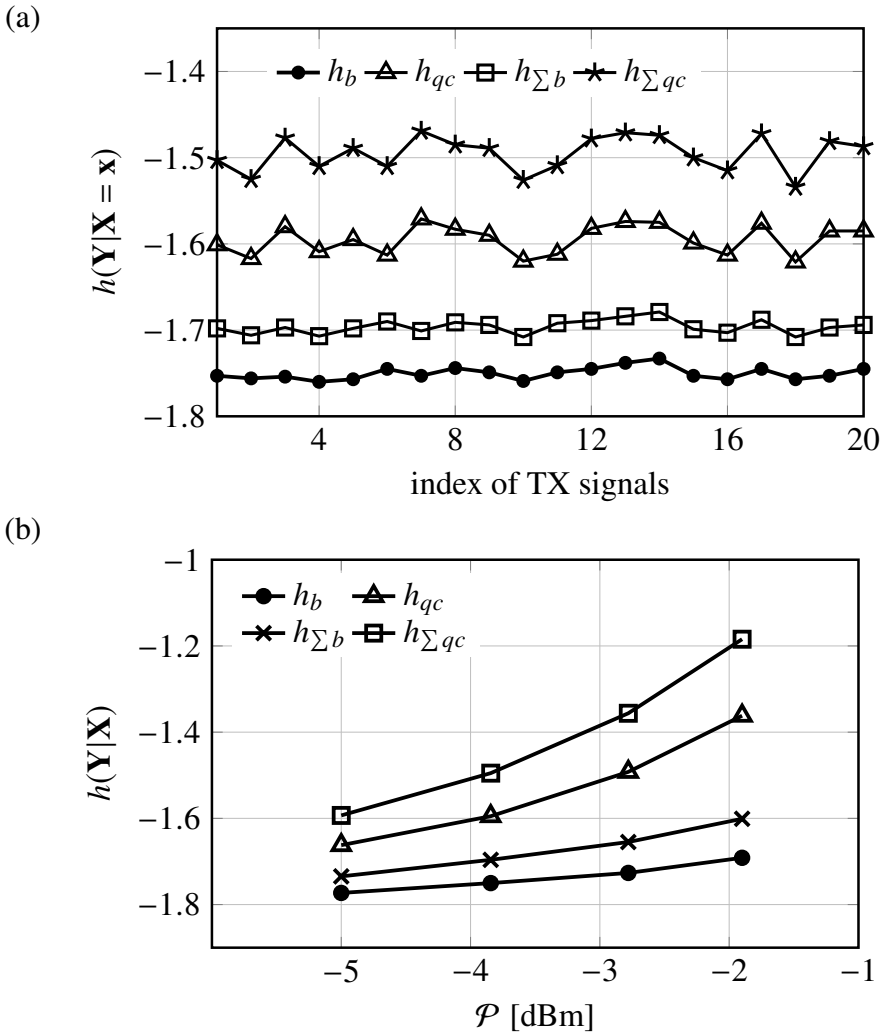

Figure 6.7: Conditional differential entropy of received symbols in both q_c and b-modulated DP-NFDM systems with the assumption of Gaussian distribution, for different transmitted symbols (a) and powers (b).

Remark 1. To verify the validity of entropy approximations, we further applied the entropy estimator in [11], explained in detail in Sec. 2.4.3, which can rather precisely estimate the differential entropy of a low-dimensional random variable from its random samples. To have robust estimations, we estimated the differential entropies of only 10 adjacent sub-carriers (20 real-valued dimensions). For different input realisation $\mathbf{x}$, we observed that the estimations are fairly close to the corresponding entropies of the Gaussian approximation, as shown in Fig. 6.8.

Remark 2. Comparing both DP-NFDM systems in terms of correlation and conditional entropy is more qualitative than quantitative. The joint MI between transmitted and received data on all sub-carriers is the suitable metric to "quantify" the performance difference of both NFDM schemes (with joint detection). The MIs using individual detection are already compared in Sec. 6.4. However, the joint MI requires $f(\mathbf{Y}|\mathbf{X})$ as well as the probability density $f(\mathbf{Y})$. It is more challenging to estimate $f(\mathbf{Y})$ especially when the input $\mathbf{X}$ is drawn from a discrete constellation, e.g., 32-QAM in our case for the following two reasons: The probability density $f(\mathbf{Y})$ is a multi-dimensional multi-center distribution, with correlation between dimensions. Moreover, the co-variance matrix K of $f(\mathbf{Y}|\mathbf{X})$ depends on $\mathbf{X}$. Hence, the entropy $H(\mathbf{Y})$ and consequently, the estimation of MI (jointly between all sub-carriers) much more challenging.

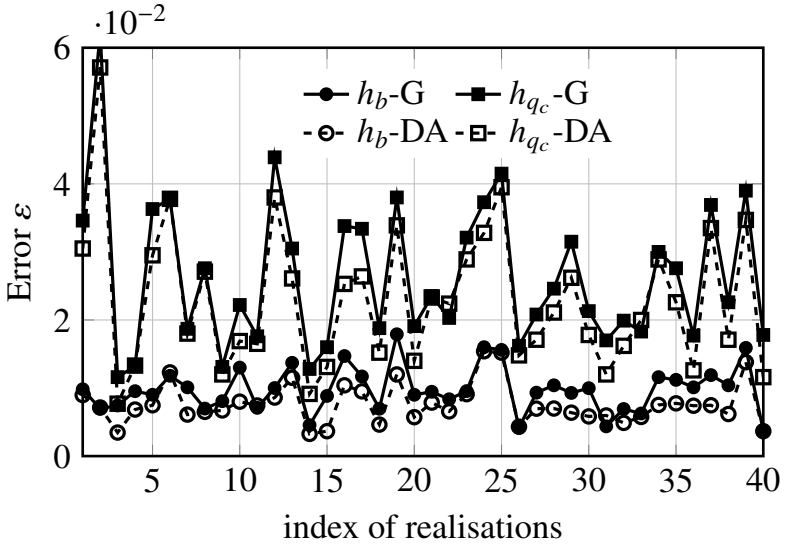

Figure 6.8: The relative error between distribution-agnostic entropy estimator [11] and the Gaussian entropy for 10 adjacent sub-carriers. G stands for Gaussian entropy, DA for distribution-agnostic entropy estimation. For both methods, the result of one-dimensional entropy $h(Y_i|\mathbf{x})$ is very close. $\varepsilon = \left(\sum_{i=1}^{10} h(Y_i|\mathbf{x}) - h(Y_1,...,Y_{10}|\mathbf{x})\right) / \sum_{i=1}^{10} h(Y_i|\mathbf{x})$.

6.6 Some drawbacks of the current system

Based on the above simulation results, we identified two main sources of error in our current system: 'transceiver distortion' in noiseless back-to-back scenario and channel mismatch owing to its non-integrability.

6.6.1 Transceiver distortion in noiseless back-to-back scenario

The transceiver distortion includes the distortion from the entire digital signal processing sequence shown in Fig. 6.2(e), such as the inaccuracy of the INFT-NFT algorithm, the application of Γ_b and the truncation of time-domain signal. Their overall effect can be quantified in an noiseless back-to-back simulation of the b-modulated DP-NFDM system. Fig. 6.9 shows the Q-factor in terms of average energy per NFDM symbol for different η. The average energy of one NFDM symbol can be calculated by $\mathcal{P}(T_0+T_G) = \mathcal{P}\eta N_C/W$. It shows clearly that the transceiver distortion is energy-dependent.

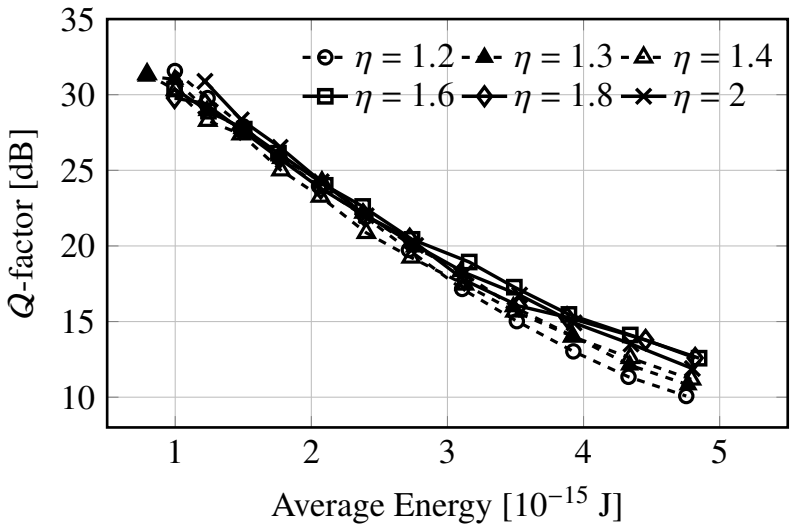

Figure 6.9: 'Transceiver distortion' in the b-modulated DP-NFDM system in the noiseless back-to-back scenario. The average energy of NFDM symbol is calculated by $\mathcal{P}(T_0+T_G) = \mathcal{P}\eta N_C/W$.

6.6.2 Channel mismatch due to non-integrability

NFT is only exact for integrable channels. To successfully apply NFT to the non-integrable channel (6.1), an approximation step has to be taken. To quantify the approximation distortion, one should simulate noiseless systems in Fig. 6.3 with a transceiver DSP that does not cause any distortion. This is unfeasible with our current transceiver DSP because of the transceiver distortion mentioned in the previous subsection. To adopt a different approach, we first clarify the concepts of different DBP schemes, as DBP is the main tool in this channel mismatch quantification.

We recall the Manakov equation (6.1). The ideal DBP refers to the process of solving the Manakov equation to recover $\mathbf{A}(0,t)$ (input) from the boundary condition $\mathbf{A}(z,t)$ (output), using SSFM with fine step size (0.1 km). The EDFAs are replaced with attenuators of the opposite gain. In a single-channel scenario, the ideal DBP should fully cancel the deterministic distortion.

We recall also the integrable path-averaged Manakov equation. The path-averaged DBP refers to the process of solving the path-averaged Manakov equation to recover $\mathbf{A}(0,t)$ from the boundary condition $\mathbf{A}(z,t)$, using SSFM with fine step size (0.1 km). Since the signal propagation happens in the Manakov equation with attenuation and EDFAs, applying path-averaged DBP inevitably causes distortion on the recovered signal $\mathbf{A}(0,t)$ even in the single-channel scenario.

In single-channel scenario, if viewed as a compensation scheme, the combination of NFT, back-rotation in non-linear frequency domain, and INFT is somewhat equivalent to the path-averaged DBP. Therefore, we propose to use the residual distortion in a noiseless non-integrable system that is compensated by path-averaged DBP to estimate the distortion caused by channel mismatch. Fig. 6.10(a)(b) describe two *noiseless* systems with ideal and path-averaged DBP compensation, using the parameters in Tab. 6.2. These systems employ DP 32-QAM Nyquist signal with 56 GHz bandwidth. In general, any signal whose transceiver DSP causes no distortion suits the purpose. The residual distortion of the system (a) in Fig. 6.10 is considered equivalent to the distortion caused by channel mismatch in systems of Fig. 6.3. Fig. 6.10(c) shows that in the power range of our interest (≤ -0 dBm), the distortion caused by channel mismatch (lower curve in Fig. 6.10(c)) is negligible. We also extend this analogy to a multi-channel scenario. Systems described in Fig. 6.10 are simulated with various number of channels. The residual distortions of the channel-of-interest (COI) are measured in terms of SNR. The residual distortion can be then considered as a rough estimate of the ICI in NFDM networks. The results are shown in Fig. 6.10(d). The estimated ICI in NFDM is weaker than ICI in WDM with SC-DBP when N_{ch} is larger than 15.

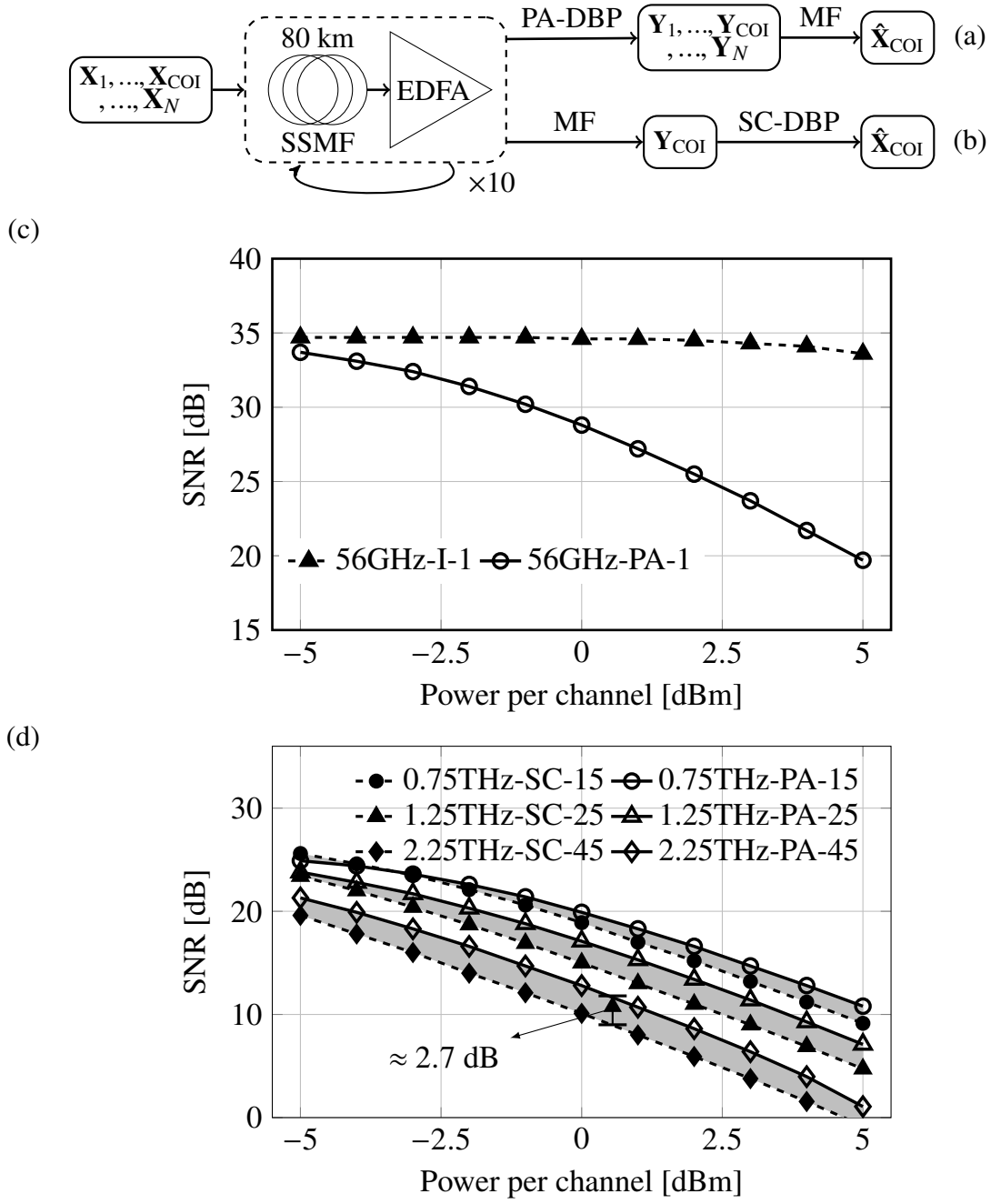

Figure 6.10: (a) (b) Simulation diagram of non-integrable models. PA-DBP stands for path-averaged DBP, IDBP for ideal DBP, COI for channel-of-interest. (c) residual distortion of IDBP (shortened in the legend to I, preceded by total bandwidth and followed by number of channels) and PA-DBP (shortened in the legend to PA) systems.

6.7 *b*-modulated DP-NFDM experiment

In the previous sections, the *b*-modulated DP-NFDM transmission system was simulated, and carefully optimised for high data rate. It was showed that the SE of *b*-modulated system achieved 1.2 bit/s/Hz higher SE than q_c-modulated system. In the remaining sections of the chapter, we implemented in experiment the DP-NFDM transmission system simulated in [186]. By developing efficient DSP, we could successfully modulate 196 non-linear sub-carriers in each polarisation, achieving a net data rate of 220 Gbit/s over 976 km of EDFA-based fibre link. To the best of our knowledge, this is the highest data rate of any NFDM systems in experiment up to date as shown in Fig. 6.1.

In the experiment, we transmit DP-NFDM signals of roughly 55 GHz (N_C/T_0) linear bandwidth. Transmitted symbols are randomly drawn from a 32-QAM constellation. The guard interval T_G between DP-NFDM symbols is estimated by $T_G \approx \pi W \beta_2 L_{\text{sp}} N_{\text{sp}} = 3.57$ ns. We set the burst duration $T_0 = T_G$ for simplicity (instead of $T_0 = 5T_G$ for maximum data rate as in Sec. 6.4). It results in $N_c = 196$ sub-carriers in the non-linear frequency domain. The system loses half of the SE due to the GI. The loss can be reduced by increasing T_0. We show one of the signal realisations in Fig. 6.11.

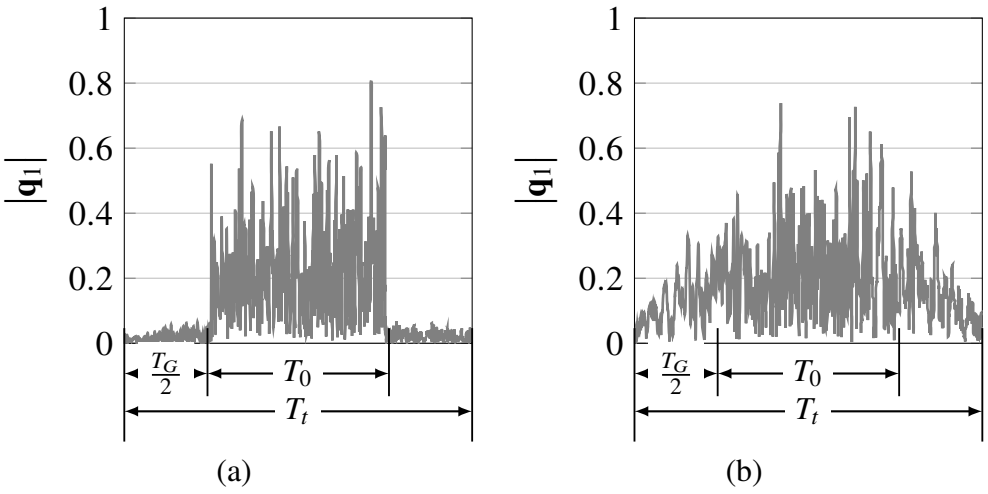

Figure 6.11: A realisation of INFT symbols (a) without PDC (b) with PDC.

6.7.1 Experimental setup and DSP

For the transmission experiment, at transmitter, we used four DAC and a PDM-IQM to generate the waveform. The output signal was subsequently amplified by an EDFA. Signals were then coupled in and out of the recirculating loop by a 50:50 coupler and travel three loops in our experiment. Each loop has three spans of SMF-28 (Corning) and three EDFAs, making the total distance of the experiment

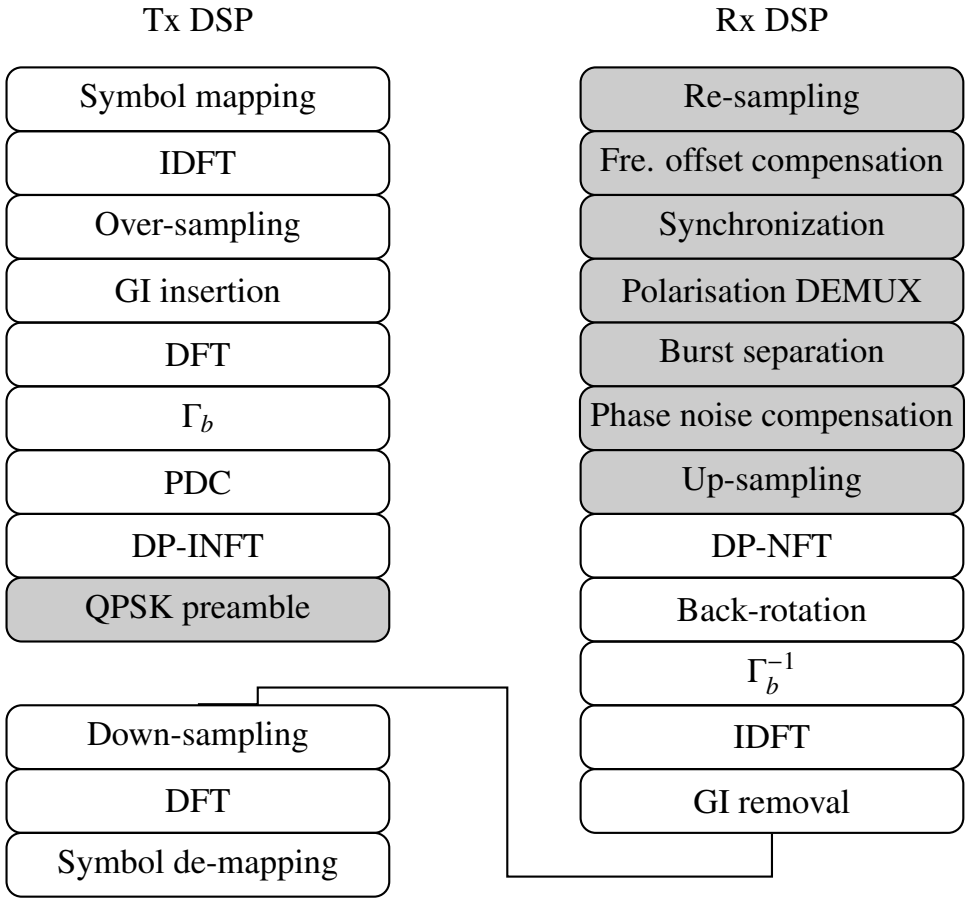

Figure 6.12: Transceiver DSP chain.

975.6 km. We chose the same distance as in [174] to allow a direct comparison. As receiver, preceded by another EDFA, an oscilloscope (DSOZ334A) with four channels was used. The sampled data taken by the oscilloscope were saved and processed offline. Two external cavity lasers were used as the optical carrier at the transmitter and as the local oscillator at the receiver. The wavelength-seletive switch (WSS) functioned as a optical band-pass filter to remove the out-of-band amplified spontaneous emission noise. Some important experimental parameters are listed in Tab. 6.3. The experimental setup is plotted in Fig. 6.14. In the following, we first go through the DSP blocks in grey box in Fig. 6.12.

Table 6.3: System parameters

fibre type	Corning SMF-28
DSP oversampling rate (Tx/Rx)	8
EDFA noise figure	5.5 dB
cavity laser line-width	100 kHz
DAC Effective number of bits	4.5 ~ 5

QPSK preamble

For synchronisation and polarisation-demultiplexing (POL-DEMUX) purposes, single-carrier quadrature phase-shift keying (QPSK) signals (55 Gbaud RRC pulses with roll-off factor of 0.5) are prepended to the DP-NFDM signals. The QPSK signals occupy 4.3% of the transmission period. The choice of the preamble signal was motivated by the existence of versatile adaptive filters for polarisation-demultiplexing (POL-DEMUX) and detection of QPSK signals. The resulted signal in one duty cycle is illustrated in Fig. 6.13. The NFDM training sequence is to remove any scaling or phase rotation in the non-linear frequency domain.

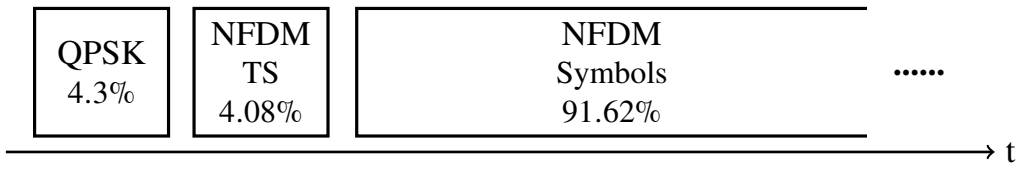

Figure 6.13: Signal structure in one duty cycle. TS stands for training symbols.

Frequency Offset Compensation

The frequency offset compensation was done with the help of the residual carrier. Details were described by (2.78) in Sec. 2.3.4.

Synchronisation

The QPSK preambles were used for synchronisation. The CD compensation (see Sec. 2.3.4) is first applied to the whole signal. By calculating the cross-correlation between the amplitude of the known QPSK 'pilot' and the amplitude of the CD-compensated signal, the beginning of the DP-NFDM signal can be easily located.

POL-DEMUX

The RLS multi-sample per symbol CMA [7, Sec. V-3)] was performed on the CD-compensated QPSK 'pilot' to estimate the 2×2 MIMO filter that tracks the arbitrary polarisation rotation and compensates the residual inter-symbol interference (ISI); see also Sec. 2.3.4 and Alg. 1. The obtained filter is then applied on the non-compensated DP-NFDM signals. Without phase noise compensation, a ring constellation is expected from the RLS-CMA.

Phase Noise Compensation

The phase noise is assumed to be perfectly tracked, and constant within one DP-NFDM symbol. The assumption is realised by comparing the phase of the theoretical waveform (from the noiseless simulation) with the received one, then average within one symbol period. It is equivalent to compensate in the non-linear frequency domain as $e^{j\phi}q(t) \leftrightarrow e^{-j\phi}\{q_{c1}(\lambda), q_{c2}(\lambda)\}$.

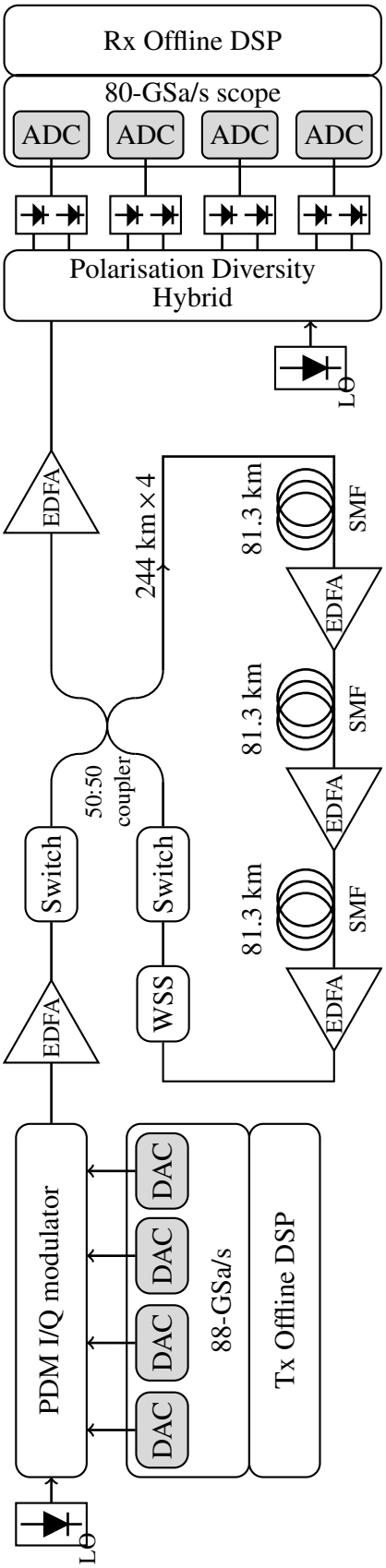

Figure 6.14: Recirculating fibre loop experiment setup. The fibre is Corning SMF-28.

6.7.2 Experimental results

To verify our transceiver DSP chain, we first carried out a back-to-back experiment. The waveforms generated by INFT for each launch power are different. In our back-to-back experiment, however, the transmitted signal power was fixed (set by the EDFA following IQM). We changed only the waveform, designed for a given launch power. To collect statistics, we collect at every designed launch power 5 traces of 512 NFDM symbols, each having 196 sub-carriers. The SNRs of the received signal can be calculated from the error vector magnitude (EVM). Fig. 6.15(a) shows the calculated SNR and their average in terms of the designed launch power at the receiver in the back-to-back scenario. It can be seen that, when we use the waveforms designed for launch powers larger than -3 dBm, the measured SNR starts to decrease. This is attributed to the reduced accuracy of INFT-NFT algorithm at high power, and DAC quantisation noise.

We perform the transmission experiment over 12 spans of 81.3 km SSMF. Waveforms were launched at their designed launch powers, and the polarisation state is also randomly changed at the transmitter. Fig. 6.15(b) shows the calculated SNRs and their average in terms of the launch power. To compute the SE, we estimate the MI (of individual sub-carrier detection) by assuming the channel conditional distribution as a Gaussian distribution [9]. The net data rate and SE are calculated from

$$\text{Net rate} = \frac{\text{MI} \cdot N_c \cdot 2}{(T_0 + T_G)\ \text{ns}}\ \text{(Gb/s)},\ \text{SE} = \frac{\text{Net rate}}{55\ \text{GHz}}\ \text{(bit/s/Hz)}. \tag{6.15}$$

This net data rate can be achieved by a soft-decision capacity-approaching binary code such as spatially coupled LDPC codes [187]. At the launch power of -1.3 dBm, the system reached its maximum net rate of 220 Gbit/s and SE of 4 bit/s/Hz. The corresponding constellation is also plotted in Fig. 6.15(b).

6.8 Summary

The research presented in this chapter extended b-modulated SP-NFDM scheme to DP system. Extensive simulations of the b-modulated DP-NFDM were carried out, and the results confirmed the viability and benefits of the b-modulation in the DP-NFDM system.

The q_c- and b-modulated DP-NFDM systems were compared in terms of Q-factor, correlation of sub-carriers, joint and individual entropy. The b-modulated DP-NFDM system shows 1 dB Q-factor improvement over q_c-modulated DP-NFDM system due to a weaker correlation of sub-carriers and less effective noise. Finally, the b-modulated system was optimised for higher data rate, achieving a

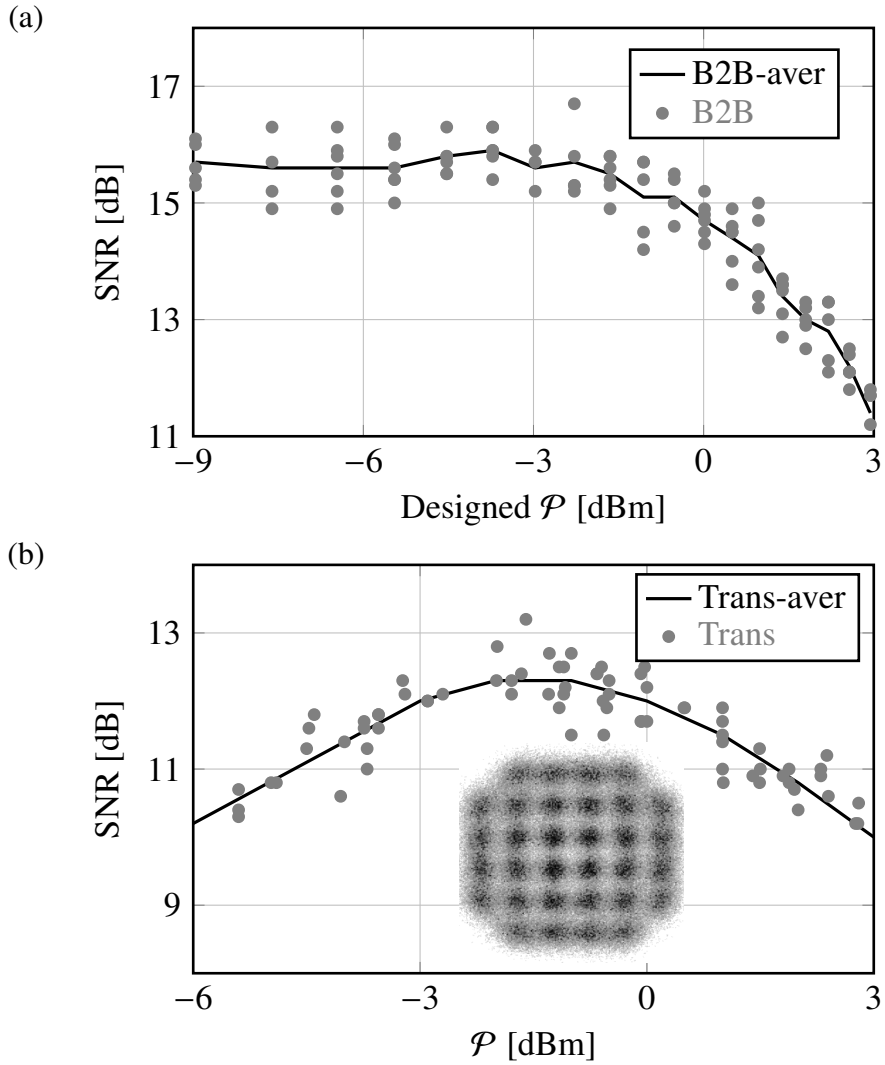

Figure 6.15: Measured SNRs and their average (a) in terms of the designed launch power in the back-to-back scenario, (b) in terms of the launch power over 12 spans of 81.3 km SMF. Below the curve is one received constellation at signal power of -1.3 dBm, Q factor = 3.7 dB, SE ≈ 2 bit/s/Hz/pol.

record net data rate of 400 Gbit/s (SE of 7.2 bit/s/Hz) over 12×80 km of SSMF, amplified by EDFA.

The system design developed in the aforementioned simulations was also verified in experiment. Extra DSP blocks were added to handle frequency offset compensation, synchronisation, POL-DEMUX and phase noise compensation through a QPSK 'pilot' preamble. The transmission experiment demonstrated a 220 Gbit/s *b*-modulated DP-NFDM system, achieving a SE of 4 bit/s/Hz. The data rate can potentially be improved by increasing the number of sub-carrier in each DP-NFDM symbol.

7

Thesis summary and future work

As mentioned at the beginning of the thesis, the main goal of the research described here was to improve SE of optical fibre channel with non-linear transmission schemes. And in particular, two questions are of interest, 1) whether it is possible to achieve a monotonic increasing SE in the ORN, in contrast to the saturating behaviour shown in Fig. 1.7, and 2) if such a transmission scheme exists, whether it is viable in practice.

The research results in the thesis presented negative evidence to these two questions, but cannot provide an absolute answer as such answers can only be given by mathematical proofs.

The study of spectral broadening in Chapter 3 showed that, for a Gaussian process and specific types of receivers, the upper bound on the SE of the optical fibre channel increases monotonically with signal launch power only if the signal power and receiver bandwidth increase at the same time, and the signal power grows at a slightly faster rate than the signal bandwidth does. Note that the spectral broadening effect is significant only in the highly non-linear regime, i.e., at powers of roughly 6 dB higher than the optimal launch power and beyond in a multi-channel transmission link. Although no practical system operates at such high powers, the study remains valuable for information-theoretic analysis of the optical fibre channel capacity over all range of powers.

In Chapter 4, 5 and 6, the focus of research shifts to the study of NFT-based

transmission schemes as they can, in theory, achieve a monotonic increasing AIR in ORN. Initially, we demonstrated that in the simplest scenario, i.e., lossless fibre and one symbol per channel use, NFDM outperformed WDM and showed a non-saturating AIR over the simulated power range. Although these simulation settings are unrealistic, they single out the important channel effects such as inter-channel interference and channel mismatch due to non-integrability. The results can be interpreted that if distortion-free INFT-NFT algorithms were available, an optical network operating on NFDM would be able to outperform current WDM systems in terms of data rate. In the subsequent simulations with multiple symbols per channel use, the NFDM system did not show any benefit over conventional WDM system. This was attributed to the inefficient modulation method used to generate the NFT-signal.

To tackle the inefficient modulation method, we adopted the dual polarisation (DP) NFT-INFT and the so-called *b*-modulation, and successfully increased the AIR and SE of an NFT-based transmission system over 12 spans of 80 km optical fibre link amplified by EDFAs to 400 Gbit/s and 7.2 bit/s/Hz in simulations and 220 Gbit/s and 4 bit/s/Hz in experiments. To put the results into perspective, one should look at Fig. 1.7 and Fig. 6.1, and see that although the data rate of the proposed NFT-based system is the highest among all existing NFT-based systems, it is still below the AIR of conventional WDM system with single-channel DBP.

The AIR of current NFT-based transmission schemes is limited mainly for two reasons. Firstly, the transceiver distortion introduced by the inaccuracy of INFT-NFT algorithms is much larger than conventional DSP chains at high signal powers, despite that the INFT-NFT algorithm already operates at a significant higher oversampling rate than conventional DSPs. This is similar to one of the difficulties faced in multi-mode fibre communication systems where high-complexity DSP schemes are needed. Secondly, based on my estimation, the distortion introduced by channel-mismatch is also severe. Currently, there is no DSP technique that can cancel the channel mismatch and the problem appears to be only solvable through installation of new optical fibre infrastructure. Similar requirement is encountered in any spatial-division multiplexing scheme.

Admittedly, NFT-based transmission schemes are still facing great challenges in many aspects, and the possibility of their implementation in commercial systems is very much debatable. However, NFT-based transmission is also a relatively young research topic and requires more research until it is fully understood.

For further investigations into the NFT-based transmission schemes, the following directions have been identified:

Impact of distributed Raman amplifier on NFT-based transmission schemes

It has been shown in Sec. 6.6.2 that the distortion introduced by channel mismatch due to non-integrability in NFDM is larger than the WDM inter-channel interference in an EDFA-amplified optical fibre link when the number of channels is lower than 15. Therefore, alternative amplification methods that are closer to the ideal lossless optical fibre and yet still practically deployable should be considered in the future research, such as the distributed Raman amplification.

The path-average (PA) model for optical fibre links with distributed Raman amplifiers was examined in [156]. With this PA model and the b-modulated DP-NFT scheme proposed in this thesis, it would be interesting to see, in simulation and experiment, whether higher data rate could be achieved. In addition, one can use the method developed in Sec. 6.6.2 to estimate the distortion from channel mismatch in large-scale NFDM networks.

Develop of INFT-NFT algorithms of high accuracy

The most critical building block of a NFT-based transmission system is the INFT-NFT algorithm. As shown in Fig. 6.9, the INFT-NFT algorithm used in this thesis introduced significant distortion at high signal powers in spite of the high oversampling rate. It is desirable to have an INFT-NFT algorithm that can operate at moderate oversampling rate and still provide acceptable accuracy at high signal powers. There has been progress in the development for accurate and fast NFT algorithms with high-order integrator [188], but none for the INFT. Developing a low-complex and accurate INFT-NFT algorithm would be the most important step for the advancement of NFT-based transmission systems.

Drawing on the history of Fourier transform, we see that it is actually the discrete Fourier transform of a discrete signal that are used in all signal analysis, not the discretised calculation of the original Fourier transform in the form of an integral. The assumption that the signal is periodic and the Nyquist sampling theorem guarantee that the original continuous signal is fully equivalent to the discrete version. Perhaps a non-linear Nyquist sampling theorem is needed for further development of efficient INFT-NFT algorithms. In retrospect, INFT-NFT successfully transfers the non-linearity in the optical fibre channel to the transceiver, albeit not making the treatment of non-linearity any easier. Instead of developing a non-linear channel model for linearly-generated signals, one can focus on the non-linear signal generation process. To this end, the related questions could be: What are the properties of the non-linear-band limited signal in time domain? How to sufficiently represent such a signal with discrete samples? Is the non-linear spectrum of an AWGN still an AWGN?

Optical add-drop multiplexer in the non-linear frequency domain

One important optical component of a WDM optically routed network is the

optical add-drop multiplexing (OADM). The device adds or drops channels in certain wavelength without optoelectronic conversion. Similarly, to build a NFDM network in scale, one would naturally need an OADM that adds and drops channels in certain non-linear frequency band without optoelectronic conversion. However, [189] is probably the only work that addresses the optical signal processing in the non-linear frequency domain. Although the proof-of-concept system is encouraging, it only operates on the discrete spectrum. It is still far away from an functional non-linear OADM.

The development of a non-linear OADM would be an incredibly innovative and difficult task. It would also enable us to potentially move the digital NFT signal processing completely to the optical domain. Therefore, research in this direction should be highly encouraged.

A

Probability distribution of a random variable

In this chapter, we explain the definition of the probability distribution of a random variable in both discrete and continuous cases and some related concepts. The main reference used here is the book [190]. For a rigorous description of probability theory, one should start with measure theory [191]. But it does not bring any benefit to the development of ideas in the thesis.

A.1 Probability distribution of a discrete random variable

A random variable Y is said to be discrete if it can assume only a finite or countably infinite number of distinct values. The probability that Y takes on the value y, often denoted as $P(Y = y)$, is defined as the sum of the probabilities of all sample points that are assigned the value y.

The probability distribution for a discrete random variable Y can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.

For any discrete probability distribution, the following must be true:

- $0 \leq p(y) \leq 1$ for all y.

- $\sum_y p(y) = 1$, where the summation is over all values of y with non-zero probability.

Then, the *expected value* of such a discrete random variable Y, often denoted as $\mathbb{E}[Y]$, is defined to be

$$\mathbb{E}[Y] = \sum_y yp(y). \tag{A.1}$$

A.2 Probability distribution of a continuous random variable

Before we can state a formal definition of a continuous random variable, we must define the the *cumulative distribution function* of any random variable.

Let Y denote any random variable. The cumulative distribution function of Y, denoted by $F(y)$, is such that $F(y) = P(Y \leq y)$ for $-\infty < y < \infty$. If $F(y)$ is a cumulative distribution function, then the following must be true:

- $F(-\infty) = \lim_{y \to -\infty} F(y) = 0$.
- $F(\infty) = \lim_{y \to \infty} F(y) = 1$.
- $F(y)$ is a non-decreasing function of y.

We can now formally define a continuous random variable. A random variable Y with the cumulative distribution function $F(y)$ is said to be *continuous* if $F(y)$ is continuous, for $-\infty < y < \infty$. For a continuous random variable, the *probability density function* is defined as

$$f(y) = \frac{dF(y)}{dy} = F'(y), \tag{A.2}$$

wherever the derivative exists. For such a density function of a continuous random variable, the following are true:

- $0 \leq f(y)$ for all y, $-\infty < y < \infty$.
- $\int_{-\infty}^{\infty} f(y)dy = 1$.

The *expected value* of a continuous random variable can be defined as

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} yf(y)dy, \tag{A.3}$$

provided that the integral exists.

B

Power spectral density

Let $y(t)$ be a complex-valued zero-mean stochastic process. The autocorrelation function of $y(t)$ is defined as

$$R(t_1, t_2) = \mathbb{E}\big[y(t_1)y^*(t_2)\big]. \tag{B.1}$$

If a stochastic process $y(t)$ is strongly stationary, then

$$R(t) = \mathbb{E}\big[y(t_0)y^*(t_0 - t)\big], \tag{B.2}$$

for any reference point t_0.

The **first definition** of power spectral density (PSD) $S(\omega)$ can be stated as the Fourier transform of the autocorrelation function [192]

$$S(\omega) = \int_{-\infty}^{\infty} R(t)e^{-i\omega t}dt. \tag{B.3}$$

The **second definition** of PSD starts with a truncated continuous-time stochastic process

$$y_T(t) = y(t)\,\mathrm{rect}(t/T) = \begin{cases} y(t), |t| \leq T/2, \\ 0, |t| > T/2. \end{cases} \tag{B.4}$$

The Fourier transform is

$$\mathcal{F}(y_T(t)) = \int_{-\infty}^{\infty} y_T(t) e^{-j2\pi ft} dt, \;\; T < \infty. \tag{B.5}$$

Using Parseval's theorem and dividing both side by T,

$$\frac{1}{T}\int_{-T/2}^{T/2} |y_T(t)|^2 dt = \frac{1}{T}\int_{-\infty}^{\infty} |\mathcal{F}(y_T(t))|^2 df \tag{B.6}$$

If we let T approach infinity, the left side becomes the average power over all time. On the right side, the Fourier transform is not defined in that limit. But it can be shown that even though the Fourier transform does not exist, its expected value does. Then

$$\mathbb{E}\left[\frac{1}{T}\int_{-T/2}^{T/2} y_T^2(t) dt\right] = \mathbb{E}\left[\frac{1}{T}\int_{-\infty}^{\infty} |\mathcal{F}(y_T(t))|^2 df\right], \tag{B.7}$$

taking the limit as T approaches infinity,

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\mathbb{E}\left[y_T^2(t)\right]dt &= \lim_{T\to\infty}\frac{1}{T}\int_{-\infty}^{\infty}\mathbb{E}\left[|\mathcal{F}(y_T(t))|^2\right]df, \\ &= \int_{-\infty}^{\infty}\lim_{T\to\infty}\mathbb{E}\left[\frac{|\mathcal{F}(y_T(t))|^2}{T}\right]df. \end{aligned} \tag{B.8}$$

The integrand on the right side is identified as PSD

$$S(f) = \lim_{T\to\infty}\mathbb{E}\left[\frac{|\mathcal{F}(x_T(t))|^2}{T}\right]. \tag{B.9}$$

The two definitions (B.3) and (B.9) are equivalent under the mild assumption that the autocorrelation function $R(t)$ decays sufficiently fast, so that

$$\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}|t||R(t)| = 0. \tag{B.10}$$

C

The Gaussian Noise model

Recall the first-order perturbation solution of NLSE in the frequency domain derived earlier in (2.31)

$$\tilde{A}^{(1)}(z,\omega) = e^{(-\frac{\alpha}{2}+j\frac{\beta_2\omega^2}{2})z}\Bigg\{\tilde{A}(0,\omega)+ j\gamma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(\Omega_{123\omega},z)\tilde{A}(0,\omega_1)\tilde{A}(0,\omega_2)\tilde{A}^*(0,\omega_3)\delta_{123\omega}d\omega_{123}\Bigg\}, \quad \text{(C.1)}$$

where

$$H(\Omega_{123\omega},z) = \frac{e^{-z\alpha+jz\Omega_{123\omega}}-1}{-\alpha+j\Omega_{123\omega}}, \quad \text{(C.2)}$$

$$\Omega_{123\omega} = \frac{\beta_2}{2}(\omega_1^2+\omega_2^2-\omega_3^2-\omega^2). \quad \text{(C.3)}$$

We denote the PSD as $S(z,\omega)$ and calculate it according to its second definition with the Gaussian assumption on the input signal $\tilde{A}^{(0)}(0,\omega)$ [56, Eq. (27)]. To save

space, we replace $\tilde{A}^{(0)}(0,\omega)$ with $\tilde{A}(\omega)$ in the following derivation

$$
\begin{aligned}
&S(z,\omega)^{GN} = \mathbb{E}\left[\tilde{A}^{(1)}(z,\omega)\tilde{A}^{(1)*}(z,\omega)\right] \\
&= e^{-\alpha z}\mathbb{E}\left[|\tilde{A}(\omega)|^2\right] + \\
&\quad 2e^{-\alpha z}\gamma \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{R}\left[jH(\Omega_{123\omega}, z)\mathbb{E}\left[\tilde{A}(\omega_1)\tilde{A}(\omega_2)\tilde{A}^*(\omega_3)\tilde{A}^*(\omega)\right]\delta_{123\omega}\right] d\omega_{123} + \\
&\quad e^{-\alpha z}\gamma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |H(\Omega_{123\omega}, z)|^2 \times \\
&\quad \mathbb{E}\left[\tilde{A}(\omega_1)\tilde{A}(\omega_2)\tilde{A}^*(\omega_3)\tilde{A}^*(\omega_1')\tilde{A}^*(\omega_2')\tilde{A}(\omega_3')\right]\delta_{123\omega}\delta_{1'2'3'\omega} d\omega_{1231'2'3'}. \qquad \text{(C.4)}
\end{aligned}
$$

With the Gaussian assumption on the input spectrum components, the 4-point moments break down to

$$
\mathbb{E}\left[\tilde{A}(\omega_1)\tilde{A}(\omega_2)\tilde{A}^*(\omega_3)\tilde{A}^*(\omega)\right]\delta_{123\omega} = \mathbb{E}\left[|\tilde{A}(\omega_1)|^2\right]\mathbb{E}\left[|\tilde{A}(\omega_2)|^2\right](\delta_{13}\delta_{2\omega} + \delta_{1\omega}\delta_{23}). \qquad \text{(C.5)}
$$

The second term on the r.h.s of (C.4) is supported on the trivial subset ($\omega_1 = \omega_3,\ \omega_2 = \omega$) and ($\omega_1 = \omega,\ \omega_2 = \omega_3$), where $jH(\Omega_{123\omega}, z)$ is purely imaginary. Hence, it should be zero.

Furthermore, due to the Gaussian assumption, the 6-point moments break down to

$$
\begin{aligned}
&\mathbb{E}\left[\tilde{A}(\omega_1)\tilde{A}(\omega_2)\tilde{A}^*(\omega_3)\tilde{A}^*(\omega_1')\tilde{A}^*(\omega_2')\tilde{A}(\omega_3')\right]\delta_{123\omega}\delta_{1'2'3'\omega} \\
&= \mathbb{E}\left[|\tilde{A}(\omega_1)|^2\right]\mathbb{E}\left[|\tilde{A}(\omega_2)|^2\right]\mathbb{E}\left[|\tilde{A}(\omega_3)|^2\right](\delta_{11'\delta_{22'}} + \delta_{12'} + \delta_{21'})\delta_{33'}. \qquad \text{(C.6)}
\end{aligned}
$$

Substituting (C.6) into (C.4), we obtain

$$
\begin{aligned}
S(z,\omega)^{GN} &= e^{-\alpha z}S(0,\omega) + \\
&\quad 2e^{-\alpha z}\gamma^2 \int |H(\Omega_{123\omega}, z)|^2 S(0,\omega_1)S(0,\omega_2)S(0,\omega_3)\delta_{123\omega} d\omega_{123}. \qquad \text{(C.7)}
\end{aligned}
$$

Bibliography

[1] P. J. Winzer and D. T. Neilson, "From scaling disparities to integrated parallelism: A decathlon for a decade," J. Lightw. Technol., vol. 35, pp. 1099–1115, Mar. 2017.

[2] P. S. Andre, A. N. Pinto, J. L. Pinto, T. Almeida, and M. Pousa, "Tunable transparent and cost effective optical add-drop multiplexer based on fiber Bragg grating for DWDM networks," in 2001 Digest of LEOS Summer Topical Meetings: Advanced Semiconductor Lasers and Applications/Ultraviolet and Blue Lasers and Their Applications/Ultralong Haul DWDM Transmission and Networking/WDM Compo, (Copper Mountain, CO, USA), Aug. 2001.

[3] A. V. Tran, C.-J. Chae, and R. S. Tucker, "Bidirectional optical add-drop multiplexer based on Bragg gratings and circulators," Appl. Opt., vol. 42, pp. 3495–3499, Jun. 2003.

[4] G. Wild, "Optical add-drop multiplexer, based on a fiber Bragg grating (FBG), and two circulators.," 2007. `https://en.wikipedia.org/wiki/File:Fbg_oadm.GIF`, accessed on 12th May 2020.

[5] R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Lightw. Technol., vol. 28, pp. 662–701, Feb. 2010.

[6] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," J. Lightw. Technol., vol. 28, pp. 939–951, Mar. 2010.

[7] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," J. Lightw. Technol., vol. 35, pp. 1125–1141, Mar. 2017.

[8] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," J. Select. Topics Quantum Electron., vol. 16, pp. 1164–1179, Oct. 2010.

[9] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable information rates for fiber optics: Applications and computations," J. Lightw. Technol., vol. 36, pp. 424–439, Jan. 2018.

[10] I. Tavakkolnia and M. Safari, "Dispersion pre-compensation for NFT-based optical fiber communication systems," in Proc. of Conf. Laser and Electro-Optics, (San Jose, CA, USA), pp. 4–5, Jun. 2016.

[11] D. Lombardi and S. Pant, "Nonparametric k-nearest-neighbor entropy estimator," Phys. Rev. E, vol. 93, p. 013310, Jan. 2016.

[12] International Telecommunication Union, Optical fibres, cables and systems. 2010. `https://www.itu-ilibrary.org/content/publication/pub-8035a9bb-en`, accessed on the 11th May 2020.

[13] Corning Inc., "The evolution of optical fiber," 2017. `https://www.corning.com/media/worldwide/global/documents/evolution_of_optical_fiber.pdf`, accessed on the 11th May 2020.

[14] K. C. Kao and G. A. Hockham, "Dielectric-fibre surface waveguides for optical frequencies," Proceedings of the Institution of Electrical Engineers, vol. 113, pp. 1151–1158, Jul. 1966.

[15] OFS Fitel LLC, "Allwave optical fiber - zero water peak," 2017. `https://www.ofsoptics.com/wp-content/uploads/AllWave-117-web-7.pdf`, accessed on the 11th May 2020.

[16] R. J. Mears, L. Reekie, I. M. Jauncey, and D. N. Payne, "Low-noise erbium-doped fibre amplifier operating at 1.54 μm," Electron. Lett., vol. 23, pp. 1026–1028, Sep. 1987.

[17] E. Desurvire, J. R. Simpson, and P. C. Becker, "High-gain erbium-doped traveling-wave fiber amplifier," Opt. Lett., vol. 12, pp. 888–890, Nov. 1987.

[18] S. B. Alexander, "Wavelength division multiplexed optical communication systems employing uniform gain optical amplifiers," 1995. United States, US5696615A.

[19] A. R. Chraplyvy, J. A. Nagel, and R. W. Tkach, "Equalization in amplified WDM lightwave transmission systems," Photon. Technol. Lett., vol. 4, pp. 920–922, Aug. 1992.

[20] A. R. Chraplyvy, R. W. Tkach, K. C. Reichmann, P. D. Magill, and J. A. Nagel, "End-to-end equalization experiments in amplified WDM lightwave systems," Photon. Technol. Lett., vol. 5, pp. 428–429, Apr. 1993.

[21] Fujitsu Ltd., "White paper: Next generation roadm architecture & benefits," Mar 2015. `https://www.fujitsu.com/us/Images/Fujitsu-NG-ROADM.pdf`, accessed on the 12th May 2020.

[22] R. A. Linke and A. H. Gnauck, "High-capacity coherent lightwave systems," J. Lightw. Technol., vol. 6, pp. 1750–1769, Nov. 1988.

[23] A. R. Chraplyvy, A. H. Gnauck, R. W. Tkach, and R. M. Derosier, "8 × 10 Gb/s transmission through 280 km of dispersion-managed fiber," Photon. Technol. Lett., vol. 5, pp. 1233–1235, Oct. 1993.

[24] A. H. Gnauck, R. W. Tkach, A. R. Chraplyvy, and T. Li, "High-capacity optical transmission systems," J. Lightw. Technol., vol. 26, pp. 1032–1045, Jun. 2008.

[25] M. S. Erkılınç, D. Lavery, K. Shi, B. C. Thomsen, R. I. Killey, S. J. Savory, and P. Bayvel, "Comparison of low complexity coherent receivers for UDWDM-PONs (λ-to-the-user)," J. Lightw. Technol., vol. 36, pp. 3453–3464, Aug. 2018.

[26] A. Mecozzi, C. Antonelli, and M. Shtaif, "Kramers–Kronig coherent receiver," Optica, vol. 3, pp. 1220–1227, Nov. 2016.

[27] Z. Li, M. S. Erkılınç, K. Shi, E. Sillekens, L. Galdino, B. C. Thomsen, P. Bayvel, and R. I. Killey, "SSBI mitigation and the Kramers–Kronig scheme in single-sideband direct-detection transmission with receiver-based electronic dispersion compensation," J. Lightw. Technol., vol. 35, pp. 1887–1893, May 2017.

[28] A. Splett, C. Kurtzke, and K. Petermann, "Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Montreux, Switzerland), MoC2.4, Sep. 1993.

[29] 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, pp. 694–721, Feb. 2014.

[30] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," Opt. Express, vol. 22, pp. 16335–16362, Jun. 2014.

[31] L. Yan, E. Agrell, H. Wymeersch, and M. Brandt-Pearce, "Resource allocation for flexible-grid optical networks with nonlinear channel model [invited]," J. Opt. Commun. Netw., vol. 7, no. 11, pp. B101–B108, 2015.

[32] Y. Yu, L. Jin, Z. Xiao, F. Yu, Y. Lu, L. Liu, W. Wu, and L. Li, "100.5Tb/s MLC-CS-256QAM transmission over 600-km single mode fiber with C+L band EDFA," in Asia Comm. Photo. Conf. (ACP), (Hangzhou, China), 2018.

[33] A. Lord, A. Soppera, and A. Jacquet, "The impact of capacity growth in national telecommunications networks," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, p. 20140431, Mar. 2016.

[34] R. W. Tkach, "Scaling optical communications for the next decade and beyond," Bell Labs Technical Journal, vol. 14, pp. 3–9, Winter 2010.

[35] "Metro network traffic growth: An architecture impact study," Bell Labs Strategic White Paper, 2013. https://www.tmcnet.com/tmc/whitepapers/documents/whitepapers/2013/9378-bell-labs-metro-network-traffic-growth-an-architecture.pdf, accessed on the 5th June 2020.

[36] A. Mecozzi and R. Essiambre, "Nonlinear Shannon limit in pseudolinear coherent systems," J. Lightw. Technol., vol. 30, pp. 2011–2024, Apr. 2012.

[37] M. Secondini and E. Forestieri, "Scope and limitations of the nonlinear Shannon limit," J. Lightw. Technol., vol. 35, pp. 893–902, Feb. 2017.

[38] C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948.

[39] S. Yoshima, Z. Liu, Y. Sun, K. R. H. Bottrill, F. Parmigiani, P. Petropoulos, and D. J. Richardson, "Nonlinearity mitigation for multi-channel 64-QAM signals in a deployed fiber link through optical phase conjugation," in Proc. of

Opt. Fiber Commun. Conf. and Exhib. (OFC), (Anaheim, CA, USA), Th4F.4, Mar. 2016.

[40] X. Yangzhang, M. I. Yousefi, A. Alvarado, D. Lavery, and P. Bayvel, "Non-linear frequency-division multiplexing in the focusing regime," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Los Angeles, CA, USA), Tu3D.1, Mar. 2017.

[41] J. Sahu, "New fibre-doped amplifier technologies for extended transmission bands," in Invited talk at Beyond state-of-the-art optical communications, (London, UK), Jun. 2017.

[42] M. A. Melkumov, S. V. Alyshev, S. V. Firstov, and E. M. Dianov, "Bismuth-doped fiber lasers and amplifiers: Review and prospects," in Int. Conf. Laser Optics (LO), (St. Petersburg, Russia), Jun. 2016.

[43] J. K. Sahu, N. K. Thipparapu, A. A. Umnikov, and P. Barua, "Recent advances in Bi-doped fiber lasers and amplifiers," in Proc. of Int. Symp. on SiO2, Advanced dielectrics and Related Devices, (Nice, France), Jun. 2016.

[44] E. Agrell, M. Karlsson, A. R. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton, M. Secondini, F. R. Kschischang, A. Lord, J. Prat, I. Tomkos, J. E. Bowers, S. Srinivasan, M. Brandt-Pearce, and N. Gisin, "Roadmap of optical communications," J. Opt., vol. 18, p. 063002, May 2016.

[45] T. Kobayashi, M. Nakamura, F. Hamaoka, K. Shibahara, T. Mizuno, A. Sano, H. Kawakami, A. Isoda, M. Nagatani, H. Yamazaki, Y. Miyamoto, Y. Amma, Y. Sasaki, K. Takenaga, K. Aikawa, K. Saitoh, Y. Jung, D. J. Richardson, K. Pulverer, M. Bohn, M. Nooruzzaman, and T. Morioka, "1-Pb/s (32 SDM/46 WDM/768 Gb/s) C-band dense SDM transmission over 205.6-km of single-mode heterogeneous multi-core fiber using 96-Gbaud PDM-16QAM channels," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Los Angeles, CA, USA), Th5B.1, Mar. 2017.

[46] K. Igarashi, T. Tsuritani, I. Morita, Y. Tsuchida, K. Maeda, M. Tadakuma, T. Saito, K. Watanabe, K. Imamura, R. Sugizaki, and M. Suzuki, "1.03-Exabit/skm super-Nyquist-WDM transmission over 7,326-km seven-core fiber," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (London, UK), PD3.E.3, Sep. 2013.

[47] N. K. Fontaine, R. Ryf, Haoshuo Chen, A. Velazquez Benitez, J. E. Antonio Lopez, R. Amezcua Correa, Binbin Guan, B. Ercan, R. P. Scott, S. J. Ben Yoo, L. Grüner-Nielsen, Yi Sun, and R. J. Lingle, "30×30 MIMO transmission over 15 spatial modes," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Los Angeles, CA, USA), Th5C.1, Mar. 2015.

[48] E. Ip, M. Li, K. Bennett, Y. Huang, A. Tanaka, A. Korolev, K. Koreshkov, W. Wood, E. Mateo, J. Hu, and Y. Yano, "146 $\lambda \times 6 \times 19$-Gbaud wavelength- and mode-division multiplexed transmission over 10×50-km spans of few-mode fiber with a gain-equalized few-mode EDFA," J. Lightw. Technol., vol. 32, pp. 790–797, Feb. 2014.

[49] S. Mumtaz, R. J. Essiambre, and G. P. Agrawal, "Nonlinear propagation in multimode and multicore fibers: Generalization of the Manakov equations," J. Lightw. Technol., vol. 31, pp. 398–406, Feb. 2013.

[50] G. Rademacher and K. Petermann, "Nonlinear Gaussian noise model for multimode fibers with space-division multiplexing," J. Lightw. Technol., vol. 34, pp. 2280–2287, May 2016.

[51] G. Agrawal, Fiber-optic Communication Systems. Wiley Series in Microwave and Optical Engineering, Wiley, Fourth ed., 2012.

[52] W. Kester, "Taking the mystery out of the infamous formula, "SNR = 6.02N + 1.76dB," and why you should care," 2009. `https://www.analog.com/media/en/training-seminars/tutorials/MT-001.pdf`, accessed on 14th May 2020.

[53] I. Fatadin, D. Ives, and M. Wicks, "Numerical simulation of intensity and phase noise from extracted parameters for CW DFB lasers," J. Quantum Electron., vol. 42, pp. 934–941, Sep. 2006.

[54] C. R. S. Fludger and T. Kupfer, "Transmitter impairment mitigation and monitoring for high baud-rate, high order modulation systems," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Düsseldorf, Germany), Tu.2.A.2 , Sep. 2016.

[55] G. Agrawal, Lightwave Technology: Telecommunication Systems. Wiley, Jul. 2005.

[56] M. I. Yousefi, "The Kolmogorov–Zakharov model for optical fiber communication," Trans. Inf. Theory, vol. 63, pp. 377–391, Jan. 2017.

[57] S. J. Farlow, Partial Differential Equations for Scientists and Engineers. Dover Books on Mathematics, Newburyport: Dover Publications, 2012.

[58] P. Poggiolini, A. Carena, V. Curri, G. Bosco, and F. Forghieri, "Analytical modeling of nonlinear propagation in uncompensated optical transmission links," Photon. Technol. Lett., vol. 23, pp. 742–744, Jun. 2011.

[59] V. Zakharov, V. L'vov, and G. Falkovich, Kolmogorov Spectra of Turbulence I. Springer series in nonlinear dynamics, Springer-Verlag, 1992.

[60] S. Chibbaro, F. De Lillo, and M. Onorato, "Weak versus strong wave turbulence in the Majda-McLaughlin-Tabak model," Phys. Rev. Fluids, vol. 2, p. 052603, May 2017.

[61] J. P. Gordon and H. Kogelnik, "PMD fundamentals: Polarization mode dispersion in optical fibers," Proc. of the National Academy of Sciences, vol. 97, pp. 4541–4550, Apr. 2000.

[62] C. B. Czegledi, Modeling and Compensation of Polarization Effects in Fiber-optic Communications Systems. Chalmers University of Technology, 2018.

[63] G. J. Foschini and C. D. Poole, "Statistical theory of polarization dispersion in single mode fibers," J. Lightw. Technol., vol. 9, pp. 1439–1456, Nov. 1991.

[64] C. D. Poole and D. L. Favin, "Polarization-mode dispersion measurements based on transmission spectra through a polarizer," J. Lightw. Technol., vol. 12, pp. 917–929, Jun. 1994.

[65] A. Papoulis, Probability, random variables, and stochastic processes / Athanasios Papoulis, S. Unnikrishna Pillai. McGraw-Hill series in electrical and computer engineering, fourth edition, international edition. ed., 2014.

[66] 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, pp. 1735–1745, Sep. 1997.

[67] G. P. Agrawal, Nonlinear Fiber Optics. Boston: Academic Press, Fifth ed., 2013.

[68] S. Sternberg, Lie Algebras. Harvard University, 2004. `http://people.math.harvard.edu/~shlomo/docs/lie_algebras.pdf`, accessed on the 16th May 2020.

[69] 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," Photon. Technol. Lett., vol. 12, pp. 489–491, May 2000.

[70] H. Sun, K.-T. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," Opt. Express, vol. 16, pp. 873–879, Jan. 2008.

[71] K. Roberts, S. H. Foo, M. Moyer, M. Hubbard, A. Sinclair, J. Gaudette, and C. Laperle, "High capacity transport—100G and beyond," J. Lightw. Technol., vol. 33, pp. 563–578, Feb. 2015.

[72] I. Fatadin, S. J. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical QPSK coherent receiver," Photon. Technol. Lett., vol. 20, pp. 1733–1735, Oct. 2008.

[73] R. Dar and P. J. Winzer, "On the limits of digital back-propagation in fully loaded WDM systems," Photon. Technol. Lett., vol. 28, pp. 1253–1256, Jun. 2016.

[74] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," J. Lightw. Technol., vol. 29, pp. 2570–2576, Sep. 2011.

[75] Y. Gao, J. C. Cartledge, A. S. Karar, and S. S. . Yam, "Reducing the complexity of nonlinearity pre-compensation using symmetric EDC and pulse shaping," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (London, UK), PD3.E.5, Sep. 2013.

[76] F. P. Guiomar, J. D. Reis, A. L. Teixeira, and A. N. Pinto, "Mitigation of intrachannel nonlinearities using a frequency-domain Volterra series equalizer," Opt. Express, vol. 20, pp. 1360–1369, Jan. 2012.

[77] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part I: Mathematical tools," Trans. Inf. Theory, vol. 60, pp. 4312–4328, Jul. 2014.

[78] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part II: Numerical methods," Trans. Inf. Theory, vol. 60, pp. 4329–4345, Jul. 2014.

[79] M. I. Yousefi and F. R. Kschischang, "Information transmission using the nonlinear Fourier transform, part III: Spectrum modulation," Trans. Inf. Theory, vol. 60, pp. 4346–4369, Jul. 2014.

[80] D. Wang, M. Zhang, M. Fu, Z. Cai, Z. Li, H. Han, Y. Cui, and B. Luo, "Nonlinearity mitigation using a machine learning detector based on k-nearest neighbors," Photon. Technol. Lett., vol. 28, pp. 2102–2105, Oct. 2016.

[81] Y. Cai, D. G. Foursa, C. R. Davidson, J. . Cai, O. Sinkin, M. Nissov, and A. Pilipetskii, "Experimental demonstration of coherent MAP detection for nonlinearity mitigation in long-haul transmissions," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (San Diego, CA, USA), OTuE1, Mar. 2010.

[82] S. Haykin, Adaptive Filter Theory. Prentice-Hall information and system sciences series, Prentice Hall, Second ed., 1991.

[83] Yuxin Chen, T. Le-Ngoc, B. Champagne, and Changjiang Xu, "Recursive least squares constant modulus algorithm for blind adaptive array," Trans. Signal Processing, vol. 52, pp. 1452–1456, May 2004.

[84] F. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers," Trans. Commun., vol. 34, pp. 423–429, May 1986.

[85] Meng Yan, Zhenning Tao, Liang Dou, Lei Li, Ying Zhao, T. Hoshida, and J. C. Rasmussen, "Digital clock recovery algorithm for Nyquist signal," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Anaheim, CA, USA), OTu2I.7, Mar. 2013.

[86] Kuang-Tsan Wu and H. Sun, "Frequency-domain clock phase detector for Nyquist WDM systems," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (San Francisco, CA, USA), Th3E.2, Mar. 2014.

[87] Y. Wang, E. Serpedin, P. Ciblat, and P. Loubaton, "Non-data aided feedforward cyclostationary statistics based carrier frequency offset estimators for linear modulations," in Proc. of IEEE Global Telecomm. Conf., vol. 2, (San Antonio, TX, USA), pp. 1386–1390, Nov. 2001.

[88] D. S. Millar, R. Maher, D. Lavery, T. Koike-Akino, M. Pajovic, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel, "Design of a 1 Tb/s superchannel coherent receiver," J. Lightw. Technol., vol. 34, pp. 1453–1463, Mar. 2016.

[89] Pooi Kam, "Maximum likelihood carrier phase recovery for linear suppressed-carrier digital data modulations," Trans. Commun., vol. 34, pp. 522–527, Jun. 1986.

[90] X. Zhou, "An improved feed-forward carrier recovery algorithm for coherent receivers with M-QAM modulation format," Photon. Technol. Lett., vol. 22, pp. 1051–1053, Jul. 2010.

[91] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," in Proc. of Int. Symp. on Inf. Theory (ISIT), (Honolulu, HI, USA), pp. 2426–2429, Jun. 2014.

[92] X. Yangzhang, V. Aref, D. Lavery, and P. Bayvel, "Spectral broadening of gaussian process in optical fibre and implication on the spectral efficiency," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Rome, Italy), Th2.38, Sep. 2018.

[93] R. G. Gallager, Information Theory and Reliable Communication. USA: John Wiley & Sons, Inc., Jan. 1968.

[94] E. T. Jaynes, "Prior probabilities," Trans. Syst. Sci. Cybern., vol. 4, pp. 227–241, Sep. 1968.

[95] D. Commenges, "Information theory and statistics: An overview," arXiv:1511.00860, Nov. 2015.

[96] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," Trans. Inf. Theory, vol. 18, pp. 14–20, Jan. 1972.

[97] R. Blahut, "Computation of channel capacity and rate-distortion functions," Trans. Inf. Theory, vol. 18, pp. 460–473, Jul. 1972.

[98] O. Vasicek, "A test for normality based on sample entropy," Journal of the Royal Statistical Society. Series B (Methodological), vol. 38, no. 1, pp. 54–59, 1976.

[99] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Phys. Rev. E, vol. 69, p. 066138, Jun. 2004.

[100] J. P. Cunningham, P. Hennig, and S. Lacoste-Julien, "Gaussian probabilities and expectation propagation," arXiv:1111.6832, Nov. 2013.

[101] S. M. Moser, Advanced topics in information theory (lecture notes), 4th Edition. Signal and Information Processing Laboratory, ETH Zürich, Switzerland, and Institute of Communications Engineering, National Chiao Tung University (NCTU), Hsinchu, Taiwan, 2019.

[102] H. Sato, "Two-user communication channels," Trans. Inf. Theory, vol. 23, pp. 295–304, May 1977.

[103] A. E. Gamal and M. Costa, "The capacity region of a class of deterministic interference channels (corresp.)," Trans. Inf. Theory, vol. 28, pp. 343–346, Mar. 1982.

[104] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber coherent communication," Trans. Inf. Theory, vol. 52, pp. 5008–5022, Nov. 2006.

[105] K. V. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," J. Lightw. Technol., vol. 15, pp. 2232–2241, Dec. 1997.

[106] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," Trans. Inf. Theory, vol. 54, pp. 3425–3441, Aug. 2008.

[107] G. Kramer, "Autocorrelation function for dispersion-free fiber channels with distributed amplification," Trans. Inf. Theory, vol. 64, pp. 5131–5155, Jan. 2018.

[108] H. Ghozlan and G. Kramer, "Models and information rates for multiuser optical fiber channels with nonlinearity and dispersion," Trans. Inf. Theory, vol. 63, pp. 6440–6456, Oct. 2017.

[109] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," J. Lightw. Technol., vol. 32, pp. 2862–2876, Aug. 2014.

[110] G. Kramer, M. I. Yousefi, and F. R. Kschischang, "Upper bound on the capacity of a cascade of nonlinear and noisy channels," in Proc. of Inf. Theory Workshop (ITW), Apr. 2015.

[111] M. I. Yousefi, G. Kramer, and F. R. Kschischang, "Upper bound on the capacity of the nonlinear Schroedinger channel," Can. Workshop Inf. Theory (CWIT), Jul. 2015.

[112] B. Goebel, R. J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with applications to fiber optics," Trans. Inf. Theory, vol. 57, pp. 5720–5736, Sep. 2011.

[113] T. Xu, B. Karanov, N. A. Shevchenko, D. Lavery, G. Liga, R. I. Killey, and P. Bayvel, "Digital nonlinearity compensation in high-capacity optical communication systems considering signal spectral broadening effect," Scientific Reports, vol. 7, pp. 2045–2322, Oct. 2017.

[114] E. Agrell and M. Karlsson, "Influence of behavioral models on multiuser channel capacity," J. Lightw. Technol., vol. 33, pp. 3507–3515, Sep. 2015.

[115] M. I. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing," in Proc. of Eur. Conf. Opt. Commun. (ECOC), Sep. 2016.

[116] 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, pp. 566–572, Jan. 2016.

[117] S. A. Babin, D. V. Churkin, A. E. Ismagulov, S. I. Kablukov, and E. V. Podivilov, "Four-wave-mixing-induced turbulent spectral broadening in a long raman fiber laser," J. Opt. Soc. Am. B, vol. 24, pp. 1729–1738, Aug. 2007.

[118] S. K. Turitsyn, S. A. Babin, E. G. Turitsyna, G. E. Falkovich, E. V. Podivilov, and D. V. Churkin, Optical Wave Turbulence, pp. 113–163. 2013.

[119] D. Soh and J. Koplow, "Analysis of spectral broadening of incoherent light in optical fibers with nonzero dispersion," Optical Engineering, vol. 50, 11 2011.

[120] D. B. S. Soh, J. P. Koplow, S. W. Moore, K. L. Schroder, and W. L. Hsu, "The effect of dispersion on spectral broadening of incoherent continuous-wave light in optical fibers," Opt. Express, vol. 18, pp. 22393–22405, Oct. 2010.

[121] D. Slepian, "On bandwidth," Proc. IEEE, vol. 64, pp. 292–300, Mar. 1976.

[122] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fiber communications," Nature, vol. 411, pp. 1027–1030, Jun. 2001.

[123] J. Tang, "The channel capacity of a multispan DWDM system employing dispersive nonlinear optical fibers and an ideal coherent optical receiver," J. Lightw. Technol., vol. 20, pp. 1095–1101, Nov. 2002.

[124] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Achievable information rates for high-speed long-haul optical transmission," J. Lightw. Technol., vol. 23, pp. 3755–3763, Dec. 2005.

[125] A. Hasegawa and T. Nyu, "Eigenvalue communication," J. Lightw. Technol., vol. 11, no. 3, pp. 395–399, 1993.

[126] M. Ablowitz and H. Segur, Solitons and the Inverse Scattering Transform. SIAM studies in applied and numerical mathematics: Society for Industrial and Applied Mathematics, SIAM, 1981.

[127] X. Yangzhang, D. Lavery, P. Bayvel, and M. I. Yousefi, "Impact of perturbations on nonlinear frequency-division multiplexing," J. Lightw. Technol., vol. 36, pp. 485–494, Jan. 2018.

[128] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Method for solving the Korteweg-deVries equation," Phys. Rev. Lett., vol. 19, pp. 1095–1097, Nov. 1967.

[129] P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves," Commun. Pure Appl. Math., vol. 21, no. 5, pp. 467–490, 1968.

[130] Symmetries and Integrability of Difference Equations. London Mathematical Society Lecture Note Series, Cambridge University Press, 2011.

[131] N. Hitchin, "Integral systems in riemannian geometry," Surveys in Differential Geometry, vol. 4, no. 1, pp. 21–81, 1998.

[132] V. E. Zakharov and A. B. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media," Soviet Journal of Experimental and Theoretical Physics, vol. 34, p. 62, Jan. 1972.

[133] J.-W. Goossens, M. I. Yousefi, Y. Jaouën, and H. Hafermann, "Polarization-division multiplexing based on the nonlinear Fourier transform," Opt. Express, vol. 25, pp. 26437–26452, Oct. 2017.

[134] S. V. Manakov, "On the theory of two-dimensional stationary self-focusing of electromagnetic waves," Soviet Journal of Experimental and Theoretical Physics, vol. 38, pp. 505–516, Feb. 1974.

[135] M. J. Ablowitz, B. Prinari, and A. D. Trubatch, Nonlinear Schrödinger equation (NLS), p. 18–45. London Mathematical Society Lecture Note Series, Cambridge University Press, 2003.

[136] L. Faddeev, A. Reyman, and L. Takhtajan, Hamiltonian Methods in the Theory of Solitons. Classics in Mathematics, Springer Berlin Heidelberg, 2007.

[137] M. Klaus and J. Shaw, "On the eigenvalues of Zakharov-Shabat systems," SIAM J. on Math. Analysis, vol. 34, no. 4, pp. 759–773, 2003.

[138] M. Yousefi and X. Yangzhang, "Linear and nonlinear frequency-division multiplexing," Trans. Inf. Theory, vol. 66, pp. 478–495, Jan. 2020.

[139] A. Hasegawa and Y. Kodama, Solitons in Optical Communications. Oxford series in optical and imaging sciences, Clarendon Press, 1995.

[140] J. E. Prilepsky, S. A. Derevyanko, and S. K. Turitsyn, "Nonlinear spectral management: Linearization of the lossless fiber channel," Opt. Express, vol. 21, pp. 24344–24367, Oct. 2013.

[141] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers," Opt. Express, vol. 22, pp. 26720–26741, Nov. 2014.

[142] S. Wahls, S. T. Le, J. E. Prilepsk, H. V. Poor, and S. K. Turitsyn, "Digital backpropagation in the nonlinear Fourier domain," in 16th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC), (Stockholm, Sweden), pp. 445–449, Jun. 2015.

[143] V. Aref, S. T. Le, and H. Buelow, "Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Düsseldorf, Germany), Th.3.B.2, Sep. 2016.

[144] I. Tavakkolnia and M. Safari, "Signalling over nonlinear fibre-optic channels by utilizing both solitonic and radiative spectra," in Euro. Conf. on Networks and Commun. (EuCNC), (Paris, France), pp. 103–107, Jul. 2015.

[145] S. T. Le et al., "Demonstration of nonlinear inverse synthesis transmission over transoceanic distances," J. Lightw. Technol., vol. 34, pp. 2459–2466, May 2016.

[146] S. Le, V. Aref, and H. Buelow, "Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit," Nat. Photon., vol. 11, pp. 570–576, Jul. 2017.

[147] S. Hari, M. I. Yousefi, and F. R. Kschischang, "Multieigenvalue communication," J. Lightw. Technol., vol. 34, pp. 3110–3117, Jul. 2016.

[148] Z. Dong, S. Hari, T. Gui, K. Zhong, M. I. Yousefi, C. Lu, P. A. Wai, F. R. Kschischang, and A. P. T. Lau, "Nonlinear frequency division multiplexed transmissions based on NFT," Photon. Technol. Lett., vol. 27, pp. 1621–1623, Aug. 2015.

[149] T. Gui, S. K. Lo, X. Zhou, C. Lu, A. P. T. Lau, and P. Wai, "4 bits/symbol phase and amplitude modulation on a single discrete eigenvalue for transmissions based on nonlinear Fourier transform," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Los Angeles, CA, USA), Th2A.58, Mar. 2017.

[150] S. T. Le, H. Buelow, and V. Aref, "Demonstration of 64× 0.5 Gbaud nonlinear frequency division multiplexed transmission with 32QAM," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (Los Angeles, CA, USA), W3J.1, Mar. 2017.

[151] A. Geisler and C. G. Schaeffer, "Experimental nonlinear frequency division multiplexed transmission using eigenvalues with symmetric real part," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Düsseldorf, Germany), Th.2.P2.SC5.56, Sep. 2016.

[152] H. Buelow, V. Aref, and W. Idler, "Transmission of waveforms determined by 7 eigenvalues with PSK-modulated spectral amplitudes," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Düsseldorf, Germany), Tu.3.E.2, Sep. 2016.

[153] I. Tavakkolnia and M. Safari, "Capacity analysis of signaling on the continuous spectrum of nonlinear optical fibers," J. Lightw. Technol., vol. 35, pp. 2086–2097, Jun. 2017.

[154] S. K. Turitsyn, J. E. Prilepsky, S. T. Le, S. Wahls, L. L. Frumin, M. Kamalian, and S. A. Derevyanko, "Nonlinear Fourier transform for optical data processing and transmission: advances and perspectives," Optica, vol. 4, pp. 307–322, Mar. 2017.

[155] T. Gui, T. H. Chan, C. Lu, A. P. T. Lau, and P. A. Wai, "Alternative decoding methods for optical communications based on nonlinear Fourier transform," J. Lightw. Technol., vol. 35, pp. 1542–1550, May. 2017.

[156] S. Le, Y. Prylepskiy, M. Kamalian, P. Rosa, M. Tan, J. Ania-Castanon, P. Harper, and S. Turitsyn, "Modified nonlinear inverse synthesis for opti-

cal links with distributed Raman amplification," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Valencia, Spain), 7341904, Sep. 2015.

[157] M. K. Kopae, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, "Optical communication based on the periodic nonlinear Fourier transform signal processing," in 6th Int. Conf. Photon. (ICP), (Kuching, Malaysia), Mar. 2016.

[158] 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," Photon. Technol. Lett., vol. 28, pp. 1803–1806, Sep. 2016.

[159] S. A. Derevyanko, J. E. Prilepsky, and S. K. Turitsyn, "Capacity estimates for optical transmission based on the nonlinear Fourier transform," Nature Commun., vol. 7, Sep. 2016.

[160] O. Kotlyar, M. Pankratova, M. Kamalian-Kopae, A. Vasylchenkova, J. E. Prilepsky, and S. K. Turitsyn, "Combining nonlinear Fourier transform and neural network-based processing in optical communications," Opt. Lett., vol. 45, pp. 3462–3465, Jul 2020.

[161] M. Pankratova, A. Vasylchenkova, S. A. Derevyanko, N. B. Chichkov, and J. E. Prilepsky, "Signal-noise interaction in optical-fiber communication systems employing nonlinear frequency-division multiplexing," Phys. Rev. Applied, vol. 13, p. 054021, May 2020.

[162] L. F. Mollenauer and J. P. Gordon, Solitons in Optical Fibers: Fundamentals and Applications. Amsterdam, The Netherlands: Elsevier Academic Press, First ed., 2006.

[163] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, "Nonlinear inverse synthesis technique for optical links with lumped amplification," Opt. Express, vol. 23, pp. 8317–8328, Mar. 2015.

[164] G. Liga et al., "Ultra-wideband nonlinearity compensation performance in the presence of PMD," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Düsseldorf, Germany), W.4.P1.SC3.28, Sep. 2016.

[165] 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, pp. 148–157, Feb. 1996.

[166] S. Civelli, S. K. Turitsyn, M. Secondini, and J. E. Prilepsky, "Polarization-multiplexed nonlinear inverse synthesis with standard and reduced-complexity NFT processing," Opt. Express, vol. 26, pp. 17360–17377, Jun 2018.

[167] I. T. Lima, V. S. Grigoryan, M. O'Sullivan, and C. R. Menyuk, "Computational complexity of nonlinear transforms applied to optical communications systems with normal dispersion fibers," in Int. Photon. Conf. (IPC), vol. 3, (Reston, VA, USA), pp. 277–278, Oct. 2015.

[168] S. T. Le, V. Aref, and H. Buelow, "125 Gbps pre-compensated nonlinear frequency-division multiplexed transmission," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Gothenburg, Sweden), W.3.C.3, Sep. 2017.

[169] S. Wahls, "Generation of time-limited signals in the nonlinear Fourier domain via b-modulation," in Proc. of Eur. Conf. Opt. Commun. (ECOC), (Gothenburg, Sweden), W.3.C.6, Sep. 2017.

[170] V. Vaibhav, "Fast inverse nonlinear Fourier transformation using exponential one-step methods, part I: Darboux transformation," arXiv:1704.00951, Apr. 2017.

[171] M. I. Yousefi, "The asymptotic capacity of the optical fiber," arXiv:1610.06458, Oct. 2016.

[172] W. A. Gemechu, T. Gui, J. Goossens, M. Song, S. Wabnitz, H. Hafermann, A. P. T. Lau, M. I. Yousefi, and Y. Jaouën, "Dual polarization nonlinear frequency division multiplexing transmission," Photon. Technol. Lett., vol. 30, pp. 1589–1592, Sep. 2018.

[173] S. T. Le and H. Buelow, "64 $\times$ 0.5 Gbaud nonlinear frequency division multiplexed transmissions with high order modulation formats," J. Lightw. Technol., vol. 35, pp. 3692–3698, Sep. 2017.

[174] S. T. Le, V. Aref, and H. Buelow, "High speed precompensated nonlinear frequency-division multiplexed transmissions," J. Lightw. Technol., vol. 36, pp. 1296–1303, Mar. 2018.

[175] S. T. Le, K. Schuh, F. Buchali, and H. Buelow, "100 Gbps b-modulated nonlinear frequency division multiplexed transmission," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), (San Diego, CA, USA), W1G.6, Mar. 2018.

[176] T. Gui, G. Zhou, C. Lu, A. P. T. Lau, and S. Wahls, "Nonlinear frequency division multiplexing with b-modulation: shifting the energy barrier," Opt. Express, vol. 26, pp. 27978–27990, Oct. 2018.

[177] S. Civelli, E. Forestieri, and M. Secondini, "Decision-feedback detection strategy for nonlinear frequency-division multiplexing," Opt. Express, vol. 26, pp. 12057–12071, Apr. 2018.

[178] X. Yangzhang, V. Aref, S. T. Le, H. Bülow, and P. Bayvel, "400 Gbps dual-polarisation non-Linear frequency-division multiplexed transmission with b-Modulation," in Proc. of Eur. Conf. Opt. Commun. (ECOC), We4F.1, Sep. 2018.

[179] A. Span, V. Aref, H. Bulow, and S. ten Brink, "Precoding for dual polarization soliton transmission," in Opto-Electronics and Commun. Conf. (OECC), (Jeju Island, Korea), Jul. 2018.

[180] V. Aref, S. T. Le, and H. Buelow, "Modulation over nonlinear Fourier spectrum: Continuous and discrete spectrum," J. Lightw. Technol., vol. 36, pp. 1289–1295, Mar. 2018.

[181] S. Gaiarin, A. M. Perego, E. P. da Silva, F. D. Ros, and D. Zibar, "Dual-polarization nonlinear Fourier transform-based optical communication system," Optica, vol. 5, pp. 263–270, Mar. 2018.

[182] F. Da Ros, S. Civelli, S. Gaiarin, E. P. da Silva, N. De Renzis, M. Secondini, and D. Zibar, "Dual-polarization NFDM transmission with continuous and discrete spectral modulation," J. Lightw. Technol., vol. 37, pp. 2335–2343, May 2019.

[183] J. Armstrong, "OFDM for optical communications," J. Lightw. Technol., vol. 27, pp. 189–204, Feb. 2009.

[184] O. V. Sinkin, R. Holzlohner, J. Zweck, and C. R. Menyuk, "Optimization of the split-step fourier method in modeling optical-fiber communications systems," J. Lightw. Technol., vol. 21, pp. 61–68, Jan. 2003.

[185] D. M. Arnold, H. . Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," Trans. Inf. Theory, vol. 52, pp. 3498–3508, Aug. 2006.

[186] X. Yangzhang, V. Aref, S. T. Le, H. Buelow, D. Lavery, and P. Bayvel, "Dual-polarization non-linear frequency-division multiplexed transmission with *b*-modulation," J. Lightw. Technol., vol. 37, pp. 1570–1578, Mar. 2019.

[187] L. Schmalen, V. Aref, J. Cho, D. Suikat, D. Rösener, and A. Leven, "Spatially coupled soft-decision error correction for future lightwave systems," J. Lightw. Technol., vol. 33, pp. 1109–1116, Mar. 2015.

[188] S. Chimmalgi, P. J. Prins, and S. Wahls, "Fast nonlinear fourier transform algorithms using higher order exponential integrators," IEEE Access, vol. 7, pp. 145161–145176, Oct. 2019.

[189] S. Li, J. Koch, and S. Pachnicke, "Optical signal processing in the discrete nonlinear frequency domain," in Proc. of Opt. Fiber Commun. Conf. and Exhib. (OFC), p. W2A.40, 2018.

[190] D. Wackerly, W. Mendenhall, and R. Scheaffer, Mathematical Statistics with Applications. Cengage Learning, 2014.

[191] J. Rosenthal, A First Look at Rigorous Probability Theory. World Scientific, 2006.

[192] P. Stoica and R. Moses, Spectral Analysis of Signals. Pearson Prentice Hall, 2005.