

DIGITAL SIGNAL PROCESSING FOR TRANSMISSION SYSTEMS WITH FIBRE OPTICAL PARAMETRIC DEVICES

Long Hoang Nguyen

Doctor of Philosophy

ASTON UNIVERSITY

September 2024

© Long Hoang Nguyen, 2024

Long Hoang Nguyen asserts their moral right to be identified as the author of this thesis

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright belongs to its author and that no quotation from the thesis and no information derived from it may be published without appropriate permission or acknowledgement.

Abstract

Digital Signal Processing for Transmission Systems with Fibre Optical Parametric Devices

Long Hoang Nguyen

Doctor of Philosophy

2024

Fibre-optical parametric devices, including two candidates fibre-optical parametric amplifier and optical phase conjugator, play crucial roles in the future of optical communications. They offer solutions for high gain with ultra-wide bandwidth, low noise and all-optical mitigation of optical channel impairments. However, the stimulated Brillouin scattering effect poses a fundamental challenge as it limits the pump power these devices need. Among suppression techniques, pump-phase modulation is widely used because of its modest cost although it introduces a temporal variation in the devices' transfer functions. In this thesis, we propose digital signal processing algorithms, including both parametric and non-parametric approaches, which can compensate for penalties induced by pump-phase modulation schemes. We focus on the linear and nonlinear regression tasks, where the latter is solved by kernel methods. By carrying out numerical simulations, we prove the effectiveness of our proposed schemes against the existing compensation algorithms. Therefore, we believe the proposed algorithms will unlock the potentials of optical parametric devices, simplify the experimental design and become a vital part of parametric devices enabled transmission systems.

Keywords:

Digital signal processing, Optical parametric process, Fibre-optical parametric amplifier, Optical phase conjugator, Kernel methods

To my beloved family who is always in my heart.

Personal Acknowledgements

Foremost, I would like to express my deepest gratitude to my academic supervisor, Dr. Stylianos Sygletos, for his invaluable support and encouragement throughout my research journey. His expertise, vision, and enthusiasm have been a constant source of motivation, guiding me in the right direction. I am sincerely grateful for his unwavering support, particularly during the more challenging phases of my research. It has been a great honor to work under his supervision, and I am thankful for the opportunity to benefit from his knowledge and mentorship.

I would also like to extend my heartfelt thanks to my co-supervisor, Dr. Sonia Boscolo, for her insightful lessons on parametric amplification and pump phase modulation schemes. I am deeply appreciative of the time and dedication she invested in providing thorough feedback on my paper drafts. Her guidance and support have been invaluable to the progress of my research.

My sincere thanks also go to Prof. Sergei Turitsyn for his unwavering support. He is one of the key individuals who provided me with the opportunity to conduct research at the Aston Institute of Photonic Technologies (AiPT) and has ensured that my research journey proceeded as smoothly as possible. I am truly grateful for his encouragement throughout this period.

I would also like to thank everyone at the AiPT: the project managers, researchers and other PhD students for insightful discussions and collaborative work.

Last but not least, I would like to express my heartfelt gratitude to my family for encouraging me during the challenging moments of the past few years. I am especially thankful to my wife for her love, understanding, and endless support throughout this journey.

List of Contents

List of Abbreviations	viii
List of Tables	xi
List of Figures	xii
List of Publications	xvi
1 Introduction	1
1.1 Motivation	1
1.2 Thesis outline	2
2 Fibre-optical parametric devices	4
2.1 Single-mode fibre properties	5
2.1.1 Chromatic dispersion	5
2.1.2 Fibre nonlinearity	6
2.2 Fibre-optical parametric amplifier	7
2.2.1 Dual-pump scheme	10
2.2.2 Single-pump scheme	16
2.3 Optical phase conjugation	19
2.4 SBS suppression techniques	21
2.4.1 SBS suppression through nonuniform Brillouin frequency shift	23
2.4.2 SBS suppression through pump-phase modulation	24
2.5 Optimisation of pump-phase modulation with sinusoidals	27
2.6 Conclusion	32

3	Digital signal processing	34
3.1	Coherent transceiver	35
3.2	Chromatic dispersion compensation	36
3.3	Carrier phase recovery	38
3.3.1	LMS-based phase recovery	39
3.3.2	Blind phase search	40
3.4	Digital signal processing with kernel methods	42
3.4.1	Feature mapping	42
3.4.2	Kernel ridge regression	44
3.4.3	Gaussian process	46
3.4.4	Kernel functions	47
3.4.5	Kernel hyperparameter optimisation	49
3.5	Conclusion	50
4	Amplitude and phase distortion compensation in systems with optical phase conjugator	51
4.1	Mid-link optical phase conjugation transmission	51
4.2	Residual dithering-induced distortion	53
4.2.1	Residual dithering-induced phase distortion	54
4.2.2	Phase-to-phase and phase-to-amplitude distortion conversion	55
4.3	Phase and amplitude distortion compensation scheme	56
4.3.1	Operating principle	57
4.3.2	Compensation scheme	58
4.3.3	Estimation of dithering-induced phase distortion	59
4.4	Compensation result	60
4.4.1	Performance over transmission distance	62
4.4.2	Impact of phase mismatch	63
4.5	Conclusion	64
5	Phase and amplitude distortion compensation in optical parametric amplified systems	65
5.1	Fibre-optical parametric amplified transmission	66
5.2	Influence of pump-phase modulation	66

5.3	Phase and amplitude distortion compensation schemes	69
5.3.1	Operating principle	69
5.3.2	Multi-branch compensation scheme	71
5.3.3	Single-branch compensation scheme	73
5.4	Compensation results	75
5.4.1	Performance over number of spans	76
5.4.2	Amplitude and phase suppression capability	77
5.4.3	Tolerance to dithering frequency deviations	78
5.4.4	Base frequency	79
5.4.5	Filter optimisation process	80
5.5	Conclusion	81
6	Kernel-based compensation in optical parametric amplified systems	83
6.1	Combining kernels	84
6.2	Sliding-window kernel-based method	85
6.3	Kernel-based phase compensation using real combined kernel	87
6.3.1	Compensation scheme and kernel selection	88
6.3.2	Compensation results	91
6.4	Kernel-based phase and amplitude compensation using complex kernel	93
6.4.1	Compensation scheme	94
6.4.2	Compensation results	96
6.5	Kernel-based phase and amplitude compensation using complexification of RKHS	97
6.5.1	Learning with complexification of Hilbert space	97
6.5.2	Compensation results	98
6.6	Conclusion	101
7	Conclusions and future work	102
7.1	Key findings	102
7.2	Future work	104
	List of References	105

List of Abbreviations

ADC analog-to-digital converter

ALD approximate linear dependency

ASE amplified spontaneous emission

AWG arbitrary waveform generator

BER bit-error-rate

BPS blind phase search

CDC chromatic dispersion compensation

CPR carrier phase recovery

CV complex-valued

DAC digital-to-analog converter

DBP digital-back-propagation

DDC dithering-induced distortion compensation

DSP digital signal processing

DWDM dense wavelength-division multiplexing

EDFA Erbium-doped fibre amplifier

FIR finite impulse response

FOPA fibre-optical parametric amplifier

- FWM** four-wave mixing
- GD** gradient descent
- GP** Gaussian process
- GSP** grid search on phase
- GVD** group-velocity dispersion
- HNLF** highly nonlinear fibre
- IM-DD** intensity modulation-direct detection
- LMS** least mean squares
- LO** local oscillator
- LPF** low-pass filter
- MAP** maximising a posteriori
- MMF** multi-mode fibre
- MSE** mean square error
- MZM** Mach-Zehnder modulator
- NLSE** nonlinear Schrodinger equation
- OA** optical amplifier
- OPC** optical phase conjugator
- PBC** polarisation beam combiner
- PBS** polarisation beam splitter
- PD** photodiode
- PM** phase modulator
- PMD** polarisation mode dispersion
- PRBS** pseudo-random binary sequence

- PSA** phase-sensitive amplifier
- PSD** power spectral density
- PSK** phase shift keying
- QAM** quadrature amplitude modulation
- RBF** radial basis function
- RF** radio-frequency
- RKHS** reproducing kernel Hilbert space
- RMS** root-mean-square
- RRC** root-raised-cosine
- Rx** receiver
- SBS** stimulated Brillouin scattering
- SMF** single-mode fibre
- SPM** self-phase modulation
- SSFM** split-step Fourier method
- SWKRLS** sliding-window kernel recursive least square
- TIA** transimpedance amplifier
- Tx** transmitter
- XPM** cross-phase modulation
- WDM** wavelength-division multiplexing

List of Tables

2.1	Dual-pump scheme parameters	15
2.2	Single-pump scheme parameters	18
4.1	SBS-limited conversion efficiency for different levels of SBS suppression.	52
4.2	OPC transmission numerical simulation parameters	61
5.1	FOPA transmission numerical simulation parameters	76
6.1	Single-pump scheme parameters	88

List of Figures

2.1	General scheme of fibre-optical parametric amplifier.	8
2.2	Diagram of frequency locations in four-wave mixing in two scenarios: (a) dual-pump FOPA and (b) single-pump FOPA.	9
2.3	Theoretical graphs in the FOPA dual-pump scenario with $\beta^{(4)} > 0$ of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$	14
2.4	Theoretical graphs in the FOPA dual-pump scenario with $\beta^{(4)} < 0$ of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$	14
2.5	FOPA gain spectrum in the dual-pump scheme.	16
2.6	Theoretical graphs in the FOPA single-pump scenario of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$	17
2.7	FOPA gain spectrum in the single-pump scheme.	18
2.8	Schematic of phase-conjugated wave generation in FWM.	20
2.9	Schematic of mid-link OPC transmission and spectrum changes along the link.	21
2.10	Schematic diagram about stimulated Brillouin scattering. The pump light at ν_p scatters from Bragg diffraction induced by the acoustic wave at ν_B to generate the Stokes wave at $\nu_s = \nu_p - \nu_B$	22
2.11	(a) Setup diagram of pump-phase modulation using multiple sinusoidal tones. (b) Pump spectrum after being phase modulated by a sinusoidal tone ω_{m_i}	25
2.12	(a) Estimated pump spectrum after phase modulation with one tone at ω_{m_1} . (b) Estimated pump spectrum after phase modulation with two tones at ω_{m_1} and $\omega_{m_2} = 3\omega_{m_1}$	26

2.13	Ideal rectangular pump spectrum used as the reference.	28
2.14	The convergence of the mean square error (MSE) cost function J in the modulation schemes with(a) 3 tones and (b) 4 tones.	29
2.15	The evolution of the amplitudes and phases during the optimisation process in the modulation schemes with(a) 3 tones and (b) 4 tones.	30
2.16	The broadened spectra after pump-phase modulation using the optimal parameters in the modulation schemes with (a) 3 tones and (b) 4 tones.	31
2.17	The SBS power threshold increase evolution during the optimisation process in the modulation schemes with (a) 3 tones and (b) 4 tones.	32
3.1	Configuration of coherent transmitter.	35
3.2	Configuration of coherent receiver.	36
3.3	LMS-based phase recovery scheme.	39
3.4	BPS phase recovery scheme.	41
3.5	Demonstration of feature mapping to $\phi(\cdot)$ to Hilbert space.	43
3.6	Gaussian kernel with its (a) covariance matrix and (b) example functions sampled from GP prior.	48
3.7	Periodic kernel with its (a) covariance matrix and (b) example functions sampled from GP prior.	49
4.1	Scheme of OPC transmission.	52
4.2	Dual-pump configuration for OPC implementation.	53
4.3	Constellation diagrams of presumed OPC transmission without ASE and laser phase noise in (a) back-to-back configuration (b) 10-span configuration.	54
4.4	Phase-to-phase and phase-to-amplitude modulation effects in dispersive channel.	56
4.5	Baseband equivalent model of transmission with single-stage OPC.	57
4.6	The compensation scheme for OPC transmission.	59
4.7	Power spectral density of the residual dithering-induced phase distortion.	59
4.8	Regression learning of residual dithering-induced phase distortion.	60
4.9	Launched power optimisation in mid-link OPC transmission.	62
4.10	Performance of OPC transmission system with compensation schemes.	62
4.11	Performance of OPC transmission system with compensation schemes considering different phase mismatches.	63

5.1	Schematic diagram of a transmission system with N cascaded FOPA stages.	66
5.2	(a) The amplitude and (b) phase spectral responses of the FOPA complex gain, along with the corresponding dithering-induced RMS amplitude and phase fluctuations as a function of signal wavelength detuning.	67
5.3	The baseband equivalent model of FOPA transmission.	69
5.4	Proposed multi-branch dithering-induced distortion compensation scheme.	71
5.5	Performance comparison between the conventional carrier phase recovery algorithm and the proposed compensation algorithm. The inset illustrate the constellation diagrams after 1500 km.	72
5.6	Dithering-induced distortion compensation in the DSP chain.	73
5.7	Dithering-induced distortion estimation filter.	74
5.8	Flow diagram of dithering-induced distortion compensation algorithm.	75
5.9	The performance of the proposed compensation scheme in terms of Q^2 -factor versus number of spans.	77
5.10	The evolution of the RMS phase and amplitude fluctuations.	78
5.11	Performance improvement due to DDC algorithm versus the dithering frequency variation.	79
5.12	Q^2 -factor performance as a function of the base frequency ν_{m1}	80
5.13	The filter optimisation for BPS and DDC algorithms.	81
6.1	Example functions sampled from GP prior when using a combined kernel from Gaussian and periodic kernels by (a) multiplying them and (b) adding them.	84
6.2	(a) Kernel matrices \mathbf{K}_n with growing sizes. (b) Kernel matrices \mathbf{K}_n with fixed sizes in sliding-window approach.	86
6.3	FOPA link with multiple cascaded stages.	88
6.4	Block diagram of the proposed kernel-based phase recovery scheme.	89
6.5	Hyperparameter optimisation with log marginal likelihood for kernel-based phase recovery.	92
6.6	Data fitting within the training set.	92
6.7	Performance result comparison for kernel-based phase compensation versus conventional phase recovery.	93

6.8 (a) The amplitude and (b) phase spectral responses of the FOPA complex gain and the corresponding dithering-induced RMS amplitude and phase fluctuations versus the signal wavelength detuning.	94
6.9 Block diagram of the proposed complex-valued kernel-based recovery scheme.	95
6.10 Hyperparameter optimisation for the kernel method using complex kernel function.	96
6.11 Performance comparison between complex-valued kernel-based compensation using complex kernel function and the conventional phase recovery algorithm. .	97
6.12 Hyperparameter optimisation for kernel-based method using complexification of RKHS.	98
6.13 Performance comparison between complex-valued kernel-based compensation using complexification of RKHS and the conventional phase recovery algorithm.	99
6.14 Amplitude and phase suppression capability of kernel-based compensation using complexification of RKHS compared to the conventional phase recovery algorithm.	100
6.15 Performance comparison between two complex-valued kernel-based compensation schemes and the conventional phase recovery algorithm.	100

List of Publications

Publications Arising from this Thesis:

- [1] L. H. Nguyen, S. Boscolo, A. D. Ellis, and S. Sygletos, "Machine learning enabled compensation of phase-to-amplitude distortion in opc systems," in *2023 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, 2023.
- [2] L. H. Nguyen, T. T. Nguyen, A. D. Ellis, S. Sygletos, and S. Boscolo, "Digital signal processing for optical phase conjugation assisted coherent systems," in *2023 23rd International Conference on Transparent Optical Networks (ICTON)*, pp. 1–4, 2023.
- [3] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Machine learning enabled digital compensation of phase-to-amplitude distortion in fibre-optical parametric amplifier based transmission links," *EPJ Web Conf.*, vol. 287, p. 13022, 2023.
- [4] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Design of optical transmission link with cascaded fibre-optical parametric amplifiers," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, 2024.
- [5] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Digital signal processing for fibre-optical parametrically amplified transmission links," in *2024 24th International Conference on Transparent Optical Networks (ICTON)*, pp. 1–4, 2024.
- [6] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Online kernel-based phase recovery for parametrically amplified optical transmission," in *2024 14th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 95–98, 2024.
- [7] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Online phase and amplitude distortion compensation in fopa transmission systems," in *2024 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, pp. 1–2, 2024.
- [8] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Complex-valued kernel-based phase and amplitude distortion compensation in parametrically amplified optical links," in *2024 50th European Conference on Optical Communication (ECOC)*, 2024.
- [9] L. H. Nguyen, S. Boscolo, and S. Sygletos., "Complex-valued kernels for compensation of pump dithering induced distortions in fopa links," in *2024 IEEE Photonics Conference (IPC)*, pp. 1–2, 2024.
- [10] L. H. Nguyen, S. Boscolo, and S. Sygletos., "Online digital compensation of pump dithering induced phase and amplitude distortions in transmission links with cascaded fibre-optical

parametric amplifiers,” *Opt. Express*, vol. 32, pp. 13467–13477, Apr 2024.

Additional Publications Completed During my Studies:

[11] M. Bastamova, V. Gordienko, S. Sygletos, M. Tan, A. Donodin, L. H. Nguyen, F. Bessin, S. Boscolo, N. Doran, and A. D. Ellis, “Performance evaluation of a polarisation insensitive Mach-Zehnder fiber parametric amplifier with 38 channel transmission,” in *2024 50th European Conference on Optical Communication (ECOC)*, 2024.

Chapter 1

Introduction

1.1 Motivation

The increasing growth of data traffic has driven the demand for more bandwidth and higher transmission speeds, accelerating advancements in optical communication systems. While traditional optical devices, such as semiconductor lasers and conventional amplifiers, have been able to address many of these needs, they face inherent limitations in bandwidth and tunability. These constraints pose significant challenges to the performance and scalability of optical transmissions, making it difficult to keep pace with the rapidly growing data demands of future networks.

Fibre-optical parametric devices, which utilise the fibre's nonlinear optical properties, especially four-wave mixing (FWM) effect, have emerged as promising candidates to overcome existing limitations of traditional systems and advance the next generation of optical networks. Two notable examples of such devices are the optical phase conjugator (OPC), used for signal recovery and distortion compensation, and the fibre-optical parametric amplifier (FOPA), which offers a novel approach to signal amplification. Both devices have attracted significant research attention due to their wide range of advantages and have strong potential to become pivotal technologies in future optical communication systems. Their advantages include the ability to operate over a broad spectral range, enabling higher data throughput and more efficient use of the available optical spectrum. The FOPA, in particular, can achieve very high gain with low noise figures, making it an effective solution for significantly extending transmission reach. The OPC offers an interesting approach to signal processing by optically compensating for signal

distortions along the transmission link. Both can be dynamically tuned across a wide range of wavelengths, making them suitable for wavelength-division multiplexing (WDM) systems.

While their potential is immense, fundamental challenges, such as stimulated Brillouin scattering (SBS), have hindered the full integration of fibre-optical parametric devices into existing optical systems. Significant progress has been made in SBS suppression research for fibre-optical parametric devices, with several key milestones achieved. In addition to proposed techniques related to fibre design and fabrication, the introduction of phase modulation to the pump source in fibre-optical parametric devices has emerged as a widely adopted and effective approach due to its efficiency and relatively low implementation cost. However, this approach presents its own challenge, which causes temporal variation in the device's transfer function, potentially leading to signal degradation, especially in systems using coherent optical transmission. Designing such systems while ensuring the signal quality, particularly in scenarios with cascaded devices like the cascaded FOPA link, remains a complex and ongoing challenge.

The thesis aims to mitigate this problem through the advancements in digital signal processing (DSP). By analysing the nature of the signal distortion induced by pump-phase modulation, we propose DSP algorithms, including both parametric and non-parametric approaches, to estimate and mitigate signal distortion in fibre-optical parametric devices-based systems. These proposed DSP algorithms, which are designed for both OPC and FOPA, will enable the potential for further SBS suppression improvement using the pump-phase modulation approach while ensuring robust end-to-end system performance. These solutions will pave the way for more effective integration of fibre-optical parametric devices into modern optical networks.

1.2 Thesis outline

The motivation behind designing DSP algorithms for fibre-optical parametric devices-based transmission, along with the structure of this thesis, is outlined in chapter 1.

Following this introduction, chapter 2 presents the necessary background knowledge and explains the operating principles of FOPA and OPC. This chapter also reviews existing techniques for SBS suppression aimed at enhancing their efficiency. Notably, a novel approach for optimising the pump-phase modulation parameters is proposed, along with calculations demonstrating the increase in the SBS power threshold.

The fundamentals of key DSP blocks are covered in chapter 3. This chapter discusses

existing DSP algorithms, including chromatic dispersion compensation (CDC) and carrier phase recovery (CPR), which are essential to support the proposed algorithms for optical transmission using fibre-optical parametric devices. We will also cover the theory of the kernel-based method framework, which will be used in the latter part of the thesis.

In chapter 4, we develop a compensation scheme for mitigating both signal phase and amplitude distortions in transmission systems using mid-link OPC. The proposed algorithm is compared with phase-only compensation schemes to highlight the critical importance of addressing distortions in both signal domains.

In chapter 5, we propose an online digital compensation scheme for the optical system with cascaded FOPA. The pump-dithering distortion in this scenario requires a more detailed investigation, particularly their evolution across multiple cascaded stages. Despite this challenge, the proposed approach demonstrates an effective improvement in signal quality over the conventional CPR algorithm.

In chapter 6, we apply the kernel-based methods to advance the compensation algorithm. As a non-parametric learning model, this approach offers more flexibility within the DSP chain by providing a compensation solution without prior knowledge of the pump-phase modulation frequencies. Our compensation algorithms are done in both real and complex domains, with the latter offering more potential for both amplitude and phase compensation.

Finally, we summarise the key findings of this thesis in chapter 7 and outline potential areas for future research.

Chapter 2

Fibre-optical parametric devices

Fibre-optical parametric devices are optical components that leverage parametric processes within fibre to perform various functionalities. A parametric process involves nonlinear interactions where the properties of light are modified by interactions with other optical waves, driven by the effects of nonlinear fibre. For efficient interactions, the phase-matching condition must be satisfied, which is significantly influenced by the fibre's dispersion characteristics. Therefore, to fully understand the operating principle of fibre-optical parametric devices, it is crucial to first discuss the fibre's chromatic dispersion and nonlinearity, as these are fundamental fibre properties affecting performance.

We will then examine different types of fibre-optical parametric devices, starting with FOPA. This analysis involves deriving its behaviour under different pump schemes based on the fibre parameters, allowing us to assess its performance and optimise its gain spectrum effectively. We will also discuss the OPC, focusing on its behaviour and its application in mitigating signal impairment.

Another major topic in this chapter is the SBS effect, which imposes fundamental limitations on the operation of fibre-optical parametric devices. We will explain its principle and review existing suppression techniques, focusing on the pump-phase modulation approach. This technique provides significant suppression with a relatively modest cost. Finally, we will propose an optimisation procedure, which is a part of our published ONDM conference paper [4] and Optics Express article [10], for the selection of pump-phase modulation parameters. This procedure aims to achieve a significant SBS power threshold increase while minimising the modulating bandwidth.

2.1 Single-mode fibre properties

While fibre-optical parametric devices can also be implemented in multi-mode fibre (MMF) [12], they encounter challenges due to intermodal dispersion effects, where different fibre modes travel at varying speeds, leading to reduced coherence in the parametric processes. In contrast, the single-mode fibre (SMF), which supports only a single propagation mode, simplifies the design of these processes, making them more predictable and easier to optimise. Furthermore, SMF is well-established and widely used in telecommunications and other applications, contributing to its accessibility and reliability. As a result, most advancements in fibre-optical parametric devices have been achieved using SMF. Therefore, in this chapter, we will focus on studying the properties of SMF and their implications for fibre-optical parametric devices.

Fibres used in transmission links exhibit power attenuation, characterised by a loss coefficient, denoted as α , with the value for SMF typically around 0.2 dB/km at 1550 nm. However, for this discussion, we will assume that fibres, including highly nonlinear fibre (HNLF) used within the fibre-optical parametric devices, are ideally lossless. Under this assumption, we will delve into the principles of chromatic dispersion, which plays a crucial role in shaping the gain spectra of the fibre-optical parametric devices, and then fibre nonlinearity, with a particular focus on FWM, the fundamental mechanism driving the operation of the fibre-optical parametric devices.

2.1.1 Chromatic dispersion

For an optical wave with angular frequency ω propagating along a lossless SMF¹, the electric field component at a distance z and time t , denoted as $E(z, t)$, can be expressed as [13]

$$E(z, t) = E(0, t)e^{j(\beta z - \omega t)}, \quad (2.1)$$

where $\beta = 2\pi n/\lambda$ is the propagation constant or wavevector, with λ representing the wavelength and n the fibre's refractive index. It is important to note that β depends on the angular frequency ω , and is also denoted as $\beta(\omega)$.

For a point where the phase $\beta z - \omega t$ remains constant, the velocity at this point, known as the phase velocity, is defined as $v_p = dz/dt = \omega/\beta = c/n$, where c is the speed of light in a vacuum. The group velocity, which describes the speed of the overall pulse envelope, is

¹Assuming a lossless SMF simplifies the analysis, especially for short transmission distances, which are common in the design of fibre-optical parametric devices. In such cases, the signal attenuation due to fibre loss is negligible. Therefore, the lossless assumption is made to isolate the effects of dispersion, avoiding the additional complexity of signal attenuation in the analysis.

defined as $v_g = d\omega/d\beta$. When different spectral components of the wave travel at different velocities, they arrive at different times. This variation in speed causes a short optical pulse, which contains multiple frequency components, to experience pulse broadening. Chromatic dispersion, which describes the dependence of the optical fibre's response on the frequency ω , is therefore a critical factor in the propagation of modulated signals. To quantify chromatic dispersion effects, we define $\beta^{(m)} = d^m\beta/d\omega^m$, with the group velocity given by $v_g = 1/\beta^{(1)}$. Optical pulse dispersion occurs when $\beta^{(1)}$ changes with frequency ω , i.e. the second-order derivative, $\beta^{(2)} \neq 0$; this parameter is known as group-velocity dispersion (GVD) parameter. When wavelength is considered rather than frequency, the effect is usually quantified by the chromatic dispersion coefficient, which is defined as

$$D = -\frac{2\pi c\beta^{(2)}}{\lambda^2}. \quad (2.2)$$

The wavelength at which $\beta^{(2)} = 0$ is called the zero-dispersion wavelength, denoted as λ_0 . In practice, λ_0 can vary along the fibre length due to variations in the fibre core diameter, typically caused by imperfections in the fabrication process. Moreover, temperature fluctuations can also affect the value of λ_0 . To minimise these variations, shorter fibre lengths are often preferred. Therefore, any reduction in fibre length must be compensated by an increase in pump power². Achieving a high pump power thus becomes a critical aspect of device design, a topic that will be explored in the second half of this chapter.

When the pulse signal propagates at the wavelength close to λ_0 , it experiences minimal distortion. The distortion, when $\beta^{(2)} = 0$, is caused by the higher-order dispersion terms, such as third-order $\beta^{(3)}$ or fourth-order $\beta^{(4)}$ dispersion parameters. These higher-order terms, along with the relative frequency locations of the pump signals concerning λ_0 , determine the phase-matching conditions and thus determine the frequency response of the fibre-optical parametric devices.

2.1.2 Fibre nonlinearity

The response of an optical medium to an intense light wave is highly dependent on the optical intensity I . This nonlinear behaviour, driven by the third-order susceptibility $\chi^{(3)}$, is commonly known as third-order nonlinearity or Kerr nonlinearity. The fibre refractive index can be ex-

²Using shorter fibres necessitates higher pump power to achieve the desired performance. We will see later that the transfer functions of fibre-optical parametric devices are dependent on the product of the fibre length and the pump power.

pressed as [12]

$$n = n_0 + n_2 I, \quad (2.3)$$

where n_0 is the refractive index under weak-field conditions and n_2 is the nonlinear refractive index. For convenience, an alternative parameter, the nonlinear coefficient γ , is often used and is defined as [12]

$$\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}}, \quad (2.4)$$

where A_{eff} is the fibre's effective area, which represents the cross-sectional area over which the optical power is distributed in the fibre core. As discussed later, the nonlinear coefficient γ is a key parameter in determining the maximum gain of fibre-optical parametric devices. Similar to pump power, a higher γ is desirable, because it enables the use of shorter fibre lengths, thereby reducing the longitudinal variation of the zero-dispersion wavelength λ_0 .

In a typical SMF, γ is around $1 - 2 (\text{W} \cdot \text{km})^{-1}$, while in a HNLF, γ is typically in the range of $20 - 30 (\text{W} \cdot \text{km})^{-1}$ [14]. The HNLF achieves this high nonlinear coefficient due to their reduced fibre core radius, which decreases A_{eff} in Eq. (2.4) and the increased n_2 , which is enhanced by incorporating a higher proportion of germanium dioxide in the fibre core material. As a result, HNLF is often used as the nonlinear medium for fibre-optical parametric devices.

Recall that $\beta = 2\pi n/\lambda$, so any modulation of the refractive index n (due to the Kerr effect) leads to a corresponding modulation of the propagation constant β , which in turn causes phase modulation of the optical signal. This phase modulation can occur in two primary forms: self-phase modulation (SPM), where the effect is due to the signal itself, or cross-phase modulation (XPM), where it is influenced by other co-propagating waves. When phase modulation occurs in the presence of multiple optical waves, their interaction generates new frequency components through the process known as FWM. The interaction of typically three optical waves gives rise to a fourth wave, with the energy exchanged across different frequencies. We will discuss the principles of FWM in the context of fibre-optical parametric devices in the following section.

2.2 Fibre-optical parametric amplifier

Unlike conventional optical amplifiers that rely on stimulated emission such as the Erbium-doped fibre amplifier (EDFA), the FOPA leverages the nonlinear properties of fibre to amplify optical signals, offering enhanced capabilities for optical networks. This nonlinear mechanism

enables FOPA to surpass conventional amplifiers in certain key aspects. One of its most notable advantages is its adjustable gain spectrum. The shape of the optical gain spectrum can be tailored by tuning the fibre dispersion properties, allowing bandwidths to extend to several hundred nanometers, far exceeding the 35nm bandwidth of an EDFA[15]. By carefully designing the fibre and selecting the appropriate pump wavelength, the centre frequency can be adjusted to fully exploit this broad gain spectrum [16], whereas EDFA offers only limited tuning capabilities. Additional advantages of FOPA include high gain [17] and a low noise figure. In its phase-insensitive configuration, FOPA can achieve a noise figure below 4 dB, comparable to that of EDFA [12], and, when operated in a phase-sensitive amplification mode, it can approach the quantum limit of sub-3 dB [18]. Furthermore, optical parametric processes in FOPA enable applications beyond amplification, such as wavelength conversion and phase conjugation, which are essential for the design of OPC, as discussed in a later subsection.

In this section, we will explore its fundamental operating principle based on FWM, derive its transfer function and then examine its gain spectra under different pump schemes. In a FOPA setup, a strong pump signal co-propagates with the input signal along a HNLF, as illustrated in Fig. 2.1. The nonlinear interaction between the intense pump wave and the signal wave generates new frequency components, resulting in the amplification of the signal. After this interaction, an optical filter is used to extract the amplified signal from the other generated components.

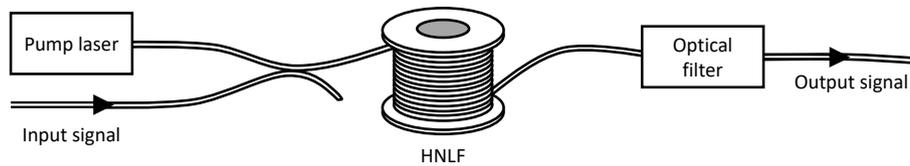


Figure 2.1: General scheme of fibre-optical parametric amplifier.

A crucial assumption for the following analysis is that all interacting waves share the same state of polarisation and maintain it throughout the entire fibre length. We begin by examining the non-degenerate FOPA scenario, illustrated in Fig. 2.2(a), where two distinct pump waves at frequencies ω_1 and ω_2 are present along with a signal at frequency ω_3 . As a result of the FWM process, a new wave, known as the idler, emerges at frequency ω_4 , satisfying the relation $\omega_1 + \omega_2 = \omega_3 + \omega_4$. In terms of photon exchange, two photons at pump frequencies combine to amplify the signal photon, while simultaneously generating the idler photon. Introducing the central frequency $\omega_c = (\omega_1 + \omega_2)/2$, this relation implies that the signal ω_3 and idler ω_4

are symmetric with respect to ω_c . The degenerate FOPA scenario, depicted in Fig. 2.2(b), represents a special case of this interaction where both pumps have identical frequencies, i.e. $\omega_1 = \omega_2 = \omega_c$. In this case, the pump loses a photon pair to the signal and the idler.

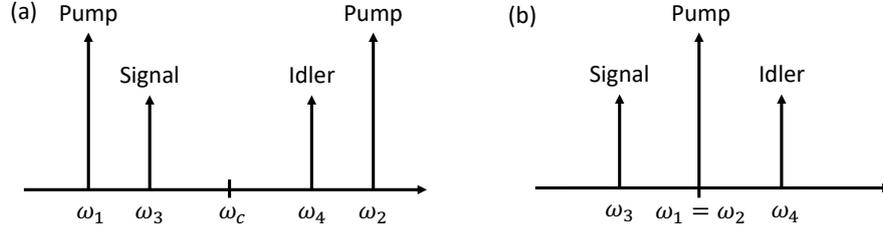


Figure 2.2: Diagram of frequency locations in four-wave mixing in two scenarios: (a) dual-pump FOPA and (b) single-pump FOPA.

The phase-matching condition, which ensures momentum conservation, is given by $\Delta\beta \triangleq \beta_3 + \beta_4 - \beta_1 - \beta_2 = 0$, where β_k represents the wavevector of the wave at frequency ω_k . In the single-pump scenario, the linear phase mismatch simplifies to $\Delta\beta \triangleq \beta_3 + \beta_4 - 2\beta_1 = 0$. Since FWM is a phase-sensitive process, satisfying the phase-matching condition is crucial for achieving efficient parametric amplification.

Assuming negligible GVD effects on the signal envelopes and a constant nonlinear coefficient γ applied to all nonlinear interactions, the propagation of four interacting waves along the nonlinear fibre in the two-pump FOPA scenario can be described as follows [12]

$$\frac{dA_1}{dz} = j\gamma \left[|A_1|^2 A_1 + 2 \sum_{i=1, i \neq 1}^4 |A_i|^2 A_1 + 2A_3 A_4 A_2^* e^{j\Delta\beta z} \right], \quad (2.5)$$

$$\frac{dA_2}{dz} = j\gamma \left[|A_2|^2 A_2 + 2 \sum_{i=1, i \neq 2}^4 |A_i|^2 A_2 + 2A_3 A_4 A_1^* e^{j\Delta\beta z} \right], \quad (2.6)$$

$$\frac{dA_3}{dz} = j\gamma \left[|A_3|^2 A_3 + 2 \sum_{i=1, i \neq 3}^4 |A_i|^2 A_3 + 2A_1 A_2 A_4^* e^{-j\Delta\beta z} \right], \quad (2.7)$$

$$\frac{dA_4}{dz} = j\gamma \left[|A_4|^2 A_4 + 2 \sum_{i=1, i \neq 4}^4 |A_i|^2 A_4 + 2A_1 A_2 A_3^* e^{-j\Delta\beta z} \right], \quad (2.8)$$

where A_i is the slowly varying envelope of the wave at frequency ω_i . In the single-pump FOPA scenario, a special case where $\omega_1 = \omega_2$, we can substitute A_2 in Eq. (2.5) with A_1 . Consequently, the equations for the third and fourth waves are given by [12]

$$\frac{dA_3}{dz} = j\gamma \left[|A_3|^2 A_3 + 2 \sum_{i=1, i \neq 3}^4 |A_i|^2 A_3 + (A_1)^2 A_4^* e^{-j\Delta\beta z} \right], \quad (2.9)$$

$$\frac{dA_4}{dz} = j\gamma \left[|A_4|^2 A_4 + 2 \sum_{i=1, i \neq 4}^4 |A_i|^2 A_4 + (A_1)^2 A_3^* e^{-j\Delta\beta z} \right]. \quad (2.10)$$

In each of the equations of Eq. (2.5) to Eq. (2.10), the first term on the right-hand side represents the SPM induced by the wave itself, while the second represents the XPM induced by other co-propagating waves. The final term in each above equation represents the FWM effect, which involves the generation of a fourth wave due to the interaction among three other waves.

2.2.1 Dual-pump scheme

In this section, we analyse the equations from Eq. (2.5) to Eq. (2.8), which describe the non-degenerate FOPA involving two distinct pump waves. By solving these equations for A_3 , we derive the transfer function of the FOPA, defined as the ratio of the output signal to the input signal, allowing us to plot the FOPA gain spectrum. We then discuss the selection of key parameters to achieve the desired power gain and a flat-top gain profile.

In the absence of pump depletion, the two pump powers are assumed to remain constant, i.e. $P_1(z) = P_1$ and $P_2(z) = P_2$. Consequently, the signal and idler powers are much lower compared to the pump levels. We also assume that the initial phases of all pump waves are equal to zero, i.e. $A_1(0) = \sqrt{P_1}$ and $A_2(0) = \sqrt{P_2}$. In this case, the SPM and XPM terms in Eqs. (2.5) and (2.6) dominate over the FWM terms. The propagation equations of the pump waves can be simplified to

$$\frac{dA_1}{dz} = j\gamma(P_1 + 2P_2)A_1, \quad (2.11)$$

$$\frac{dA_2}{dz} = j\gamma(P_2 + 2P_1)A_2. \quad (2.12)$$

Solving these equations yields the solutions $A_1(z) = A_1(0)e^{j\gamma(P_1+2P_2)z}$ and $A_2(z) = A_2(0)e^{j\gamma(P_2+2P_1)z}$. For Eqs. (2.7) and (2.8), which correspond to the signal and idler, we assume no pump depletion and neglect the SPM terms and the XPM terms between the signal A_3 and the idler A_4 . This simplification leads to

$$\frac{dA_3}{dz} = jp_3A_3 + jr_3A_4^*e^{jq_3z}, \quad (2.13)$$

$$\frac{dA_4}{dz} = jp_4A_4 + jr_4A_3^*e^{jq_4z}, \quad (2.14)$$

where $p_3 = p_4 = 2\gamma P_0$, $r_3 = r_4 = \gamma P_0$ and $q_3 = q_4 = 3\gamma P_0 - \Delta\beta$. It is often convenient to denote the total pump power as $P_0 = P_1 + P_2$, assuming a symmetric pump condition where $P_1 = P_2$. Since our focus is on the signal at the FOPA output, we will temporarily leave Eq. (2.14), which describes the idler. By defining the total wavevector mismatch as

$$\kappa \triangleq p_3 + p_4^* - q_3 = \Delta\beta + \gamma P_0, \quad (2.15)$$

and introducing the variable transformation $C_3 = A_3e^{j(\frac{\kappa}{2}-p_3)z}$, Eq. (2.13) simplifies to [12]

$$\frac{d^2C_3}{dz^2} - g^2C_3 = 0, \quad (2.16)$$

where the parametric gain coefficient g is defined as

$$g = \sqrt{(\gamma P_0)^2 - \left(\frac{\kappa}{2}\right)^2}. \quad (2.17)$$

We observe that κ in Eq. (2.15) arises from the linear wavevector mismatch, $\Delta\beta$, as well as the second term, $\Delta\beta_{NL} \triangleq \gamma P_0$, which can be interpreted as the nonlinear wavevector mismatch. The general solution of Eq. (2.16) has the form

$$C_3 = M_3e^{gz} + N_3e^{-gz}, \quad (2.18)$$

where M_3 and N_3 are determined by the initial conditions. To find them, we first take the derivative of the above equation and obtain

$$\frac{dC_3}{dz} = g(M_3e^{gz} - N_3e^{-gz}). \quad (2.19)$$

Using Eqs. (2.18) and (2.19) and letting $z = 0$, we solve the coefficients

$$M_3 = \frac{1}{2} \left[C_3(0) + \frac{1}{g} \frac{dC_3(0)}{dz} \right], \quad (2.20)$$

$$N_3 = \frac{1}{2} \left[C_3(0) - \frac{1}{g} \frac{dC_3(0)}{dz} \right]. \quad (2.21)$$

At length $z = 0$, we have $C_3(0) = A_3(0)$ and the derivative

$$\frac{dC_3(0)}{dz} = j\frac{\kappa}{2}A_3(0) + jr_3A_4^*(0). \quad (2.22)$$

By assuming $A_4(0) = 0$ (no idler at the fibre input), we can rewrite the coefficients in Eqs. (2.20) and (2.21) as

$$M_3 = \frac{A_3(0)}{2} \left(1 + j \frac{\kappa}{2g} \right), \quad (2.23)$$

$$N_3 = \frac{A_3(0)}{2} \left(1 - j \frac{\kappa}{2g} \right). \quad (2.24)$$

Substituting them into Eq. (2.18) and using the definitions $\cosh(x) = (e^x + e^{-x})/2$ and $\sinh(x) = (e^x - e^{-x})/2$, we obtain the final solution as [12]

$$C_3 = A_3(0) \cosh(gz) + j \frac{\kappa}{2g} A_3(0) \sinh(gz) \quad (2.25)$$

The FOPA complex parametric signal gain, defined as the ratio between $A_3(L)$ and $A_3(0)$ with L is the HNLFF length, is thus calculated as

$$\mu_s \triangleq \frac{A_3(L)}{A_3(0)} = \left[\cosh(gL) + j \frac{\kappa}{2g} \sinh(gL) \right] \exp \left[j \left(2\gamma P_0 - \frac{\kappa}{2} \right) L \right], \quad (2.26)$$

The signal power gain is given by $G_s = |\mu_s|^2$ and reaches its maximum value when the total wavevector mismatch $\kappa = 0$. The maximum parametric gain coefficient is $g_{\max} = \gamma P_0$, obtained by setting $\kappa = 0$ in Eq. (2.17). The corresponding maximum power gain is

$$G_{s,\max} = |\cosh(g_{\max}L)|^2 = |\cosh(\Phi_{\text{NL}})|^2, \quad (2.27)$$

where $\Phi_{\text{NL}} \triangleq \gamma P_0 L$ refers to the nonlinear phase shift. Equation (2.26) also indicates that the shape of the FOPA gain spectrum is modulated by κ , which depends on the linear wavevector mismatch $\Delta\beta$. The fibre dispersion parameters, namely $\beta^{(3)}$ and $\beta^{(4)}$, as well as the relative location of λ_0 with respect to the pump frequencies, contribute to this term.

We now focus on studying the gain spectrum of the FOPA. It is often convenient to define the signal frequency detuning $\Delta\omega_s$ of the signal at ω_3 from the central frequency ω_c as $\Delta\omega_s \triangleq \omega_3 - \omega_c = \omega_c - \omega_4$. Similarly, the pump frequency detuning $\Delta\omega_p$ of the pump waves at ω_1 and ω_2 from ω_c is given by $\Delta\omega_p \triangleq \omega_1 - \omega_c = \omega_c - \omega_2$. In some cases, the signal wavelength detuning, defined as $\Delta\lambda_s = \lambda_3 - \lambda_c = \lambda_c - \lambda_4$, is also used. The relationship between $\Delta\omega_s$ and $\Delta\lambda_s$ can be expressed as

$$\Delta\omega_s = -\frac{2\pi c \Delta\lambda_s}{\lambda_c(\lambda_c + \Delta\lambda_s)} \quad (2.28)$$

From the definition of $\Delta\beta = \beta_3 + \beta_4 - \beta_1 - \beta_2$ and with the above definitions of $\Delta\omega_s$ and $\Delta\omega_p$, we can write

$$\Delta\beta = \beta(\omega_c + \Delta\omega_s) + \beta(\omega_c - \Delta\omega_s) - \beta(\omega_c + \Delta\omega_p) - \beta(\omega_c - \Delta\omega_p). \quad (2.29)$$

Using the power series expansion in terms of $\Delta\omega_s$ and $\Delta\omega_p$, we observe that the odd-order derivative terms cancel each other out. Finally, we obtain

$$\Delta\beta = 2 \sum_{m=0}^{\infty} \frac{\beta^{(2m)}}{(2m)!} [(\Delta\omega_s)^{2m} - (\Delta\omega_p)^{2m}]. \quad (2.30)$$

The linear wavevector mismatch $\Delta\beta$ depends only on the even-order derivatives of β at ω_c . It is typically practical to truncate higher-order terms in Eq. (2.30) with $m > 2$. This leads to the two-term model of $\Delta\beta$ as [12]

$$\Delta\beta = \beta^{(2)} [(\Delta\omega_s)^2 - (\Delta\omega_p)^2] + \frac{\beta^{(4)}}{12} [(\Delta\omega_s)^4 - (\Delta\omega_p)^4]. \quad (2.31)$$

Here, β_2 can be approximated as follows when λ_c is adjusted close to λ_0

$$\beta^{(2)} = \beta^{(3)}(\omega_c - \omega_0) + \frac{\beta^{(4)}}{2}(\omega_c - \omega_0)^2. \quad (2.32)$$

Given the pump frequencies ω_1 and ω_2 , which determine $\Delta\omega_p$, we can observe that $\Delta\beta$ is an even function of $\Delta\omega_s$, according to Eq. (2.30) or (2.31). Consequently, the graphs of the total wavevector mismatch κ , the parametric gain coefficient g and the complex parametric signal gain μ_s versus $\Delta\omega_s$ are symmetric with respect to ω_c , as described by Eqs. (2.15), (2.17) and (2.26), respectively. However, their graphs versus $\Delta\lambda_s$ are not symmetric due to Eq. (2.28), which makes them functions of not only $(\Delta\lambda_s)^2$, but also $\Delta\lambda_s$. Nevertheless, this asymmetry will not be very pronounced unless the FOPA has a very wide gain bandwidth.

Optimising the FOPA gain is a crucial step in the design process, where the goal is to achieve a relatively flat and wide gain spectrum. Since the FOPA gain, as described by Eq. (2.26), is determined by the parametric gain coefficient g , the optimisation process involves selecting appropriate parameters to produce a flat-top curve of g versus $\Delta\lambda_s$ or $\Delta\omega_s$. The relationship between g and $\Delta\lambda_s$ or $\Delta\omega_s$ is determined by the linear wavevector mismatch $\Delta\beta$, as shown in Eqs. (2.15), (2.17) and (2.31). We plotted their relationship in Figs. 2.3 and 2.4, based on [12].

We begin by examining the curve of the linear wavevector mismatch $\Delta\beta$, as a function of the squared frequency detuning $(\Delta\omega_s)^2$ using Eq. (2.31). Assuming $\beta^{(2)} < 0$, this curve is parabolic when $\beta^{(4)} > 0$, as illustrated in Fig. 2.3(a), and semi-parabolic when $\beta^{(4)} < 0$, as depicted in Fig. 2.4(a). The initial value in both cases, at $\Delta\omega_s = 0$, denoted as $\Delta\beta'$, is

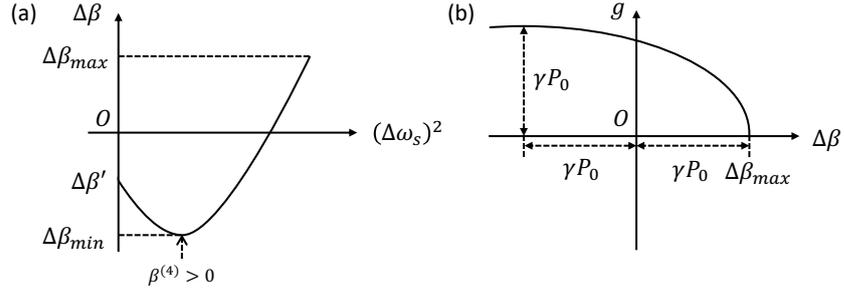


Figure 2.3: Theoretical graphs in the FOPA dual-pump scenario with $\beta^{(4)} > 0$ of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$.

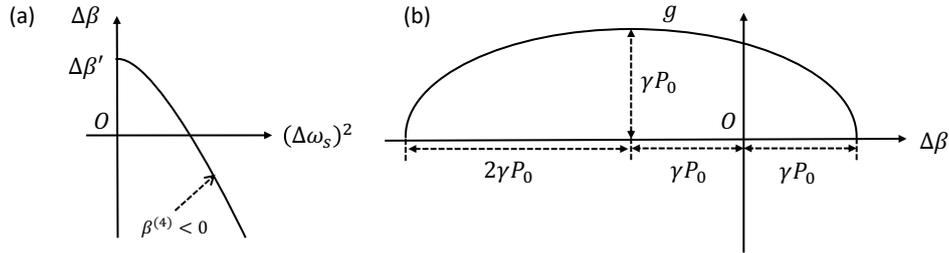


Figure 2.4: Theoretical graphs in the FOPA dual-pump scenario with $\beta^{(4)} < 0$ of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$.

influenced not only by these dispersion parameters but also by the pump frequency detuning, $\Delta\omega_p$, i.e.

$$\Delta\beta' = -\beta^{(2)}(\Delta\omega_p)^2 - \frac{\beta^{(4)}}{12}(\Delta\omega_p)^4. \quad (2.33)$$

When $\beta^{(4)} > 0$, it is important to note that only values of $\Delta\beta$ that result in a defined g in Eq. (2.17), where the expression under the square root must be greater than or equal to zero, will be considered. This constraint implies that $\Delta\beta$ cannot exceed $\Delta\beta_{max} = \gamma P_0$, where the parametric gain coefficient g reaches zero. The curve reaches its minimum at $(\Delta\omega_s)^2 = -6\beta^{(2)}/\beta^{(4)}$, and this minimum value is

$$\Delta\beta_{min} = \beta^{(2)} \left(\left(-\frac{6\beta^{(2)}}{\beta^{(4)}} \right) - (\Delta\omega_p)^2 \right) + \frac{\beta^{(4)}}{12} \left(\left(-\frac{6\beta^{(2)}}{\beta^{(4)}} \right)^2 - (\Delta\omega_p)^4 \right). \quad (2.34)$$

When $\beta^{(4)} < 0$, there is no minimum value for $\Delta\beta$. The maximum value of $\Delta\beta$ is determined by the smaller of $\Delta\beta'$ or $\Delta\beta_{max}$.

Next, we consider the graphs of g versus $\Delta\beta$, as given by Eq. (2.17), shown in Figs.

2.3(b) and 2.4(b). The curves exhibit elliptical shapes. When $\beta^{(4)} < 0$, we consider the entire ellipse, as $\Delta\beta$ can become arbitrary negative. We recall that the maximum value of g occurs at $g_{\max} = \gamma P_0$ when $\kappa = 0$ or $\Delta\beta = -\gamma P_0$. The minimum value, which drops to zero, is reached when $\Delta\beta$ approaches $\Delta\beta_{\max}$ or $-3\gamma P_0$.

By examining both graphs in Fig. 2.3 (when $\beta^{(4)} > 0$) simultaneously, we observe that as the squared signal frequency detuning $(\Delta\omega_s)^2$ increases from zero, the parametric gain coefficient g transitions from $g(\Delta\beta')$ to $g(\Delta\beta_{\min})$, eventually dropping to zero or $g(\Delta\beta_{\max})$. In contrast, considering both graphs in Fig. 2.4 (when $\beta^{(4)} < 0$), we see that g starts from $\Delta\beta'$ and decreases quadratically. Because $\Delta\beta$ decreases monotonically, it is challenging to maintain values that yield results close to g_{\max} over a wide range of $(\Delta\omega_s)^2$.

Achieving a high value of g over a broad range of signal frequency detuning is important for optimising the FOPA gain spectrum. In this case, the scenario with $\beta^{(4)} > 0$ offers more flexibility for achieving this goal. The dispersion parameters $\beta^{(2)}$ and $\beta^{(4)}$, along with the pump frequency detuning $\Delta\omega_p$ in Eqs. (2.33) and (2.34), should be selected so that $\Delta\beta'$ and $\Delta\beta_{\min}$ in Fig. 2.3(a) are close to $-\gamma P_0$. This ensures that the gain remains high across the desired range of signal detuning.

Parameters	Values
λ_0	1560.7 [nm]
λ_c	1563.7 [nm]
$\Delta\omega_p$	45×10^{-9} [rad/s]
$\beta^{(3)}$	1.09×10^{-41} [s^3m^{-1}]
$\beta^{(4)}$	3.85×10^{-55} [s^4m^{-1}]
γ	10 [$\text{W}^{-1}\text{km}^{-1}$]
L	200 [m]
P_0	1.785 [W]

Table 2.1: Dual-pump scheme parameters

A set of FOPA parameters with $\beta^{(4)} > 0$ that will be used for the dual-pump FOPA scheme throughout this dissertation is listed in Table 2.1. First, the values of the nonlinear coefficient γ , the fibre length L and the pump power P_0 are selected to achieve a maximum power gain $G_{s,\max}$ of 25 dB. Based on the optimisation process outlined earlier, our choice of other parameters yields $\Delta\beta' \approx -0.012 \text{ m}^{-1}$ and $\Delta\beta_{\min} \approx -0.017 \text{ m}^{-1}$, while $-\gamma P_0 \approx -0.018 \text{ m}^{-1}$, aligning well with the selection criteria discussed. The ability to optimise these values close to each other is due to $\Delta\omega_p \neq 0$, as is the case in the dual-pump scheme. The involvement of $\Delta\omega_p$ shifts $\Delta\beta'$ away from zero, allowing it to be adjusted appropriately. As we will see later, this property

cannot be achieved in the single-pump scheme, making the dual-pump scheme more effective for achieving a flat-top FOPA gain spectrum.

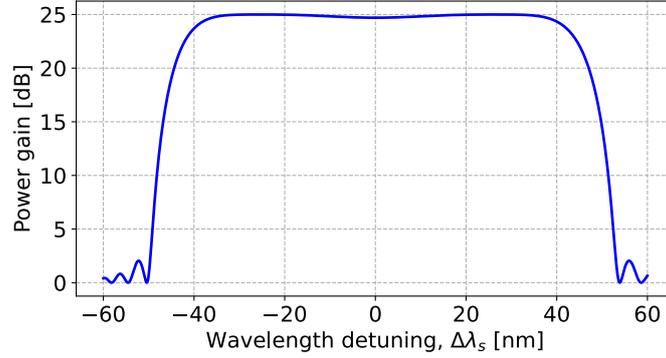


Figure 2.5: FOPA gain spectrum in the dual-pump scheme.

The FOPA gain spectrum, based on the chosen parameters, is shown in Fig. 2.5. Since the graph is plotted against the wavelength detuning, $\Delta\lambda_s$, a slight asymmetry is observed, as anticipated in our earlier discussion. The result shows a reasonably flat spectrum with a peak gain around 25 dB within the wavelength detuning range of -40 nm to 40 nm. Outside this region, the FOPA gain drops quickly due to the failure to satisfy the phase-matching condition. This broad, flat spectrum is highly beneficial for applications that require a wide wavelength operating range.

2.2.2 Single-pump scheme

We will apply a similar approach to analyse the single-pump scheme, where the propagation of the interacting waves is governed by Eqs. (2.5), (2.6) and (2.9). In this section, we will derive the transfer function for FOPA in the single-pump scheme, followed by a discussion on the optimisation method for the gain profile.

Similar to the dual-pump scheme, the assumption of negligible pump depletion allows us to neglect the FWM term in Eq. (2.5). In the single-pump scheme, this simplification is extended further by only considering the wave associated with A_1 . This means that only the SPM term in Eq. (2.5) is relevant, leading to the simplified expression

$$\frac{dA_1}{dz} = j\gamma P_1 A_1. \quad (2.35)$$

Since there is only a single pump, Eq. (2.6) does not apply. Solving the differential equation above yields $A_1(z) = A_1(0)e^{j\gamma P_1 z}$. The propagation equations for the signal and idler are

analogous to Eqs. (2.13) and (2.14) with modifications: $p_3 = p_4 = 2\gamma P_1$, $r_3 = r_4 = \gamma P_1$ and $q_3 = q_4 = 2\gamma P_1 - \Delta\beta$. Following the same procedure, the complex parametric gain μ_s in the single-pump FOPA scheme is identical to that in Eq. (2.26) with P_1 replacing P_0 . This indicates that the same power gain can be achieved if $P_1 = P_0$, but the total wavevector mismatch is

$$\kappa \triangleq p_3 + p_4^* - q_3 = \Delta\beta + 2\gamma P_1, \quad (2.36)$$

which is different from Eq. (2.15) in the dual-pump scheme. Therefore, although the single-pump scheme achieves the same maximum power gain as the dual-pump scheme, it produces a different shape for the FOPA gain spectrum. The optimal operating wavelength, which yields the maximum gain, occurs when $\Delta\beta = -2\gamma P_1$ in the single-pump scheme, compared to $\Delta\beta = -\gamma P_0$ in the dual-pump scheme. By expressing the relationship between the linear wavevector mismatch $\Delta\beta$ and the fibre dispersion, and following a similar procedure to the dual-pump scheme while considering the two-term model, we can derive

$$\Delta\beta = \beta^{(2)}(\Delta\omega_s)^2 + \frac{\beta^{(4)}}{12}(\Delta\omega_s)^4. \quad (2.37)$$

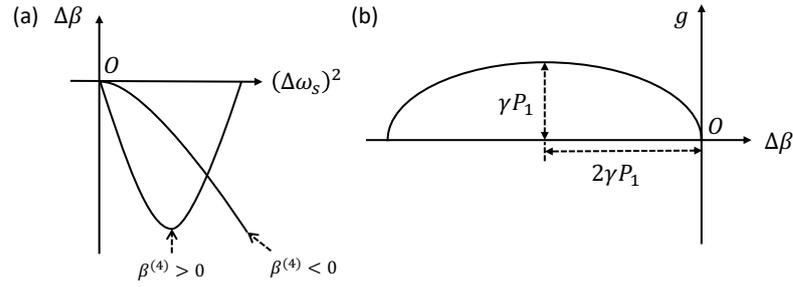


Figure 2.6: Theoretical graphs in the FOPA single-pump scenario of (a) Linear wavevector mismatch $\Delta\beta$ versus squared frequency detuning $(\Delta\omega_s)^2$ and (b) Parametric gain coefficient g versus linear wavevector mismatch $\Delta\beta$.

We plotted Fig. 2.6, based on [12], to examine the relationship between relevant quantities in the single-pump scenario. First, we analyse the relationship between $\Delta\beta$ and $(\Delta\omega_s)^2$, as given by Eq. (2.37), shown in Fig. 2.6(a). Assuming again that $\beta^{(2)} < 0$, the curvature of the $\Delta\beta$ graph depends on the sign of $\beta^{(4)}$. As $(\Delta\omega_s)^2$ increases, the graph exhibits a parabolic trajectory, with a global minimum present when $\beta^{(4)} > 0$, while it forms a semi-parabolic curve with its minimum at the origin when $\beta^{(4)} < 0$. The graph of the parametric gain coefficient g as a function of $\Delta\beta$ is presented in Fig. 2.6(b). This graph illustrates an elliptical shape, where $g_{max} = \gamma P_1$ occurs at $\Delta\beta = -2\gamma P_1$, and g approaches zero as $\Delta\beta$ reaches either zero or $-4\gamma P_1$. For optimal parametric amplification, $\Delta\beta$ must approach $-2\gamma P_1$ in Fig. 2.6(a) to

achieve the maximum g in Fig. 2.6(b). To achieve this desired behaviour, we select $\beta^{(4)} < 0$ in this case, as this ensures that the curve descends to the appropriate minimum value. Conversely, if $\beta^{(4)} > 0$, careful selection of both $\beta^{(2)}$ and $\beta^{(4)}$ is required to ensure that the global minimum of the $\Delta\beta$ curve lies below $-2\gamma P_1$.

Parameters	Values
λ_0	1560.7 [nm]
λ_p	1563.7 [nm]
$\beta^{(3)}$	1.09×10^{-41} [s^3m^{-1}]
$\beta^{(4)}$	-2.85×10^{-55} [s^4m^{-1}]
γ	10 [$\text{W}^{-1}\text{km}^{-1}$]
L	200 [m]
P	1.785 [W]

Table 2.2: Single-pump scheme parameters

The initial value $\Delta\beta'$, defined analogously to the dual-pump scheme, is zero regardless of the sign of $\beta^{(4)}$. As $(\Delta\omega_s)^2$ increases, $\Delta\beta'$ decreases quadratically. This behaviour indicates significant fluctuations in the parametric gain coefficient g and, consequently, in the FOPA power gain as $(\Delta\omega_s)^2$ increases. In the absence of $\Delta\omega_p$, it is impossible to shift $\Delta\beta'$ closer to the minimum value of $\Delta\beta$, thus preventing the realisation of a flat-top gain spectrum, as is achieved in the dual-pump scheme.

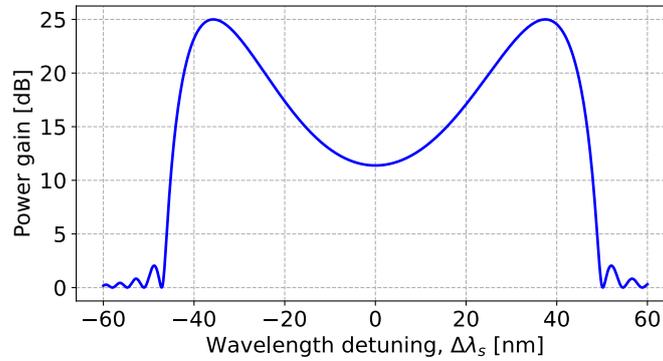


Figure 2.7: FOPA gain spectrum in the single-pump scheme.

The FOPA gain spectrum, calculated using the parameters listed in Table. 2.2, is shown in Fig. 2.7. In this case, a negative value of $\beta^{(4)}$ has been selected, as discussed earlier. The parameters were also chosen to achieve a power gain of 25 dB, consistent with the dual-pump scheme. As shown in Fig. 2.7, the power gain initially increases with the wavelength detuning, reaching an optimal level in line with our previous discussion. The spectrum exhibits two peaks on either side, before declining rapidly as the phase-matching condition is no longer satisfied.

2.3 Optical phase conjugation

Having explored the FOPA in both dual-pump and single-pump configurations and their respective gain spectra, we now shift our focus to another critical application of the FWM process: OPC. While FOPA has proven effective for signal amplification and wavelength conversion, FOPA offers a complementary approach that specifically addresses the challenges posed by fibre dispersion and nonlinearity in optical transmission systems.

In this section, we will begin with a briefly overview of key milestones in the development of OPC, highlighting its fundamental concept. Following this, we will continue our analysis of FWM to derive the transfer function for OPC, which specifically accounts for the idler signal. Finally, we will discuss the mathematical reason behind why OPC can play an important role in mitigating signal distortions in the optical transmission.

An OPC is an optical device that generates a phase-conjugated replica of an incoming optical signal [19]. Initially proposed as a method for compensating fibre dispersion [20, 21], the use of OPC has since evolved to mitigate nonlinear effects in optical networks [22, 23, 24, 25]. While other compensation techniques, such as digital-back-propagation (DBP) [26, 27], have been explored to address channel impairments, they often face significant challenges, including the complexity of DSP. This has prompted ongoing research into simplifying the complexity [28]. In contrast, mid-link OPC has attracted considerable attention as an elegant and effective solution for improving channel capacity. By passing the distorted signal through an OPC, signal distortion experienced in the first fibre can be reversed as the phase-conjugated signal propagates along the second fibre with similar characteristics. With OPC, a single device can simultaneously counteract both dispersion and nonlinearity, enabling the transmission of high bandwidth signals and supporting advanced modulation formats [29].

The initial implementation concept of OPC was based on backward FWM, where the generated idler wave travels in the opposite direction to the signal. However, it was later demonstrated that phase-conjugated waves could also be generated in the forward direction with fibre-based FWM [30]. This approach was further validated by experimental works in [21, 31]. The setup for fibre-based OPC is similar to that shown in Fig. 2.1, with the main difference being that the selected output signal is the idler at ω_4 in Fig. 2.2. A simplified diagram demonstrating the generation of the phase-conjugated wave is shown in Fig. 2.8.

To derive the transfer function of OPC, we begin with Eq. (2.14), which governs the prop-

agation of the idler wave. By performing a variable change, we set $C_4 = A_4 e^{j(\frac{\kappa}{2} - p_4)z}$. This substitution transforms the equation into a form similar to Eq. (2.16), but with C_4 replacing C_3 . The final solution, with the same assumption of $A_4(0) = 0$, is expressed as

$$C_4 = j \frac{\gamma P_0}{g} A_3^*(0) \sinh(gz). \quad (2.38)$$

The complex idler signal gain, defined as the ratio between $A_4(L)$ and $A_3^*(0)$, is [12]

$$\mu_i \triangleq \frac{A_4(L)}{A_3^*(0)} = j \frac{\gamma P_0}{g} \sinh(gL) \exp \left[j \left(2\gamma P_0 - \frac{\kappa}{2} \right) L \right], \quad (2.39)$$

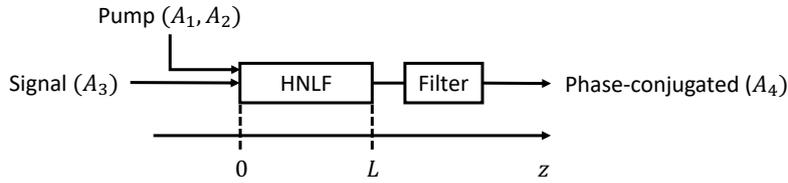


Figure 2.8: Schematic of phase-conjugated wave generation in FWM.

We thus obtain a complex conjugate version of the input signal $A_3(0)$ at the output of the fibre. In general, if the amplification sufficiently compensates for the signal attenuation due to fibre loss, we can obtain the complex conjugate version $A^*(z, t)$ of the transmitted signal $A(z, t)$. The OPC can mitigate distortions when deployed mid-span in optical transmission systems, as depicted in Fig. 2.9. This is achieved through the generation of the phase-conjugated wave, which has an inverse spectral profile, a technique known as mid-span spectral inversion [32]. In Fig. 2.9, after propagation along the first fibre segment, the signal's spectrum becomes distorted and broadened. The mid-link OPC generates a wave at a new frequency, symmetric to the original one relative to the central frequency ω_c . This new wave has an inverted spectrum compared to the pre-OPC signal. The highest frequency components become the lowest ones, and if they travel faster in the first half of the link, they will now propagate more slowly in the second half, assuming both fibres have the similar characteristics. Therefore, the chromatic dispersion caused by two fibres can be compensated at the transmission output. Regarding Kerr nonlinearity, the signal exhibits intensity-dependent phase shifts in the first half of the link followed by equivalent negative phase shifts in the second half. Therefore, spectral inversion reverses both signal distortions, resulting in an undistorted signal at the transmission output.

Additionally, the ability of OPC to simultaneously compensate for chromatic dispersion and

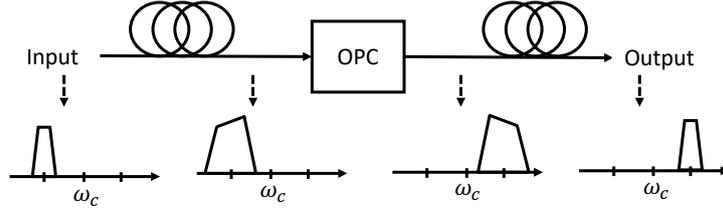


Figure 2.9: Schematic of mid-link OPC transmission and spectrum changes along the link.

Kerr nonlinearity can be explained within the framework of the nonlinear Schrodinger equation (NLSE) [13], which describes optical pulse propagation along a nonlinear fibre, i.e.

$$\frac{\partial A}{\partial z} + \frac{j\beta^{(2)}}{2} \frac{\partial^2 A}{\partial t^2} - j\Gamma|A|^2 A = 0, \quad (2.40)$$

where Γ is the transmission fibre nonlinear coefficient. Equation (2.40) demonstrates both dispersion and nonlinearity effects, which cause pulse broadening and phase modulation, respectively. In the transmission system with length L , the phase conjugate wave $A^*(z, t)$ is generated at $z = L/2$ and its propagation in the second half can be described by substituting $A^*(z, t)$ into Eq. (2.40). By taking the complex conjugate of both sides, we obtain

$$\frac{\partial A}{\partial z} - \frac{j\beta^{(2)}}{2} \frac{\partial^2 A}{\partial t^2} + j\Gamma|A|^2 A = 0, \quad (2.41)$$

This shows that the phase conjugate wave propagates in a manner that reverses both distortions experienced by the original wave, thanks to the symmetry of the OPC transmission. OPC is thus expected to become an essential tool in long-haul optical transmission, where restoring the signal quality can be theoretically achieved.

2.4 SBS suppression techniques

The performance of the optical parametric processes heavily depends on the optical pump power sent to the HNLFF. However, this power scaling is significantly constrained by a nonlinear process known as SBS [13]. SBS causes substantial pump power loss through back-scattering, effectively imposing a threshold on the amount of power that can be used for the parametric amplification process. In this section, we will first cover the fundamentals of SBS effect, establishing a foundation for understanding the approaches used to mitigate it. We will then discuss various suppression techniques, including longitudinal variations in fibre properties and the introduction of phase modulation to pump sources, with the primary focus on the latter. Finally,

we will explore how to design an effective pump-phase modulation scheme for optimal SBS suppression.

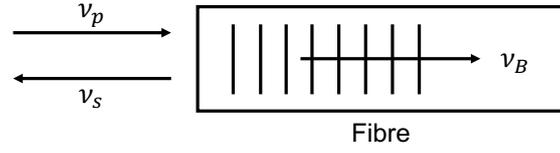


Figure 2.10: Schematic diagram about stimulated Brillouin scattering. The pump light at ν_p scatters from Bragg diffraction induced by the acoustic wave at ν_B to generate the Stokes wave at $\nu_s = \nu_p - \nu_B$.

The optical pump field induces temporal and spatial variations in the fibre's density through the process of electrostriction [33]. This density modulation generates an acoustic wave, which then leads to a periodic modulation of the refractive index within the fibre. Consequently, the pump light at frequency ν_p undergoes scattering via Bragg diffraction, resulting in the generation of a Stokes wave with a frequency ν_s . As depicted in Fig. 2.10, the frequency of the Stokes wave is downshifted due to the Doppler effect, as the grating created by the acoustic wave moves away from the fibre's input. The Brillouin frequency shift, defined as the difference between the pump and Stokes frequencies, $\nu_p - \nu_s$, is given by

$$\nu_B = \frac{2n_p v_A}{\lambda_p}, \quad (2.42)$$

where λ_p is the pump wavelength, n_p is the refractive index at λ_p and v_A is the acoustic wave velocity. In silica fibres, the Brillouin frequency shift at the pump wavelength of $\lambda_p = 1550$ nm is approximately 11.15 GHz. The scattered light interacts with the pump field, amplifying the acoustic wave responsible for the scattering process. This mutual reinforcement, driven by energy conservation, results in the depletion of the pump power. Therefore, SBS emerges as a fundamental challenge in the design of OPC and FOPA.

The Brillouin gain exhibits a Lorentzian spectral profile, with its bandwidth, $\Delta\nu_B$, being inversely proportional to the phonon lifetime. Specifically, $\Delta\nu_B$ is defined as $\Delta\nu_B = \Gamma_B/(2\pi)$, where Γ_B represents the phonon decay rate. In optical fibres, the acoustic phonon lifetime is typically on the order of 10 ns, resulting in a narrow Brillouin gain bandwidth, typically ranging from 10 to 50 MHz. The Brillouin peak gain, g_B , which occurs at the frequency ν_B , depends on $1/\Gamma_B$ and is thus inversely proportional to the Brillouin gain bandwidth $\Delta\nu_B$. When the spectral width of the pump, $\Delta\nu_p$, exceeds $\Delta\nu_B$, the Brillouin gain is significantly reduced by a factor of $1 + \Delta\nu_p/\Delta\nu_B$. Under these conditions, the SBS threshold is given by [34]

$$P_{\text{th}} = \frac{21kA_{\text{eff}}}{g_B L_{\text{eff}}} \left(1 + \frac{\Delta\nu_p}{\Delta\nu_B} \right), \quad (2.43)$$

where k denotes the pump polarisation factor, which ranges between 1 and 2, while A_{eff} and L_{eff} are the effective fibre area and effective fibre length respectively. Increasing the SBS threshold or effectively suppressing SBS effect becomes a crucial task for the design of efficient OPC and FOPA devices. In the following, we will discuss various existing techniques for SBS suppression.

2.4.1 SBS suppression through nonuniform Brillouin frequency shift

The first approach to increasing the SBS threshold P_{th} involves reducing the Brillouin peak gain g_B in Eq. (2.43). This can be achieved by introducing variations in the Brillouin frequency shift ν_B along the length of the fibre, thereby inducing a frequency mismatch. As a result, the Stokes waves generated at different points within the fibre are not perfectly aligned, disrupting the phase-matching condition. This lack of constructive interference leads to a reduced overall amplitude of the Stokes wave. According to Eq. (2.42), ν_B depends on the refractive index n_p and the acoustic velocity v_A . Modifying either of these parameters along the fibre length requires changes to the fibre design and fabrication, forming a class of techniques aimed at controlling the SBS threshold.

Temperature variations can affect both the fibre's refractive index and acoustic velocity. Since the Brillouin gain is linearly dependent on temperature [35], introducing temperature gradients along the fibre can effectively increase the SBS threshold. Experimental studies have demonstrated that periodic temperature variations along 1.6 km of polarisation-maintaining fibre, with a maximum temperature difference of 37°C, can entirely suppress the SBS effect [36]. In this case, the SBS threshold increased from 4.4 mW to close to 10 mW when a heater is applied. Another study using a multi-level temperature variation with a maximum range of 140°C demonstrated a 4.8 dB increase in the SBS threshold over 100 m of HNLF [37].

Strain variation is another effective method for broadening the Brillouin gain bandwidth. By introducing a sinusoidal fibre strain variation of $\pm 0.35\%$ through fibre twisting during the fabrication process, the Brillouin spectrum width increased from 50 MHz to 400 MHz [38]. Further broadening of the Brillouin spectrum, up to 1.7 GHz, was achieved by a programmable fibre coiling machine capable of applying arbitrary strain distributions with a maximum strain differ-

ence exceeding 3% [39]. This significant enhancement, as demonstrated by experimental data, resulted in a theoretical increase of 15.5 dB in the SBS power threshold.

The Brillouin frequency shift can also be manipulated by varying the core radius along the fibre length, which changes the acoustic velocity v_A in Eq. (2.42) [40]. This method resulted in an SBS threshold increase of 3.6 dB compared to conventional fibres, as reported in the study.

Fibre dopants can modify the optical and acoustic index profiles of the fibre, leading to a shift in the Brillouin frequency [41]. Varying the fluorine (F) concentration along the fibre length, resulting in a relative-index difference of 0.2%, produced a 7 dB increase of the SBS power threshold compared to the conventional fibres at an operating wavelength of 1550 nm [42].

2.4.2 SBS suppression through pump-phase modulation

Another effective approach for increasing the SBS threshold is modulating the pump phase, also known as dithering, which broadens the pump's spectral width $\Delta\nu_p$ as described in Eq. (2.43). It works by generating multiple frequency sidebands around the central spectral line, effectively distributing the pump power across these components. The reduced intensity of each spectral mode results in the generation of weaker acoustic waves, which in turn leads to the suppression of SBS gain. Phase modulation is preferred over amplitude modulation because it preserves the total pump power while achieving the desired broadening, making it a more efficient technique for managing the SBS threshold.

The first experimental work of this method demonstrated more than 12 dB of SBS suppression at 1320 nm over 31.6 km of SMF [43]. Since then, this technique has been widely adopted due to its simplicity and lower cost compared to more complicated fibre design modifications. It is also versatile, as it can be applied to various types of fibres without requiring physical alterations. Several strategies have been explored for selecting the radio-frequency (RF) signal used to drive the phase modulator, including arbitrary and sinusoidal waveform.

Phase modulation using a random driving electrical signal or an electrical noise source is very efficient in broadening the pump linewidth although the resulted optical spectrum is hard to control. A RF white noise source, assisted by RF low pass filters for optical bandwidth control, was used to achieve an SBS enhancement factor of around 18 dB at an optical bandwidth of 1.47 GHz within a 70 m fibre [44]. In a later work, using a 3 GHz signal modulated by a 2^5-1 pseudo-random binary sequence (PRBS) pattern, a fibre amplifier with 1.7 kW output

power could be realised [45]. In contrast to the continuum spectrum generated by a white noise source, phase modulation using PRBS produces a discrete spectrum, offering greater flexibility in tuning the overlap with the Brillouin gain spectrum and, consequently, in controlling the SBS process seeding. However, despite the effectiveness of these techniques, they introduce additional random or pseudo-random distortion into the output signal, which complicates signal compensation at the receiver and may lead to challenges in maintaining overall signal quality.

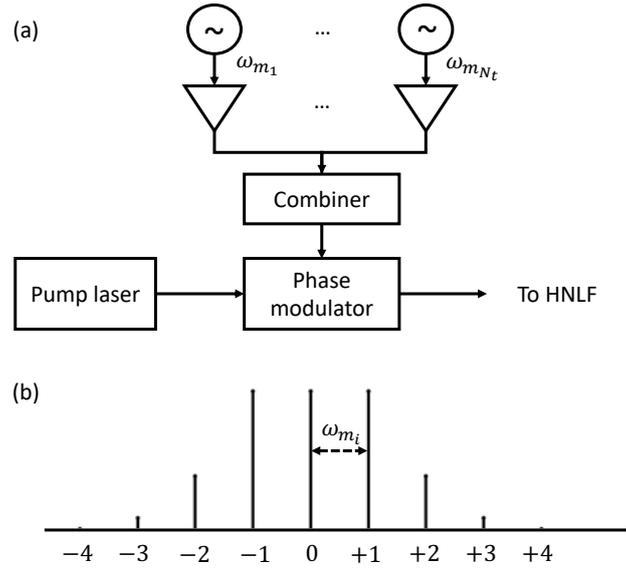


Figure 2.11: (a) Setup diagram of pump-phase modulation using multiple sinusoidal tones. (b) Pump spectrum after being phase modulated by a sinusoidal tone ω_{m_i} .

Sinusoidal waveform phase modulation, on the other hand, introduces deterministic distortion to the optical signal, which is easier to predict and manage. Additionally, sine wave generators are widely available and inexpensive, contributing to the popularity of this approach in practical applications. A diagram illustrating the pump-phase modulation scheme using a set of multiple sinusoidal tones, denoted as $[\omega_{m_1}, \dots, \omega_{m_{N_t}}]$ where N_t is the number of tones, is shown in Fig. 2.11(a). These tones are combined electronically and used to drive the phase modulator. The resulting phase-modulated pump signal is then sent to the HNLF for the FWM process.

The pump spectrum after phase modulation with an arbitrary tone at frequency ω_{m_i} , where $i \in [1, \dots, N_t]$, is demonstrated in Fig. 2.11(b). The phase-modulated pump signal contains sidebands around the pump frequency ω_p , spaced by the modulation frequency ω_{m_i} . The amplitude of each sideband can be described by the Bessel function of the first kind for a given order [46]. In Fig. 2.11(b), the horizontal axis represents the Bessel function order. We

observe that the higher-order peaks are much weaker than the zero-th and first-order peaks. The modulation index of the sinusoidal tone, set at 1.435 rad for this plot, results in nearly equal amplitudes between the zero-th and first-order peaks [47].

The modulation frequency ω_{m_i} should exceed the SBS gain bandwidth $\Delta\nu_B$, which typically ranges from 20 to 50 MHz for the HNLFF [46]. This ensures that the broadened pump spectrum extends beyond the SBS region, thus reducing the potential for SBS gain. Multiple RF tones are employed to further decrease the power of the generated frequency components in Fig. 2.11, thereby reducing the power spectral density (PSD) integrated over the SBS gain bandwidth. Several experimental studies have demonstrated the effectiveness of combining multiple RF tones for enhanced SBS suppression. For example, a FOPA with 70 dB gain was achieved by significantly surpassing the SBS power threshold, where the pump laser was modulated by a 10 GHz tone followed by four RF additional tones at 105, 325, 1000 and 3110 MHz [17]. The technique can be further enhanced by controlling the amplitudes and phase differences between the RF tones, with a suppression of 17 dB achieved with two modulation frequencies at 2.5 GHz and 3 GHz [48]. The motivation behind optimising the RF tone parameters is to produce an optimally broadened pump spectrum, which will be the main focus of section 2.5.

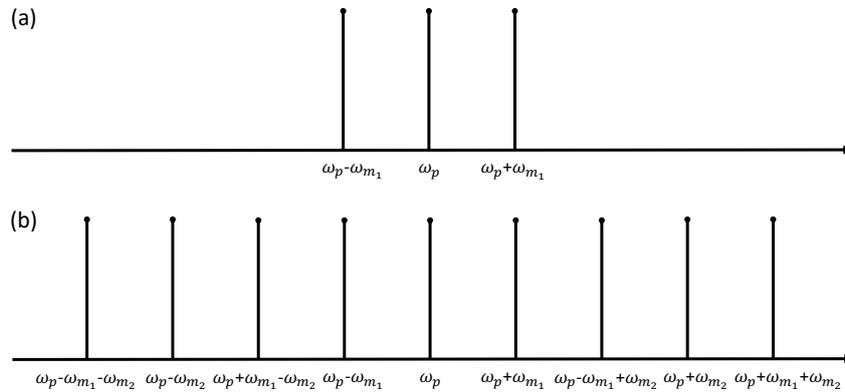


Figure 2.12: (a) Estimated pump spectrum after phase modulation with one tone at ω_{m_1} . (b) Estimated pump spectrum after phase modulation with two tones at ω_{m_1} and $\omega_{m_2} = 3\omega_{m_1}$.

The efficiency objective of the pump-phase modulation is to maximise the broadened pump bandwidth using the lowest possible modulation frequencies. Lower RF tone frequencies reduce the hardware demands on electronic equipment and minimise the impact on the FOPA transfer function. Since the power from the pump frequency is redistributed across the generated sidebands after modulation, achieving equal-amplitude spectral lines is important for optimal spectral broadening. Recent works have been done in tuning the amplitudes and phases

of RF tones to obtain close-to-equal power distribution of generated peaks [47], and using up to three modulation frequencies, a spectrum with up to 11 equal-amplitude spectral lines could be obtained [49]. The author in [50] suggested that selecting the modulation frequencies spaced by multiples of three, e.g. $[\omega_{m_1}, 3\omega_{m_1}, 9\omega_{m_1}, \dots]$, can result in more efficient broadening. The reason behind this can be found in Fig. 2.12, where the estimated broadened pump spectrum for a single modulation tone at ω_{m_1} is plotted in Fig. 2.12(a). Because the second-order sidebands are significantly lower than the zero-th and first-order components, we only consider peaks at ω_p and $\omega_p \pm \omega_{m_1}$. By introducing a second modulation frequency $\omega_{m_2} = 3\omega_{m_1}$, we can generate up to nine equal-amplitude frequency components at ω_p , $\omega_p \pm \omega_{m_1}$, $\omega_p \pm \omega_{m_2}$ and $\omega_p \pm \omega_{m_1} \pm \omega_{m_2}$, as shown in Fig. 2.12(b). These components are evenly spaced by the lower modulation frequency ω_{m_1} . The SBS power threshold can be further increased by introducing additional RF tones, each tone spaced in multiples of three from the lowest frequency, which is also known as the base frequency ω_{m_1} . The number of generated frequency components is generally 3^{N_t} , and the optical power is distributed evenly among these peaks, resulting in the increase of SBS power threshold by the same factor.

While we have discussed the frequency selection rule for the RF tones, the amplitudes and phases of these tones play crucial roles in determining the overall effectiveness of the pump-phase modulation technique. Achieving a broad and flat spectral profile, where the pump energy is evenly distributed across the generated sidebands, requires careful tuning of both the amplitudes and phases. Inaccurate choices can lead to uneven power distribution, reducing the efficiency of SBS suppression. In the next section, we will explore the procedure for optimising these parameters to achieve the desired pump spectral broadening.

2.5 Optimisation of pump-phase modulation with sinusoidals

We will present a numerical optimisation approach, where a desired rectangular spectrum, representing the theoretical upper limit of the SBS power threshold increase, is used as the learning target. The optimisation process will iteratively adjust the tone parameters, including amplitude and phases, to achieve the closest possible match to this target spectrum. Throughout the optimisation, we will monitor and calculate the corresponding increase in the SBS power threshold. We will then examine the resulted broadened spectrum in the last part.

We consider the case where the driving signal of the phase modulator, as depicted in Fig.

2.11(a), consists of multiple sinusoidal waves, defined as

$$\varphi(t) = \sum_{i=1}^{N_t} A_{m_i} \sin(\omega_{m_i} t + \xi_{m_i}), \quad (2.44)$$

where the parameters A_{m_i} , ω_{m_i} and ξ_{m_i} are the amplitude, frequency and phase of the i -th tone. The set of frequencies is $[\omega_{m_1}, 3\omega_{m_1}, 9\omega_{m_1}, \dots]$, according to the rule discussed in the section 2.4.2. In this optimisation, we selected a base frequency of $\omega_{m_1} = 60$ MHz to ensure it exceed the Brillouin gain bandwidth. Our task is to optimise A_{m_i} and ξ_{m_i} for an even broadened pump spectrum given the number of RF tones N_t .

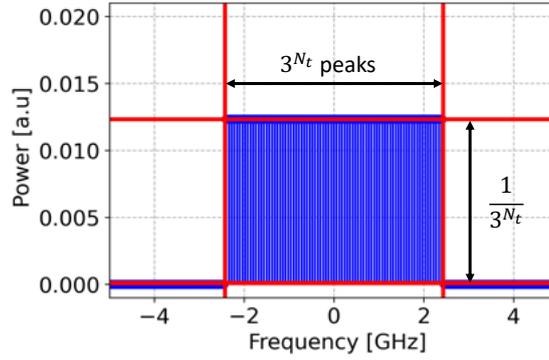


Figure 2.13: Ideal rectangular pump spectrum used as the reference.

The number of tones N_t is chosen based on the requirement to raise the SBS threshold sufficiently, and in the analysis throughout this thesis, we would like to achieve a power gain of 25 dB. In the 1550 nm wavelength region, the nonlinear phase shift $\Phi_{NL} = \gamma P_0 L$ typically ranges from 0.2 to 0.3 rad [34]. Given that the maximum achievable power gain is expressed as $G_{s,max} = |\cosh(\Phi_{NL})|^2$, this corresponds to a maximum gain of only 0.17 to 0.39 dB, which is significantly below the required 25 dB gain. Achieving 25 dB gain requires a nonlinear phase shift of approximately 3.57 rad. It is therefore necessary to increase the SBS threshold P_{th} by a factor of approximately 11.9 to 17.85, or equivalent a range of 10.76 to 12.52 dB. The increase in the SBS threshold is analytically calculated as [48]

$$\Delta P_{th} = -10 \log \left(\frac{\tilde{P}_{max}}{\tilde{P}_{0,max}} \right), \quad (2.45)$$

where $\tilde{P}_{0,max}$ and \tilde{P}_{max} are the maximum power densities over the SBS bandwidth before and after pump-phase modulation. Increasing ΔP_{th} is equivalent to reducing \tilde{P}_{max} to its minimum possible value, which theoretical occurs when all generated peaks from phase modulation have equal amplitudes. Given the frequency selection rule discussed earlier, the number

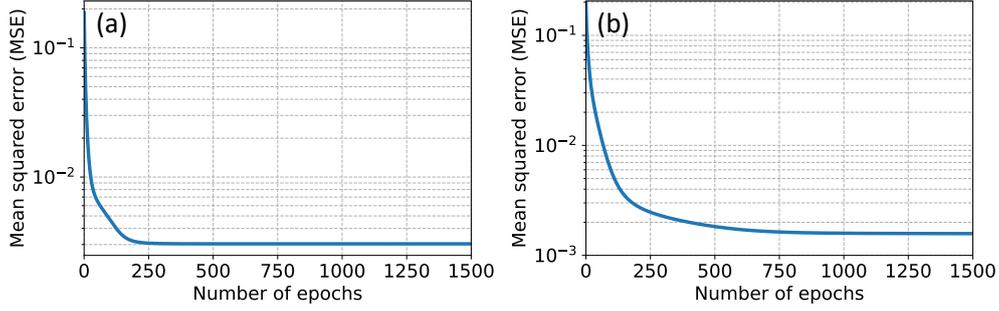


Figure 2.14: The convergence of the mean square error (MSE) cost function J in the modulation schemes with (a) 3 tones and (b) 4 tones.

of generated peaks is 3^{N_t} , so each generated peaks should have an amplitude of $1/3^{N_t}$ when $\tilde{P}_{0,\max}$ is normalised to one. With 3 and 4 RF tones corresponding to $N_t = 3$ and $N_t = 4$, the theoretical increase in SBS threshold would be 14.31 dB and 19.08 dB respectively, based on calculations from Eq. (2.45). One may notice that the use of 3 RF tones is sufficient to achieve our desired SBS increase factor. However, achieving a broadened pump spectrum with perfectly equal peaks and thereby reaching these theoretical values is impractical due to the beating effects of the RF tones, which we did not consider in the simplified analysis shown in Fig. 2.12. Moreover, the above calculation assumes a perfect conversion rate of the pump power for the FWM effect, which is not practical due to losses in the experimental setup. Therefore, the use of four tones might be a safe choice to achieve our desired performance for FOPA.

Although we cannot achieve an ideal rectangular broadened pump spectrum, we sought to optimise them as equally as possible by conducting an iterative learning process. In detail, we employed an adaptive approach to optimise A_{m_i} and ξ_{m_i} and construct the signal $\varphi(t)$ using Eq. (2.44) that results in a nearly rectangular broadened pump spectrum. We first define a reference spectrum with an ideal rectangular shape as depicted in Fig. 2.13. The number of peaks as well as the amplitude of each peaks are also shown. We then create an arbitrary learning signal in the form of $\varphi(t)$ with random initialisation of A_{m_i} and ξ_{m_i} . They are updated to minimise the cost function which is defined as

$$J = \sum_{i=1}^{3^{N_t}} |p_i - p_{i,\text{ref}}|^2, \quad (2.46)$$

where p_i and $p_{i,\text{ref}}$ represent the amplitudes of the i -th discrete peaks of the learning and reference spectra respectively. The mean square error (MSE) cost function was minimised in Tensorflow using the gradient descent method. The concept of gradient descent here is to

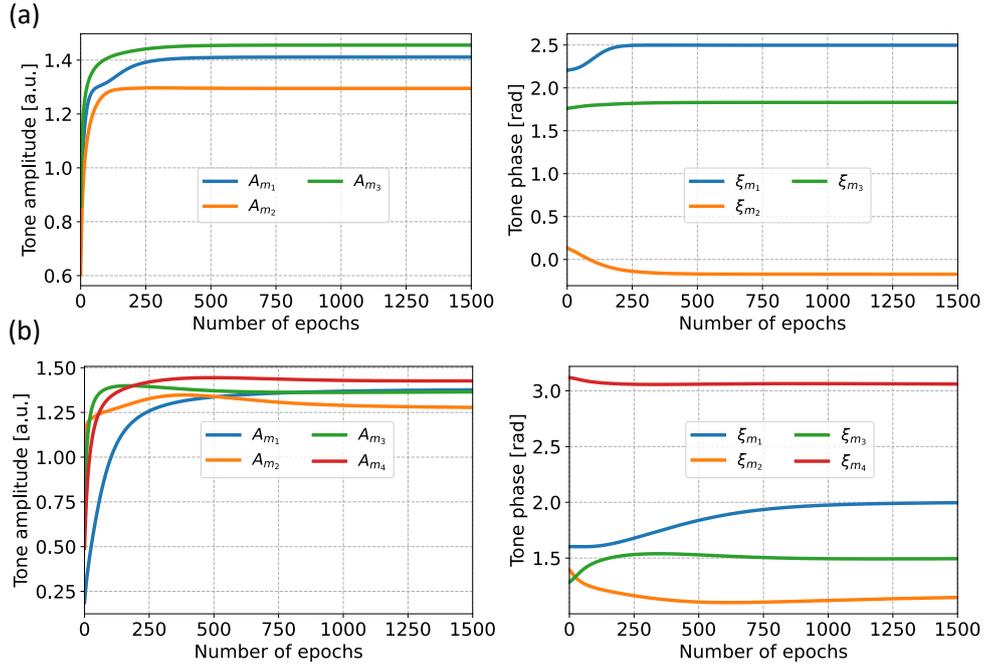


Figure 2.15: The evolution of the amplitudes and phases during the optimisation process in the modulation schemes with (a) 3 tones and (b) 4 tones.

iteratively update the variables A_{m_i} and ξ_{m_i} by moving them along the opposite direction of the gradient to reduce the cost function. The gradient is the vector of partial derivatives of the cost function J with respect to the variables. The update rule for the amplitudes and phases are

$$A_{m_i} := A_{m_i} - \mu \frac{\partial J}{\partial A_{m_i}}, \quad (2.47)$$

$$\xi_{m_i} := \xi_{m_i} - \mu \frac{\partial J}{\partial \xi_{m_i}}, \quad (2.48)$$

where $\partial J / \partial A_{m_i}$ and $\partial J / \partial \xi_{m_i}$ are the partial derivatives of the cost function J with respect to the amplitude A_{m_j} and phase ξ_{m_j} of the modulation tones respectively. The parameter μ , known as the learning rate, governs the convergence behaviour of the algorithm. It is selected as a compromise between convergence speed and algorithm stability. The convergence plots with a learning rate $\mu = 0.25$ are shown in Fig. 2.14. These plots demonstrate that the algorithm converges in both modulation schemes, with the MSE values stabilising before 1500 epochs.

The evolution of the amplitudes A_{m_i} and phases ξ_{m_i} for both pump-phase modulation schemes can be seen in Fig. 2.15. We observe that the convergence rates of these parameters closely corresponds to the MSE convergence trends shown in Fig. 2.14. The parameters in the modulation schemes with three tones reach their optimal values more quickly than those

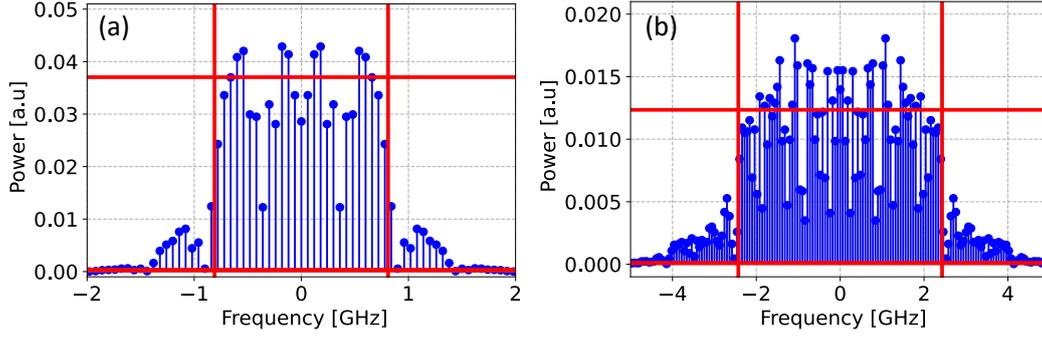


Figure 2.16: The broadened spectra after pump-phase modulation using the optimal parameters in the modulation schemes with (a) 3 tones and (b) 4 tones.

in the four-tone scheme, reflecting the faster MSE convergence in the three-tone case. These optimised parameter values, obtained after the algorithm convergence, will be applied for our numerical model of the pump-phase modulation scheme for further analysis.

The broadened pump spectra for both modulation schemes are plotted in Fig. 2.16, where the vertical and horizontal lines are aligned with the reference spectrum. These lines form the regions where the sideband peaks are intended to fit. In both schemes, the frequency components within the target region are significantly larger than those outside it. While achieving a perfectly rectangular spectrum is theoretically impossible, we observe a well-balanced distribution of generated peaks inside the target region, with suppressed peaks outside. The upper horizontal red lines indicate the theoretical minimum for the maximum power densities values \tilde{P}_{\max} in Eq. (2.45) can achieve. Using the optimised tone parameters, the resulting spectra effectively keep the maximum power densities close to these theoretical thresholds, indicating an efficient optimisation of the pump-phase modulation.

A key focus of this optimisation is determining the corresponding increase in the SBS power threshold based on the optimal RF tone parameters. Using Eq. (2.45) and the observed \tilde{P}_{\max} in Fig. 2.16, we can calculate the SBS power threshold ΔP_{th} . To track its relation with the convergence trend, these computations were performed after each epoch. Since the optimisation results depend on the random initialisation of the learning parameters A_{m_i} and ξ_{m_i} , we conducted the optimisation across 100 batches, with each batch corresponding to a different random initialisation. The results are plotted in Fig. 2.17, where the black lines represent the calculated ΔP_{th} for each batch and the green lines show the averaged curved across all batches. This graph shows both the individual batch variations and the overall trend of ΔP_{th} as the optimisation progresses.

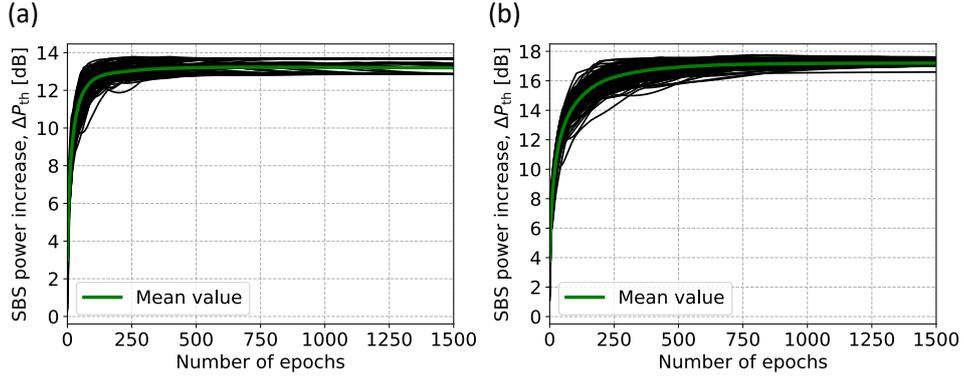


Figure 2.17: The SBS power threshold increase evolution during the optimisation process in the modulation schemes with (a) 3 tones and (b) 4 tones.

The observed values of ΔP_{th} at the final of the optimisation are approximately 13.2 dB and 17.2 dB for the three-tone and four-tone modulations schemes, respectively. These results are smaller than the theoretical values of 14.31 dB and 19.08 dB, as discussed earlier. This observation aligns with our previous discussion about the impracticality of achieving a perfectly rectangular broadened pump spectrum, which are associated with the theoretical limits. When comparing the required SBS power threshold increase of up to 12.52 dB for 25 dB power gain, the three-tone scheme appears to be a marginal choice when we decide the number of RF tones to use. We thus continue using both the three-tone and four-tone schemes in our numerical model throughout this thesis.

2.6 Conclusion

In this chapter, we have examined two primary fibre-optical parametric devices: FOPA and OPC. We began by exploring the principles of FWM and solving the corresponding propagation equations under the assumption of no pump depletion, which allowed us to derive the transfer functions for both devices.

Regarding FOPA, we emphasised the importance of fibre parameters and their influence on performance. A parameter optimisation procedure was outlined for the design stage to achieve the desired power gain spectrum. We also provide a detailed explanation of why the dual-pump FOPA scheme results in a flatter gain profile than the single-pump scheme, albeit at the cost of increased experimental complexity. This trade-off is a key consideration for practical implementation. For OPC, we highlighted its promising future in optical transmission, emphasising

its theoretical ability to mitigate both chromatic dispersion and fibre nonlinearity simultaneously.

In the latter of this chapter, the SBS effect and various suppression techniques were the main focuses. We discussed the principle of SBS, along with its power threshold, to determine key parameters that can be adjusted for suppression. A thorough review of suppression techniques provided insights into key milestones in the field. Among these, the pump-phase modulation approach stands out due to its effective suppression capability at a modest cost. We explored the efficient frequency selection method for multi-tone pump-phase modulation. Notably, we proposed and published an optimisation procedure aimed to optimise the SBS suppression efficiency further by learning the amplitudes and phases of modulation tones through an adaptive approach. Its effectiveness was demonstrated by analysing the resulted broadened pump spectrum and calculating the SBS power threshold increase.

Chapter 3

Digital signal processing

The development of coherent receivers, which mix the incoming signal with a local oscillator (LO), has transformed optical communications by enabling the extraction of both amplitude and phase information from the optical signal [51]. By providing access to full optical signal information, coherent receivers have presented numerous opportunities for researchers to develop novel DSP algorithms at the receiver (Rx), allowing for more effective mitigation of various channel impairments [52]. The combination of coherent receivers and advanced DSP techniques provides powerful solutions for modern backbone optical networks, making DSP research a focal point of interest.

In this chapter, we will explore the architecture of coherent receivers and their ability to capture both signal amplitude and phase. While a full DSP chain comprises multiple blocks, such as IQ imbalance compensation, polarisation mode dispersion (PMD) compensation or Kerr nonlinearity compensation, our focus here will be on the most relevant DSP blocks addressing fibre chromatic dispersion and laser phase noise mitigation. In each block, we will review the most widely used algorithms that will be used for further analysis in this thesis.

The latter part of this chapter will focus on the theory of kernel-based methods, which serve as the foundation for several algorithms proposed for FOPA systems in chapter 6. They are a class of algorithms designed to capture nonlinear relationships between data points without the need to explicitly compute in the higher-dimensional space. Kernel-based methods in DSP thus have attracted significant research interest due to their flexibility and effectiveness in complex signal processing tasks.

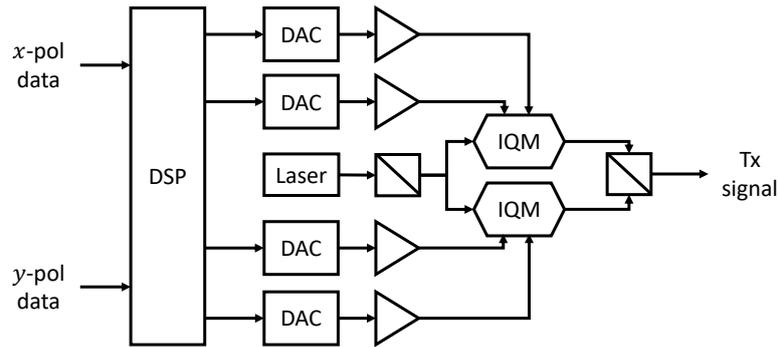


Figure 3.1: Configuration of coherent transmitter.

3.1 Coherent transceiver

During the third generation of optical fibre communications, coherent systems were extensively studied as a means to increase repeater spacing, due to their superior sensitivity compared to direct-detection systems. This sensitivity enables the receiver to operate closer to the quantum limit, whereas the performance of intensity modulation-direct detection (IM-DD) systems is strongly constrained by electronic and shot noise [53]. However, research on coherent systems was subsequently overshadowed by the advancements of WDM systems utilising IM-DD and the development of EDFA, which offers a cost-effective means of increasing repeater spacing [54]. Research into coherent systems regained momentum at the beginning of this century, driven by the increasing recognition of their benefits for data rate enhancement.

The configuration diagram of a coherent transmitter is shown in Fig. 3.1. At the transmitter (Tx), four digital signals produced by DSP, corresponding to the in-phase (I) and quadrature (Q) components of both the x - and y - signal polarisation, are converted to the analog domain by four digital-to-analog converter (DAC)s [55]. These signals are then pre-amplified before being fed into two IQ Mach-Zehnder modulator (MZM)s. In the case of Tx imperfection, imbalances in gain, phase and delay between the I and Q paths, as well as between different polarisation, may occur. The optical outputs are then combined by a polarisation beam combiner (PBC) before being transmitted through the optical fibre.

The Rx is designed to convert the incoming signal into four electrical signals, which represent four data sequences originally transmitted by the coherent transmitter [55]. The architecture of a phase and polarisation diversity coherent receiver is illustrated in Fig. 3.2. At the optical front-end, both the incoming optical signal and the LO are split into two polarisations by two polarisation beam splitter (PBS)s. These polarisation components are then mixed by

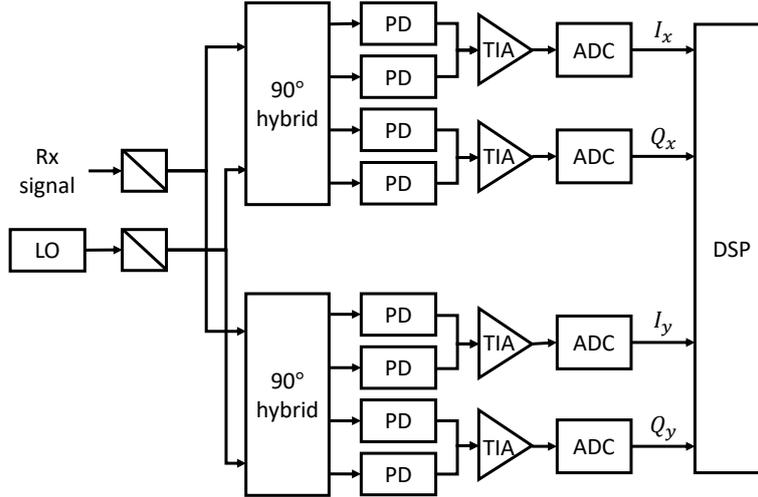


Figure 3.2: Configuration of coherent receiver.

a pair of 90° hybrids. Four balanced photodiode (PD)s detect the resulting electrical signals, which are then amplified by four transimpedance amplifier (TIA)s. The mixing due to the 90° hybrids produces beating terms - interference between the incoming optical signal and the LO - detected by the balanced PDs. Since LO serves as a reference signal, this beating allows both the amplitude and phase of the incoming signal to be extracted. Four analog-to-digital converter (ADC)s then converts the amplified signals into four digital signals I_x , Q_x , I_y and Q_y corresponding to the in-phase (I) and quadrature (Q) components for the x - and y - polarisation channels. Each signal is proportional to the beating between the corresponding polarisation component of the received signal and the LO. Specifically, I_x , Q_x , I_y and Q_y are proportional to $\Re\{E_x E_{LO}^*\}$, $\Im\{E_x E_{LO}^*\}$, $\Re\{E_y E_{LO}^*\}$ and $\Im\{E_y E_{LO}^*\}$, respectively [56], where $\Re\{\cdot\}$ and $\Im\{\cdot\}$ represent the real and imaginary parts of the signal.

The received signals are often distorted by various impairments and noise, making it the responsibility of the DSP to recover the original signal. To achieve this, the received signals are processed through a series of DSP blocks, each designed to mitigate specific distortions and enhance the signal quality.

3.2 Chromatic dispersion compensation

We focus here on the chromatic dispersion, as previously discussed in section 2.1.1. Chromatic dispersion causes the spreading of optical pulses as they propagate along the fibre, which results in inter-symbol interference (ISI) and limiting both transmission distance and bandwidth.

Therefore, CDC algorithms must be employed to preserve signal integrity in high-speed, long-distance data transmission. They ensure that the transmitted signal is accurately recovered at the receiver, enhancing the reliability of optical communication. In this section, we will first explain the frequency response of CDC, followed by its implementation in the time domain.

To compensate for chromatic dispersion, we begin by considering a linear system, where fibre nonlinearity is neglected. The propagation of an optical pulse $A(z, t)$ over a distance z and time t , originally described by Eq. (2.40), can now be expressed as

$$\frac{\partial A}{\partial z} + \frac{j\beta^{(2)}}{2} \frac{\partial^2 A}{\partial t^2} = 0. \quad (3.1)$$

By applying the Fourier transform $\tilde{A}(z, \omega)$ of $A(z, t)$, this equation becomes

$$\frac{\partial \tilde{A}(z, \omega)}{\partial z} - \frac{j\beta^{(2)}}{2} \omega^2 \tilde{A}(z, \omega) = 0. \quad (3.2)$$

Solving this equation gives us the channel response

$$G(z, \omega) = \frac{\tilde{A}(z, \omega)}{\tilde{A}(0, \omega)} = \exp\left(j \frac{\beta^{(2)} \omega^2 z}{2}\right). \quad (3.3)$$

Since chromatic dispersion is a static impairment, it can be compensated by applying an all-pass filter with an inverse channel response, i.e. $H(z, \omega) = 1/G(z, \omega)$. The transfer function for dispersion compensation over a transmission length L , denoted as $H(\omega)$, can be expressed in terms of the dispersion coefficient D , defined in Eq. (2.2), as

$$H(\omega) = \exp\left(j \frac{\lambda^2 D \omega^2}{4\pi c} L\right). \quad (3.4)$$

The Fourier transform of the dispersion-compensated signal is obtained by multiplying the Fourier transform of the distorted signal by the transfer function $H(\omega)$ defined in Eq. (3.4). The all-pass compensating filter with response $H(\omega)$ can be approximated in the time domain using a finite impulse response (FIR) filter. Although implementing the filter in the frequency domain is more efficient [57, 58, 59], we opt for the time-domain implementation due to its simplicity and straightforward approach.

In the time domain, the output of a FIR filter y_n at any time step n is calculated as

$$y_n = \sum_{i=-M}^M w_i x_{n-i}, \quad (3.5)$$

where x_n is the distorted signal due to chromatic dispersion, and x_{n-i} represents the neighbouring samples affected by the memory effect caused by dispersion. The FIR filter is char-

acterised by a set of tap coefficients w_i , with the total number of coefficients being $2M + 1$, where M is the filter delay. Since the received signal in a coherent system is complex-valued, the computation with tap coefficients also takes place in the complex domain.

The filter coefficients can be dynamically updated based on error evaluation, such as with a least mean squares (LMS) adaptive filter [60]. Alternatively, the coefficients can be pre-calculated based on the compensating filter response $H(\omega)$ in frequency domain, as defined in Eq. 3.4), given the knowledge of the channel. The i -th filter coefficient has its analytical form given by [52]

$$w_i = \sqrt{\frac{jcT^2}{\lambda^2 DL}} \exp\left(-j\frac{\pi cT^2 i^2}{\lambda^2 DL}\right), \quad (3.6)$$

where T is the sampling period. The filter delay can also be analytically determined as

$$M = \left\lfloor \frac{\lambda^2 |D| L}{2cT^2} \right\rfloor, \quad (3.7)$$

where the notation $\lfloor \cdot \rfloor$ denotes the largest integer number which does not exceed the value inside the bracket. According to Eq. (3.6), the tap weights exhibit constant amplitudes, with the real and imaginary components change periodically. It is also worth noting that the number of tap coefficients derived from Eq. (3.7) represents only the upper limit. The actual required filter length may be much smaller, and research has been conducted on filters with fewer taps and reduced implementation cost [61]. Because CDC is not the main focus of this thesis, we only use it as a fundamental block within the DSP chain used in our transmission scenarios. We thus either use the time-domain method described by Eqs. (3.6) and (3.7) or directly apply the compensating filter response $H(\omega)$ from Eq. (3.4) to achieve the goal of dispersion compensation.

3.3 Carrier phase recovery

CPR is a fundamental block within DSP that estimates and corrects the phase of the optical carrier. In coherent communication systems, the detection of phase-modulated symbols, especially when using advanced modulation formats such as phase shift keying (PSK) and quadrature amplitude modulation (QAM), is severely affected by phase noise. This phase noise arises from the non-zero linewidths of the laser sources in both Tx and Rx, as well as in optical parametric devices. CPR plays an important role to mitigate this issue, enabling the use of

advanced modulation formats for improved bandwidth utilisation. There have been numerous algorithms proposed for CPR, such as Viterbi-Viterbi algorithm [62] or method with maximising a posteriori (MAP) estimation [63], and they can broadly be classified into two classes: feedback and feed-forward approaches. In this section, we will review two candidate algorithms from each approach: LMS-based phase recovery [64, 65] and blind phase search (BPS) [66], highlighting their operational principles and contributing parameters.

3.3.1 LMS-based phase recovery

As a feedback CPR, the LMS-based phase recovery works as an one-tap adaptive filter whose tap weight is update iteratively based on symbol error evaluation [64]. In detail, at symbol step n , the algorithm computes a complex tap coefficient w_n to derotate the distorted symbol r_n , producing the equalised symbol $z_n = r_n w_n$ which is then passed to a decision circuit. An alternative method produces an estimated phase $\hat{\phi}_n$ instead [65], and the derotated symbol is calculated as $z_n = r_n e^{-j\hat{\phi}_n}$, as illustrated in Fig. 3.3. The decision circuit calculates the squared distances between the derotated symbol z_n and the constellation points, and selects the closest point as the decision-directed symbol $[z_n]_D$.

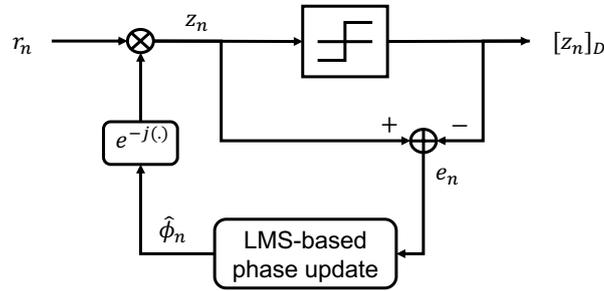


Figure 3.3: LMS-based phase recovery scheme.

The tap coefficient w_n or estimated phase $\hat{\phi}_n$ is then updated based on the squared error between the equalised symbol and its corresponding decision-directed one, i.e. $|e_n|^2 = |z_n - [z_n]_D|^2$. The tap weight w_n is updated using the normalised LMS algorithm [67] as

$$w_{n+1} = w_n - \mu \frac{e_n r_n^*}{|r_n|^2}, \quad (3.8)$$

where μ is the step size or learning rate and the superscript $*$ denotes the complex conjugate operator. The estimated phase $\hat{\phi}_n$ has its updating rule written as

$$\hat{\phi}_{n+1} = \hat{\phi}_n - \mu \Im\{z_n e_n^*\}, \quad (3.9)$$

The parameter μ is selected to balance the trade-off between the tracking capability and noise tolerance, and is thus sensitive to the signal modulation formats [68]. A high μ means a faster reaction to the phase variation but makes the algorithm vulnerable to the stochastic noise and leads to incorrect decision-directed symbols and phase estimates, potentially causing cycle slips. A small μ , in contrast, reduces noise sensitivity but reacts slowly to high phase variation, and thus makes the algorithm vulnerable to large laser linewidths. In general, the LMS-based phase recovery is computationally simple given a proper tuning of the step size μ , but its feedback nature makes it challenging for hardware realisation.

3.3.2 Blind phase search

The BPS algorithm, as a widely-used feed-forward CPR method, estimates the optimal phase by testing a set of candidate phase values, rather than relying on a feedback error signal for iterative updating [66]. The optimal phase is selected based on the criterion that minimises the squared distance between the derotated symbol and the closest constellation point. A diagram demonstrating the operation of the BPS algorithm is shown in Fig. 3.4. The BPS algorithm divides a phase quadrant, ranging from 0 and $\pi/2$, into B test phases. Each test phase φ_b is given as

$$\varphi_b = \frac{\pi b}{2B}, \quad (3.10)$$

where b is an integer number in the range $[0, B - 1]$. The number of test phases B is chosen as a trade-off between the algorithm resolution and computational complexity. In our transmission scenarios with 16-QAM signal within this thesis, we set $B = 32$.

For each received symbol r_n , the algorithm rotates it by each of test phases to generate the rotated symbols $z_{n,b} = r_n e^{j\varphi_b}$. There are B possible rotated symbols $z_{n,b}$ corresponding to B test phases. The rotated symbols are passed through decision circuits which generate the corresponding detected symbols $[z_{n,b}]_D$. The algorithm measures the squared distances between the rotated symbols and their corresponding detected symbols, i.e.

$$|d_{n,b}|^2 = |z_{n,b} - [z_{n,b}]_D|^2. \quad (3.11)$$

To improve noise tolerance, the algorithm considers a window of $2K + 1$ symbols surrounding the current symbol. The algorithm selects the estimated phase $\hat{\varphi}_n$ from possible values of φ_b that minimises the sum of squared distances over this window, i.e.

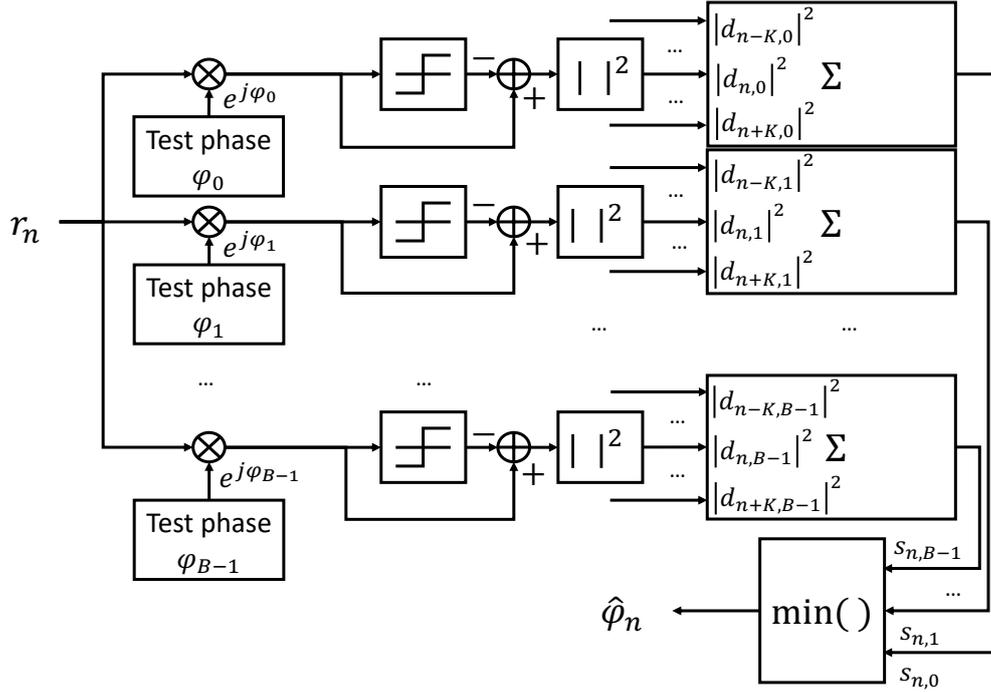


Figure 3.4: BPS phase recovery scheme.

$$\min_b \sum_{k=-K}^K |d_{n-k,b}|^2. \quad (3.12)$$

The window length $2K + 1$ is a key parameter governing the algorithm's performance, with the behaviour being similar to the learning rate μ in the feedback CPR algorithm. A longer window increases noise tolerance but limits the algorithm's ability to track high-speed changes.

Due to the symmetry of the constellation in QAM signals, the BPS algorithm exhibits a four-fold ambiguity problem. It refers to the fact that phase rotations by multiples of $\pi/2$ return equivalent constellation points. To resolve this issue, we calculate the phase difference between two consecutive estimated phases to correct phase jumps by multiples of $\pi/2$. Given the selected phase $\hat{\varphi}_n$ at symbol step n derived from Eq. (3.12), the number of cycle jump C_n is determined as

$$C_n = \left\lfloor \frac{1}{2} - \frac{\hat{\varphi}_n - \hat{\varphi}_{n-1}}{\pi/2} \right\rfloor, \quad (3.13)$$

where $\lfloor \cdot \rfloor$ operates as the floor function defined in Eq. (3.7). The second term inside the floor function represents how many $\pi/2$ steps between two consecutive estimated phases. The floor operation with threshold $1/2$ returns an integer value which represents how many multiples of $\pi/2$ the current estimated phase has shifted. The corrected estimated phase is thus

$$\hat{\varphi}_{n,\text{corrected}} = \hat{\varphi}_n + C_n \frac{\pi}{2}. \quad (3.14)$$

This corrected value is used to produce the equalised symbol instead of the initial estimated phase $\hat{\varphi}_n$ for better phase recovery.

We have discussed two typical CPR algorithms: while LMS-based algorithm can provide a simple tool in numerical simulations, BPS algorithm offers a more hardware-friendly solution, which is more practical.

3.4 Digital signal processing with kernel methods

The concept of kernel methods is mapping input signals from a finite-dimensional space into a higher-dimensional reproducing kernel Hilbert space (RKHS) using *kernel functions*. Hilbert space is a space which allows us to deal with signal of infinite dimension [69]. The kernel functions, through *kernel trick*, allow us to compute the dot products in the RKHS without explicitly defining or calculating feature vectors in that space. We can instead directly compute the kernel functions with input vectors from the original space.

On the other hand, many real-world problems are about nonlinear relationships which linear models cannot sufficiently capture. Only nonlinear models can learn the complex input-output relationships. However, nonlinear models are often computationally expensive and challenging to interpret. Feature mapping provides an elegant solution, allowing the complex nonlinear relationship to be linearly represented in the higher-dimensional space. This can be achieved implicitly through the use of kernel functions as discussed above. Therefore, kernel methods provide a powerful framework to model nonlinear functions while relying on linear algebra. This capability makes kernel methods highly valuable in machine learning for regressions and classification tasks. They are also attractive tools in the statistics and signal processing areas [70].

3.4.1 Feature mapping

An illustration of feature mapping is shown in Fig. 3.5, produced from [70], where the data is mapped from an input space, denoted as \mathcal{X} , into a higher-dimensional Hilbert space, denoted as \mathcal{H} , by a nonlinear mapping $\phi(\cdot)$. The feature mapping $\phi(\cdot)$ can be written as

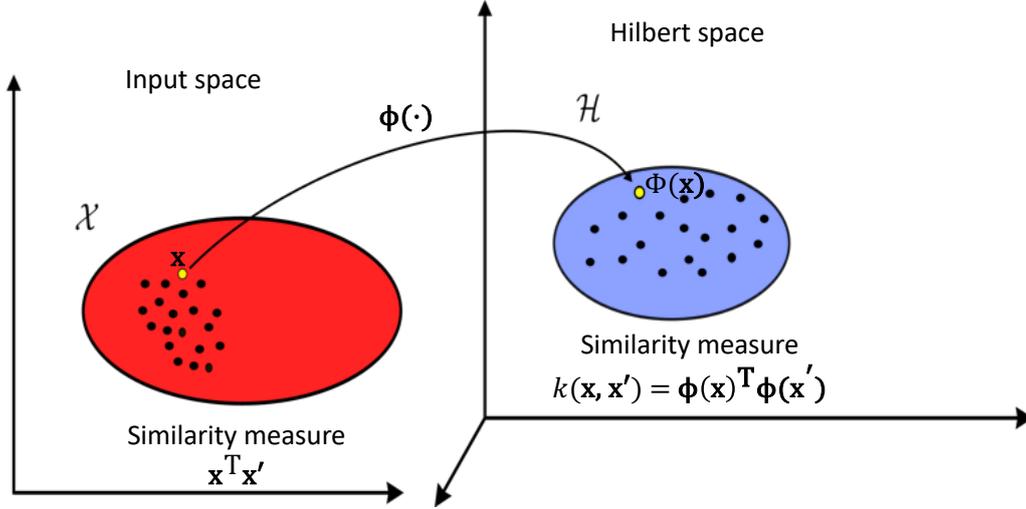


Figure 3.5: Demonstration of feature mapping to $\phi(\cdot)$ to Hilbert space.

$$\begin{aligned}\phi : \mathcal{X} &\longrightarrow \mathcal{H} \\ \mathbf{x} &\longmapsto \phi(\mathbf{x}).\end{aligned}$$

In the higher-dimensional space \mathcal{H} , we can apply a linear model to learn an arbitrary function $f(\mathbf{x})$, which is nonlinear in the original space \mathcal{X} , i.e.

$$f(\mathbf{x}) = \langle \phi(\mathbf{x}), \hat{\mathbf{w}} \rangle_{\mathcal{H}}, \quad (3.15)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ represents the dot or inner product operation and $\hat{\mathbf{w}}$ is the weight vector in the Hilbert space¹. The problem of the space \mathcal{H} is that the transformed data can be very high-dimensional, making the computation of the dot product here computationally challenging. Kernel methods introduce the kernel trick, which enables the dot product computation between transformed vectors without explicitly performing the mapping into higher-dimensional space. It instead uses a positive definite function (or a kernel function) [70] to compute the dot product directly in terms of the original input. This allows kernel methods to efficiently compute the similarities between data points, which are represented by dot products, as if they were mapped into the higher-dimensional space. The kernel function is defined as

$$k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}}, \quad (3.16)$$

where \mathbf{x} and \mathbf{x}' are two arbitrary data vectors. The feature mapping also conceptually

¹The weight vector $\hat{\mathbf{w}}$ has the size equal to the dimension of the Hilbert space.

implies that any data point \mathbf{x}' can be defined by its similarity $k(\mathbf{x}, \mathbf{x}')$ to the data point \mathbf{x} .

Although kernel methods are used for classification and regression tasks, we mainly focus on its application in regression within this thesis. In the next subsections, we will explore the learning algorithms with kernel methods for nonlinear regression task.

3.4.2 Kernel ridge regression

Kernel ridge regression extends the concept of ridge regression to model nonlinear relationships between input features and output predictions through the use of kernel functions. We define a general problem for nonlinear regression: Given an input $\mathbf{x}_n \in \mathbb{R}^L$ and an observation output $y_n \in \mathbb{R}$, we aim to model the underlying function $y_n = f(\mathbf{x}_n) + \epsilon_n$, where ϵ_n represents the noise, often assumed to be normally distributed with zero mean and variance σ^2 , i.e. $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$. We consider a training set of N observations and define $\mathbf{X} \in \mathbb{R}^{N \times L}$ is the input observation matrix, which stacks its rows with N input patterns. The corresponding output vector is denoted as $\mathbf{y} \in \mathbb{R}^N$.

We first examine linear ridge regression before extending it to nonlinear regression. In the linear regression case, we seek the weight vector $\mathbf{w} \in \mathbb{R}^L$ that minimises the least-squares cost function

$$\min_{\mathbf{w}} J(\mathbf{w}) = \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 + \lambda\|\mathbf{w}\|^2, \quad (3.17)$$

where $\|\cdot\|$ is the L_2 -norm of the vector inside the brackets and λ is the regularisation term. This λ term is chosen to balance the trade-off between data fitting and the amplitudes of the weight vector. The solution of this equation is given by

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1} \mathbf{X}^T\mathbf{y}, \quad (3.18)$$

where \mathbf{I} is the identity matrix. We now consider the case of nonlinear regression, where the relationship between \mathbf{x}_n and y_n cannot be captured by a linear model. We can map the data into a higher-dimensional feature space (possibly infinite-dimensional), where the problem becomes linear again. In the feature space \mathcal{H} , the least-squares cost function becomes

$$\min_{\hat{\mathbf{w}} \in \mathcal{H}} \hat{J}(\hat{\mathbf{w}}) = \sum_{i=1}^N \|y_i - \langle \phi(\mathbf{x}_i), \hat{\mathbf{w}} \rangle_{\mathcal{H}}\|^2 + \lambda\|\hat{\mathbf{w}}\|_{\mathcal{H}}^2, \quad (3.19)$$

where $\|\hat{\mathbf{w}}\|_{\mathcal{H}}^2 = \langle \hat{\mathbf{w}}, \hat{\mathbf{w}} \rangle_{\mathcal{H}}$ is the squared norm of $\hat{\mathbf{w}}$ in the Hilbert space. We now introduce the *Representer Theorem* [71], which shows that solutions to certain cost minimisation prob-

lems, involving quadratic regularisers like the one above, can be represented as expansions in terms of the training data. Therefore, the nonlinear function $f(\mathbf{x})$ can be written as a linear combination of the kernel functions computed with the training data points [72]

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}). \quad (3.20)$$

If we define the kernel matrix \mathbf{K} with the entry \mathbf{K}_{ij} at i -th row and j -th column as

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j), \quad (3.21)$$

and by the definition of kernel function in Eq. (3.16), we can also write

$$\mathbf{K} = \phi(\mathbf{X})\phi(\mathbf{X})^T. \quad (3.22)$$

From Eq. (3.22) and Eq. (3.20), we can now represent the nonlinear mapping f for the entire training data as

$$f(\mathbf{X}) = \mathbf{K}\boldsymbol{\alpha}, \quad (3.23)$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T \in \mathbb{R}^N$ is the coefficient vector. Because the function f can be considered as linear in the higher-dimensional space, we have $f(\mathbf{x}) = \langle \phi(\mathbf{x}), \hat{\mathbf{w}} \rangle_{\mathcal{H}}$. From Eq. (3.22), the weight vector $\hat{\mathbf{w}}$ can be written in the basis defined by rows of all mapped data points, i.e.

$$\hat{\mathbf{w}} = \sum_{i=1}^N \alpha_i \phi(\mathbf{x}_i). \quad (3.24)$$

The cost function in Eq. (3.19) can now be rewritten as

$$\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}) = \|\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}\|^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha}. \quad (3.25)$$

Its solution is given by

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}. \quad (3.26)$$

We observe that the original problem of finding the optimal weight vector $\hat{\mathbf{w}} \in \mathcal{H}$, which lies in a potentially high-dimensional space, can be reformulated into a more computationally efficient problem of finding the coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^N$. This transformation leverages the kernel trick, allowing us to work with kernel evaluations instead of explicitly computing in a high-dimensional feature space. In the next subsection, we will explore an approach that also leverages kernels, but within a probabilistic framework. This method not only provides predictions but also quantifies uncertainty, a key feature for tuning kernel hyperparameters.

3.4.3 Gaussian process

A Gaussian process (GP) defines a distribution over functions, where the function values at any finite set of input points are jointly distributed according to a multivariate Gaussian distribution. GPs are powerful because they capture not just the predicted values but also the uncertainty associated with those predictions. This makes them useful for tasks that require confidence intervals or uncertainty quantification.

In a GP, the function $f(\mathbf{x})$ is assumed to be a random variable drawn from a Gaussian process \mathcal{GP} , which is fully specified by its mean function $m(\cdot)$ and covariance function $k(\cdot, \cdot)$ i.e.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (3.27)$$

The mean function $m(\mathbf{x})$ is defined as the expected function at input \mathbf{x} , i.e.

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]. \quad (3.28)$$

The covariance function $k(\mathbf{x}, \mathbf{x}')$, which is also known as the kernel function, represents the correlation between function values at different input points \mathbf{x} and \mathbf{x}' , i.e.

$$k(\mathbf{x}, \mathbf{x}') = \text{Cov}(f(\mathbf{x}), f(\mathbf{x}')). \quad (3.29)$$

Assuming the mean function is zero for simplicity, the function values $f(\mathbf{x})$ at a finite set of training inputs $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}^T$ follow a multivariate normal distribution, i.e.

$$\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_N)]^T \sim \mathcal{N}(0, \mathbf{K}), \quad (3.30)$$

where \mathbf{K} is the kernel matrix defined in Eq. (3.21). The observation output vector \mathbf{y} , which includes noise $\epsilon_n \sim \mathcal{N}(0, \sigma^2)$, can be written as

$$\mathbf{y} \sim \mathcal{N}(0, \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}), \quad (3.31)$$

where $\mathbf{K}(\mathbf{X}, \mathbf{X})$ is the kernel matrix computed for the training set. To predict the output for a test input set \mathbf{X}_* , GP considers the joint distribution of the observation output \mathbf{y} and the test outputs \mathbf{y}_* , which is given by [73]

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right). \quad (3.32)$$

The conditional distribution of \mathbf{y}_* given \mathbf{y} provides the predictive mean and covariance

$$\bar{\mathbf{y}}_* = \mathbf{K}(\mathbf{X}_*, \mathbf{X}) [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}]^{-1} \mathbf{y}, \quad (3.33)$$

$$\text{Cov}(\mathbf{f}_*) = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) [\mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}]^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}_*). \quad (3.34)$$

This provides both the predictive mean and the covariance, allowing us to quantify the uncertainty of the predictions.

3.4.4 Kernel functions

In this section, we will discuss few commonly used examples of kernel function and which assumptions on the underlying functions are corresponded to each of them. A valid kernel function should be symmetric and positive semidefinite [73].

We first discuss about *Gaussian kernel*, which is also known with other names, such as radial basis function (RBF) kernel or the squared exponential kernel. Gaussian kernel is widely recognised for its ability to handle nonlinear relationships and its flexibility, often delivering strong performances across various datasets while relying on only a single hyperparameter. However, its drawbacks include high computational and memory costs as the dataset size increases, which are also general disadvantages of kernel-based methods, and its sensitivity to hyperparameter tuning, which can lead to overfitting if not managed properly. Despite these challenges, the Gaussian kernel is a universal approximator, meaning it can model any dataset if appropriately tuned. This makes it a reliable and effective starting point when applying kernel methods to nonlinear problems. The name "Gaussian" comes from their expression, which is similar to the probability density function of the Gaussian distribution. Its form is

$$k_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l_{\text{RBF}}^2}\right), \quad (3.35)$$

where l_{RBF} represents the kernel lengthscale, which governs the kernel's response to the distance between points. A small value of l_{RBF} means the kernel function decays quickly with the distance $\|\mathbf{x} - \mathbf{x}'\|$, so the points \mathbf{x} and \mathbf{x}' needs to be very close to be regarded as having high similarity. In contrast, a high value of l_{RBF} allows the highly separated points to be considered as similar.

A demonstration of a GP with Gaussian kernel is shown in Fig. 3.6, where the colour map showing the calculated kernel function values given the input points \mathbf{x} and \mathbf{x}' is plotted in Fig. 3.6(a). We also plotted three functions $f(\mathbf{x})$ sampled from GP prior in Fig. 3.6 to show the properties of the underlying functions which Gaussian kernel makes assumption in. The Gaussian kernel is modelled for functions with local variation structure, with the lengthscale

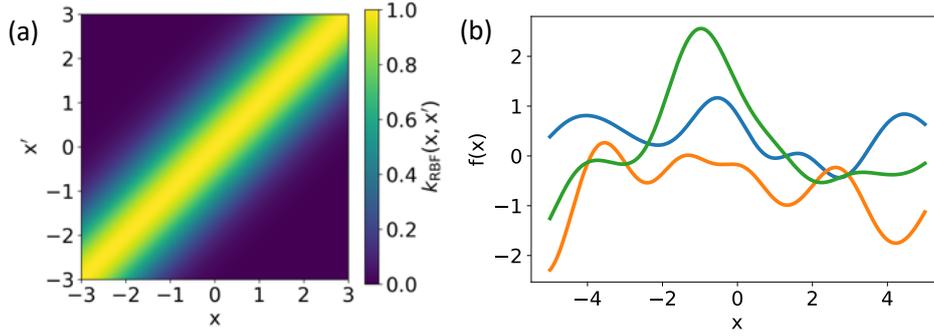


Figure 3.6: Gaussian kernel with its (a) covariance matrix and (b) example functions sampled from GP prior.

l_{RBF} defines their smoothness. Functions with rapid variations, where high similarity is only considered for close points, are modelled with a short lengthscale l_{RBF} , while a long lengthscale is selected for smoother functions.

The Gaussian kernel is most widely used choice for kernel selection due to its flexibility and simplicity. Although it is governed by a single hyperparameter - l_{RBF} , it is a universal approximator given enough data and right hyperparameter choice.

The second kernel function we consider is the *periodic kernel*. It was derived in [74], with the objective of modelling functions with repeated structures. It is also referred to by exponential sine squared due to its mathematical expression, which is

$$k_{\text{Per}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{2 \sin^2(\pi \|\mathbf{x} - \mathbf{x}'\|/p)}{l_{\text{Per}}^2}\right), \quad (3.36)$$

where l_{Per} is the lengthscale with its implicit meaning is similar to l_{RBF} within Gaussian kernel. The hyperparameter p is called periodicity, which defines the period of the oscillation within modelling functions. It ensures that the calculated function values are still dependent on the point distances but only within repeated cycles.

The covariance matrix associated with the periodic kernel is plotted in Fig. 3.7(a), where the covariance values exhibit a repeating pattern as the distance between points varies, reflecting the periodic nature of the kernel. Three sample functions drawn from a GP with the periodic kernel are shown in Fig. 3.7(b), with each clearly demonstrating the repeated structure. This makes the periodic kernel is ideal for modelling cyclic behaviour, and particularly useful for estimating the dithering-induced distortion in FOPA links, where signal distortion exhibits oscillatory characteristics.

While there are many other types of kernel functions, such as linear kernel, rational quadratic

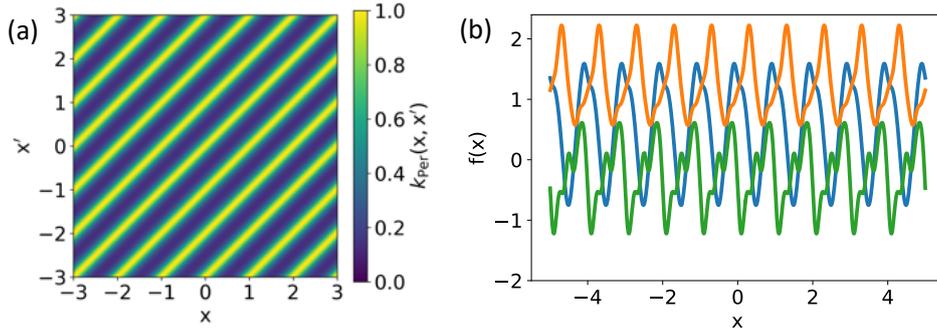


Figure 3.7: Periodic kernel with its (a) covariance matrix and (b) example functions sampled from GP prior.

kernel and constant kernel [73], this thesis only focuses on those most relevant to the specific problems involving periodicity and smoothness. We have seen that each kernel function has its own hyperparameter, such as l_{RBF} in Eq. (3.35) or l_{Per} in Eq. (3.36). These hyperparameters determine how the model generalises and selecting appropriate values for them is crucial for achieving optimal performance. Hyperparameter selection can be performed through grid search or cross-validation, but these approaches can be computationally expensive, especially if the number of hyperparameters increases. To address this, we will take a probabilistic approach by introducing the concept of GP.

3.4.5 Kernel hyperparameter optimisation

Each kernel function comes with its own hyperparameters and in GP regression, selecting a right set of hyperparameters is as crucial as choosing the kernel function itself. A common approach for hyperparameter optimisation involves maximising the marginal likelihood function $p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector of GP hyperparameters. It quantifies how likely the observed output data \mathbf{y} is given the input data \mathbf{X} and the model hyperparameters $\boldsymbol{\theta}$.

We start with the probability density function of \mathbf{y} , which follows a multivariate normal distribution as described in Eq. (3.31), is given by

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2} |\mathbf{K}_y|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y}\right), \quad (3.37)$$

where $|\cdot|$ defines the determinant operator and $\mathbf{K}_y = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I}$ is defined as the covariance matrix considering the noisy output data. It is usually more convenient to work with the log marginal likelihood instead, which is expressed as [73]

$$\log p(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}_y| - \frac{N}{2} \log(2\pi). \quad (3.38)$$

The first term, $\mathbf{y}^T \mathbf{K}_y^{-1} \mathbf{y}$, evaluates how well the model fits the observed data. The second term, $\log |\mathbf{K}_y|$, acts as a regulariser by penalising the model complexity. The last term is a normalised constant term and does not depend on the model parameters. Therefore, the optimisation process aims to maximise the log marginal likelihood with respect to hyperparameter $\boldsymbol{\theta}$. This results in a trade-off between accurately fitting the data and reducing the model complexity, allowing the model to generalise well to unseen data.

3.5 Conclusion

We have discussed in this chapter how the optical signal, thanks to the coherent receivers, is detected preserving both phase and amplitude information. We also reviewed existing algorithms for fundamental blocks, CDC and CPR blocks, within the DSP chain which will be applied to our scenario in the following chapters. In the second part of this chapter, we provided the theoretical fundamentals about kernel-based methods and discussed how it can be used for nonlinear regression tasks. We also briefly introduced the role of kernel functions within GP regression before introducing two primary families of kernel functions. A hyperparameter optimisation procedure by maximising log marginal likelihood was also discussed.

Chapter 4

Amplitude and phase distortion compensation in systems with optical phase conjugator

In this chapter, we examine the optical transmission systems with a mid-link OPC, focusing on the phase distortion induced by the residual dithering introduced at the OPC stage. The conventional DSP algorithms, typically employed to mitigate phase noise arising from non-zero laser spectral widths, are ineffective in addressing this particular form of phase distortion. To overcome this limitation, we propose a novel compensation scheme that simultaneously addresses the phase distortion in conjunction with the secondary amplitude distortion caused by fibre dispersion in the latter half of the transmission link. Notably, the ability to compensate for the amplitude distortion is an innovative aspect of this work which was published in the CLEO and ICTON paper [1, 2], as only phase compensation has been previously explored in the literature.

4.1 Mid-link optical phase conjugation transmission

We consider an optical transmission of total N spans, each with a SMF segment followed by an optical amplifier (OA), using a mid-link OPC configuration, as shown in Fig. 4.1. The OPC is deployed at the midpoint, dividing the transmission into two segments, each containing $N/2$ spans. As discussed in section 2.3, by creating a phase-conjugated copy of the signal, the OPC

can effectively reverse, at the end of the second half, any distortions accumulated in the first segment.

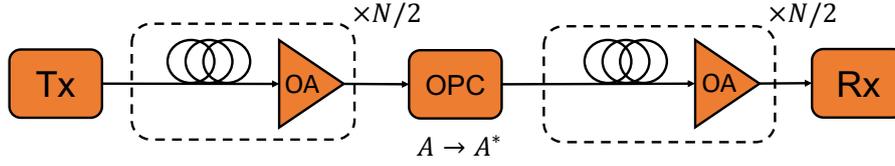


Figure 4.1: Scheme of OPC transmission.

We modelled the OPC based on the dual-pump polarisation-insensitive configuration described in [75, 76], as illustrated in Fig. 4.2. In this setup, two pump lasers are phase modulated by phase modulator (PM)s driven by electric RF tones generated from an arbitrary waveform generator (AWG). Two low-pass filter (LPF)s are used to remove any higher-order frequency components generated by the nonlinear response of the preceding RF amplifiers. The two phase-modulated pumps are then combined by a PBC before being sent to a HNLF for signal mixing.

One of key insights from Chapter 2 is that the pump-phase modulation is essential for SBS suppression, which is then important to achieve sufficient OPC conversion efficiency. The conversion efficiency is defined as $G_i = |\mu_i|^2$, where μ_i represents the complex idler gain derived from Eq. (2.39). The maximum conversion efficiency is thus given as

$$G_{i,\max} = \sinh^2(\Phi_{\text{NL}}), \quad (4.1)$$

where the nonlinear phase shift Φ_{NL} is defined in Eq. (2.27). For HNLFs in the 1550 nm region, typical nonlinear phase shifts are only around 0.2-0.3 radians [34], resulting in the maximum conversion efficiency $G_{i,\max}$ of approximately -13.9 to -10.3 dB. Additional calculated values of maximum conversion efficiency, corresponding to different SBS threshold increase factors, are provided in Table 4.1.

SBS threshold increase factor	1	3	9
Number of required RF tones	0	1	2
Max nonlinear phase shift [rad]	[0.2, 0.3]	[0.6, 0.9]	[1.8, 2.7]
Max conversion efficiency [dB]	[-13.9, -10.3]	[-7.72, 0.23]	[9.4, 17.4]

Table 4.1: SBS-limited conversion efficiency for different levels of SBS suppression.

From this table, we observe that achieving 'lossless' optical phase conjugation requires increasing the nonlinear phase shift by at least a factor of three. Considering additional external factors which may reduce the conversion efficiency, it is beneficial using a pump-phase modu-

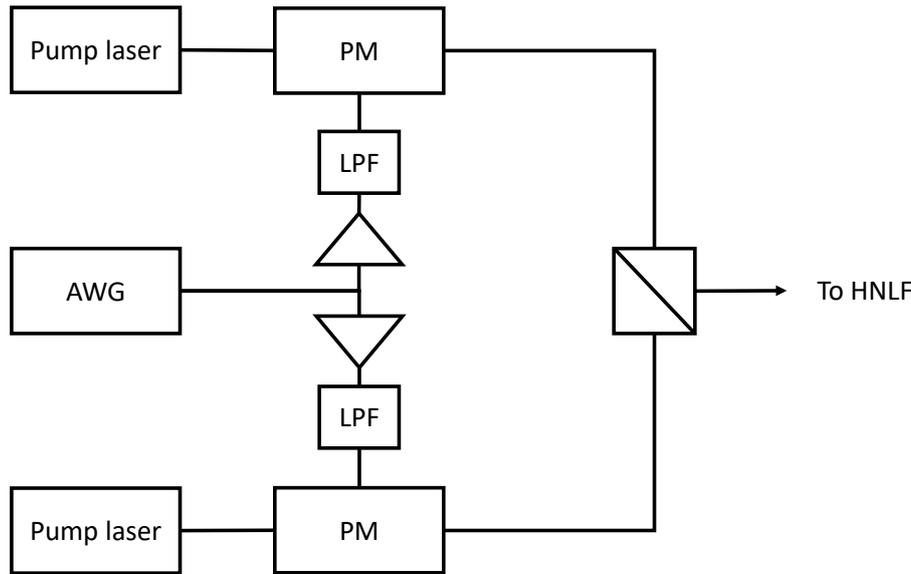


Figure 4.2: Dual-pump configuration for OPC implementation.

lation with two RF tones. In [75, 76], two sinusoidal RF tones at $f_1 = 60$ MHz and $f_2 = 600$ MHz are used to for the electrical signal driving the phase modulator.

4.2 Residual dithering-induced distortion

In the dual-pump configuration depicted in Fig. 4.2, the ideal scenario occurs when the two pumps are precisely driven by two out-of-phase signals which theoretically cancel out any phase distortions which could be transferred to the signal. However, in practice, imperfections in adjustments of these driving signals, due to mismatches in electronic components or any deviations in the amplifier characteristics, often result in residual dithering that is transferred to the output signal. In detail, when the AWG generates two sets of RF tones in Fig. 4.2, achieving perfect out-of-phase operation is challenging, resulting in a nonzero phase mismatch. This mismatch prevents complete cancellation of phase distortions, leading to residual phase fluctuations in the output signal. In practice, phase mismatches typically range from a few degrees to tens of degrees, depending on the synchronisation accuracy of the AWG. This residual dithering can degrade overall system performance, particularly in coherently detected transmission system. In this section, we will first model the phase distortion due to residual dithering, followed by a discussion in the phase-to-phase and phase-to-amplitude distortion conversion caused by chromatic dispersion in the second fibre segment.

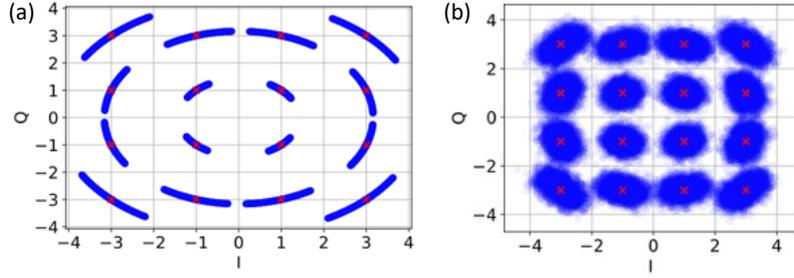


Figure 4.3: Constellation diagrams of presumed OPC transmission without ASE and laser phase noise in (a) back-to-back configuration (b) 10-span configuration.

4.2.1 Residual dithering-induced phase distortion

The pump-phase modulation scheme was numerically modelled according to [77], where the driving signals, which were generated by a two-channel AWG, are composed by two sinusoidal RF tones at f_1 and f_2 , corresponding to angular frequencies ω_1 and ω_2 respectively. The electrical signals are expressed as

$$\varphi_1(t) = (A + \delta A) [\cos(\omega_1 t + \delta\theta) + \cos(\omega_2 t + \delta\theta)], \quad (4.2)$$

$$\varphi_2(t) = -A [\cos(\omega_1 t) + \cos(\omega_2 t)], \quad (4.3)$$

where A is the modulation index in rad, δA and $\delta\theta$ denote the modulation index mismatch and phase mismatch respectively, representing the imperfect RF tone adjustments. In our analysis, we set the modulation index $A = 1.2$ rad and the modulation index mismatch $\delta A = 0.01$ rad while varying the phase mismatch $\delta\theta$. The complex envelope of the two phase-modulated pump waves are modelled as

$$A_{pi}(t) = |A_{pi}| e^{j[\delta\phi_i(t) + \phi_{mi}(t)]}, \quad (4.4)$$

where $i = 1, 2$ represents the pump index. The phase distortion $\delta\phi_i(t)$ is caused by the non-zero spectral width of the laser and is numerically modelled using the Brownian motion model. In this model, $\delta\phi_i(t)$ is composed of increments between consecutive time steps that follow a Gaussian distribution [78]. The second term, $\phi_{mi}(t)$, accounts for the modulated pump phase by RF tones. The driving signals described in Eqs. (4.2) and (4.3) are amplified by non-ideal RF amplifiers before being passed through LPFs as depicted in Fig. 4.2. To simulate the nonlinearity of these RF amplifiers, we employ a third-order polynomial model [79]. Therefore,

the pump-phase modulation by each pump $\phi_{mi}(t)$ is expressed as [77]

$$\phi_{mi}(t) = \int d\tau h(t - \tau) [\alpha_1 \varphi_i(\tau) + \alpha_2 \varphi_i(\tau)^2 + \alpha_3 \varphi_i(\tau)^3], \quad (4.5)$$

where α_1 , α_2 and α_3 are the coefficients of the RF amplifier's response, $h(t)$ is the low-pass filter's impulse response. In our simulation, the RF amplifier response's coefficients are set as $\alpha_1 = 1$, $\alpha_2 = 0.007$ and $\alpha_3 = 0.005$ to align with the experimental data in [76]. The envelope of the idler signal is proportional to $A_{p1} \cdot A_{p2}$, so the phase modulation transferred to the idler signal is

$$\Psi(t) = \sum_{i=1}^2 [\delta\phi_i(t) + \phi_{mi}(t)]. \quad (4.6)$$

We can see that under the assumption of linear response of the RF amplifiers, the dithering-induced pump phase modulation transferred to the OPC output is zero when we can achieve an ideal counter-phasing configuration for $\varphi_1(t)$ and $\varphi_2(t)$. In other words, it can be achieved when $\delta A = \delta\theta = 0$. Otherwise, the phase modulation of the idler is modulated by the frequencies f_1 , f_2 and the values of the modulation mismatch δA and phase mismatch $\delta\theta$. A constellation diagram obtained after a numerical simulation of an OPC back-to-back transmission is shown in Fig. 4.3(a). If we neglect the laser phase noise and amplified spontaneous emission (ASE) noise, the final constellation diagram only demonstrates the phase distortion arising from the effect discussed above.

4.2.2 Phase-to-phase and phase-to-amplitude distortion conversion

In the mid-link OPC transmission which is assumed to be free of laser phase noise and ASE noise, we are left with chromatic dispersion, fibre nonlinearity and the residual dithering-induced phase distortion which was discussed in the previous section. Suppose that the OPC can effectively correct the distortion from dispersion and nonlinearity, one may think we are left with only dithering-induced phase distortion. However, the received constellation diagram shows the presence of amplitude distortion, as demonstrated in Fig. 4.3(b) with a transmission of 10 fibre spans. This effect occurs when phase modulation in the idler signal at the OPC output interacts with the chromatic dispersion present in the fibre during the second half of the transmission.

The relationship between phase and amplitude modulation at the input and output of a dispersive fibre can be given by a small-signal analysis. The input signal with optical power $S_{in}(t) = \langle S \rangle + \Delta S_{in}(t)$, where $\langle S \rangle$ is the average and $\Delta S_{in}(t)$ is the small-signal intensity

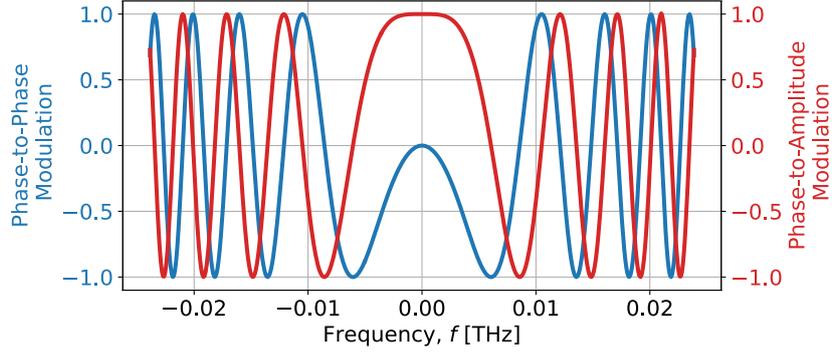


Figure 4.4: Phase-to-phase and phase-to-amplitude modulation effects in dispersive channel.

modulation, and phase $\phi_{in}(t)$ is related to the output signal with optical power $S_{out}(t) = \langle S \rangle + \Delta S_{out}(t)$ and phase $\phi_{out}(t)$ by [80]

$$\begin{bmatrix} \frac{\Delta S_{out}(j\omega)}{2\langle S \rangle} \\ \phi_{out}(j\omega) \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{\beta_2 z}{2}\omega^2\right) & \sin\left(\frac{\beta_2 z}{2}\omega^2\right) \\ \sin\left(-\frac{\beta_2 z}{2}\omega^2\right) & \cos\left(\frac{\beta_2 z}{2}\omega^2\right) \end{bmatrix} \begin{bmatrix} \frac{\Delta S_{in}(j\omega)}{2\langle S \rangle} \\ \phi_{in}(j\omega) \end{bmatrix}. \quad (4.7)$$

In our OPC model, only phase modulation is present, so we only consider the terms with $\phi_{in}(j\omega)$. We plot the phase-to-phase and phase-to-amplitude modulation based on Eq. (4.7) in Fig. 4.4. Phase-to-phase modulation reflects how the signal phase changes along the dispersive channel, while phase-to-amplitude distortion describes how phase changes cause variations in the signal amplitude. This explains why the symbols with phase distortion in Fig. 4.3(a) are converted into clouds with both phase and amplitude distortions in Fig. 4.3(b). This happens differently for different frequency components as shown in Fig. 4.4, leading to inter-symbol interference in which overlapping symbols cause degradation to the neighbouring symbols. It is an effect with memory, so we need to account for sufficient channel memory to reverse this effect. Recognising and understanding this phenomenon is important to design effective compensation schemes at the receiver, preserving signal quality over long-haul transmission.

4.3 Phase and amplitude distortion compensation scheme

In the previous section, we examined how residual pump-phase modulation is transferred to the OPC output and how this phase modulation leads to different types of distortion at the receiver. Both phase and amplitude distortions negatively affect systems using advanced modulation formats such as QAM, which relies on the accurate detection of both signal phase and amplitude. In this section, we propose a dithering-induced distortion compensation (DDC) scheme to ef-

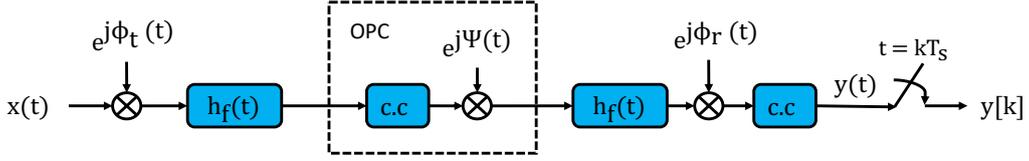


Figure 4.5: Baseband equivalent model of transmission with single-stage OPC.

fectively mitigate these distortions by estimating the residual dithering-induced phase distortion and cancelling out the phase-to-phase and phase-to-amplitude distortion conversion.

4.3.1 Operating principle

We first analyse, in time domain, the impact of residual dithering-induced phase distortion and how it contributes to the distortion conversion observed at the received signal. This can be achieved by examining a linear baseband equivalent model, shown in Fig. 4.5, of the optical transmission with a mid-link OPC. In this model, each half of the transmission link, consisting of $N/2$ fibre spans, is represented by a FIR filter, whose impulse response is denoted as $h_f(t)$. Examining the signal propagation, the transmitted signal, $x(t)$, is first affected by the transmitter laser phase noise $\phi_t(t)$ before passing through the filter represented by $h_f(t)$. After the OPC stage, the signal propagates to the second half of the link, also modelled by the same filter $h_f(t)$. At the receiver, the signal is further influenced by the LO laser phase noise $\phi_r(t)$. The signal is then undergoes complex conjugation (denoted by the block labelled c.c), followed by matched filtering and downsampling. The received symbol at time instant k is denoted as $y[k]$. The operation of OPC is modelled as a complex conjugation block followed by the presence of phase distortion $\Psi(t)$ in Eq. (4.6). The contributions of ASE noise is excluded from this analysis for simplicity.

Because the total laser phase noise, defined as $\delta\phi(t) = \phi_t(t) + \phi_r(t) + \delta\phi_1(t) + \delta\phi_2(t)$, varies on a much slower timescale than the residual dithering-induced phase distortion $\phi_{mi}(t)$ (kHz compared to GHz level), our analysis focuses on the interaction between fibre dispersion, represented by $h_f(t)$, and the faster phase distortions $\phi_{mi}(t)$. The total laser phase noise can thus be separated from this process and assumed to be compensated by a conventional CPR algorithm. Therefore, the end-to-end relationship between the input signal $x(t)$ and output signal $y(t)$ can be expressed as

$$y(t) = e^{j\delta\phi(t)} \left[\left[[x(t) * h_f(t)]^* e^{j\phi_m(t)} \right] * h_f(t) \right]^*, \quad (4.8)$$

where $*$ denotes the convolution and the superscript $*$ represents the complex conjugate operator. The total dithering-induced phase distortion is $\phi_m(t) = \phi_{m1}(t) + \phi_{m2}(t)$. Applying the linear approximation $e^{j\alpha} \simeq 1 + j\alpha$, we can simplify the above equation to

$$y'(t) = x(t) - [[x(t) * h_f(t)] j\phi_m(t)] * h_f^*(t), \quad (4.9)$$

where $y'(t) = y(t)e^{-j\delta\phi(t)}$ represents the laser phase noise-compensated signal. The participation of the residual dithering-induced phase distortion $\phi_m(t)$ leads to the deviation between the transmitted signal $x(t)$ and received signal $y'(t)$. An estimate of the second term is thus crucial to correct the received signal. From the above equation, we can solve for the residual dithering-induced phase distortion $\phi_m(t)$. Given that $\phi_m(t)$ is real-valued, it can be expressed as

$$\phi_m(t) = -\Im \left\{ \frac{[y'(t) - x(t)] * h_f(t)}{x(t) * h_f(t)} \right\} \quad (4.10)$$

At the receiver, the signal $y'(t)$, obtained after the CPR stage, is known and the channel impulse response $h_f(t)$ can be estimated based on prior knowledge of the link characteristics. However, the input signal $x(t)$ remains unknown and needs to be estimated. We propose approximating $x(t)$ using the decision-directed symbols from the output of the CPR block. This leads to a two-stage compensation scheme, which will be discussed in the next section.

4.3.2 Compensation scheme

Based on the preceding analysis, the proposed compensation scheme to mitigate the dithering-induced distortion operates in two stages, as depicted in Fig. 4.6. The first stage applies a CPR algorithm, such as the BPS method [66], to produce the symbols free from laser phase noise. The second stage focuses on estimating the second term in Eq. (4.9). In this stage, an estimate $\bar{x}[k]$ of the transmitted symbol $x[k]$ is generated by a decision circuit, which then participates in Eq. (4.9) in the role of transmitted signal $x(t)$ to recreate the residual dithering-induced complex distortion. This block also requires knowledge of the channel impulse response $h_f[k]$, which can be estimated using the method outlined in [52], given the number of spans and the fibre's dispersion coefficient. The final estimate of the transmitted symbol $\hat{x}[k]$ is then produced by adding the estimated dithering-induced complex distortion to the output of the CPR block.

However, the effectiveness of this stage significantly depends on how accurately the residual dithering-induced phase distortion $\hat{\phi}_m[k]$ is estimated. In the next section, we will explore a

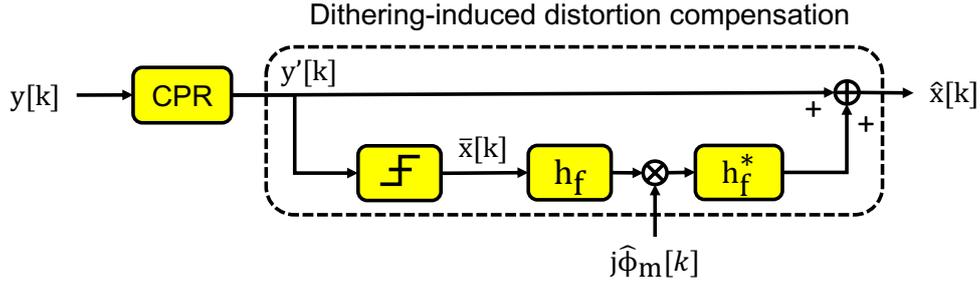


Figure 4.6: The compensation scheme for OPC transmission.

linear regression approach to learn this phase distortion.

4.3.3 Estimation of dithering-induced phase distortion

The total pump-phase modulation transferred to the idler $\phi_m[k]$ is the key term the algorithm needs to predict. Based on Eqs. (4.2), (4.3) and (4.5), this term follows a sinusoidal pattern corresponding to the frequencies used in the pump-phase modulation. This behaviour can be confirmed by analysing the PSD of the first estimated residual dithering-induced phase distortion after the CPR stage. This estimate is calculated as the angle of the complex difference between the symbols after CPR, $y'[k]$, and the corresponding detected symbols, $[y'[k]]_D$, i.e. $\angle \{y'[k] - [y'[k]]_D\}$. As illustrated in Fig. 4.7, the PSD plot clearly reveals two prominent frequency peaks at frequencies $f_1 = 60$ MHz and $f_2 = 600$ MHz. The prediction of $\phi_m[k]$ thus can be done by performing linear regression with the feature vector using the knowledge of the above frequencies. In detail, the feature vector $\mathbf{B}[k] \in \mathbb{R}^{2N_t}$ is formed as

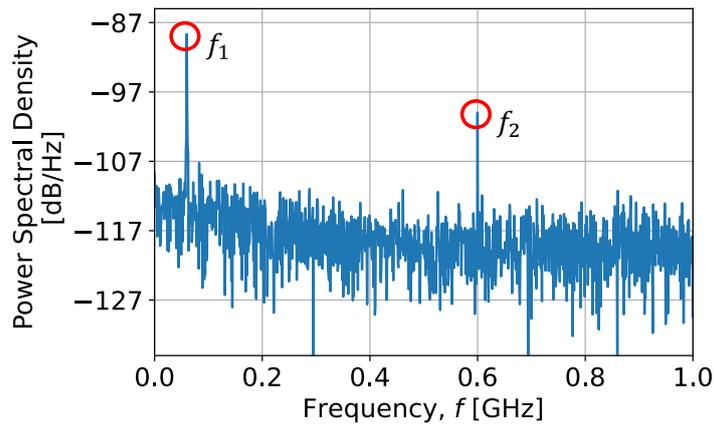


Figure 4.7: Power spectral density of the residual dithering-induced phase distortion.

$$\mathbf{B}[k] = [\sin[\omega_i k], \cos[\omega_i k], \dots]^T, i = 1, 2, \dots, N_t. \quad (4.11)$$

In our scenario, there were two modulation frequencies corresponding to $N_t = 2$, but the general case can consider up to N_t number of RF tones. The estimated residual dithering-induced phase modulation $\hat{\phi}_m[k]$ is obtained as $\hat{\phi}_m[k] = \boldsymbol{\theta}^T[k]\mathbf{B}[k]$ through a coefficient vector $\boldsymbol{\theta} \in \mathbb{R}^{2N_t}$. The linear regression model was trained using batch gradient descent (GD) algorithm with a training sequence of 1000 symbols and a learning rate of 0.1. The algorithm convergence is shown in Fig. 4.8(a) with the convergence is guaranteed after 100 epochs. After the algorithm reaches its convergence, it uses decision-directed symbols instead of the reference ones. To plot these graphs and for further analysis, we used the modulation index $A = 1.2$ rad, the modulation index mismatch $\delta A = 0.01$ rad for Eqs. (4.2) and (4.3). The phase mismatch $\delta\theta = 6^\circ$ for Fig. 4.8 and will be varied in one of the next testing scenarios.

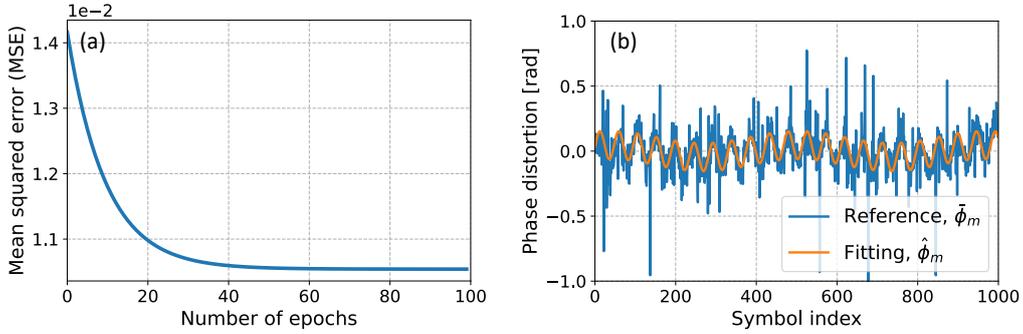


Figure 4.8: Regression learning of residual dithering-induced phase distortion.

4.4 Compensation result

The direct-count bit-error-rate (BER) is used as the evaluation metric to compare the performance between different compensation schemes across testing scenarios. The parameters used in our numerical simulation are listed in Table. 4.2 with the fibre parameters are corresponding to the standard SMF. The propagation along the SMF was modelled using the split-step Fourier method (SSFM) with maximum step width of 1 km and maximum phase change of 0.005 rad. For our analysis, we modelled a transmission of a 16-QAM signal using a root-raised-cosine (RRC) filter with a roll-off factor of 0.1 at 28 Gbaud, with 8 samples per symbol. The BER values in each scenario was averaged over 10 batches of 2^{16} symbols.

We start our analysis by finding the optimal launched power for our mid-link OPC trans-

Parameters	Values
Modulation	16-QAM
Baud rate	28 [GBaud]
Span length	100 [km]
Loss	0.2 [dB/km]
Dispersion	17 [ps/(nm.km)]
Nonlinearity	1.2 [$W^{-1}km^{-1}$]
Tx/Rx laser linewidths	100 [kHz]
OPC laser linewidths	50 [kHz]
Noise figure	4.5 [dB]

Table 4.2: OPC transmission numerical simulation parameters

mission. In this scenario, we considered the number of spans $N = 20$, which is equivalent to the transmission distance of 2000 km. It will be the maximum number of spans considered in this analysis. To isolate the impact of the residual-dithering induced phase distortion, we assumed a nearly perfect OPC model with the phase mismatch $\delta\theta = 0^\circ$. We used the LMS-based algorithm for the CPR block because of its reduced computational cost, and tried with different learning rate values. We plotted the BER curve versus the various launched power and observed the best curve obtained at the learning rate of 0.005, and this curve is plotted in Fig. 4.9. As observed from the graph, the optimal BER is achieved at a launched power of 1 dBm, which will be used for the next testing scenarios. Although mid-link OPC transmission theoretically compensates for fibre dispersion and nonlinearity, the increase in BER within the nonlinear regime is primarily caused by parametric noise amplification. This amplification arises from independent ASE sources introduced by each amplifier along the transmission. Only the ASE from the first amplifier of the link benefits from the same nonlinear compensation as the signal after OPC. For any subsequent amplifiers, this ideal compensation cannot be fully achieved due to asymmetry with respect to the receiver and the mid-link OPC. As a result, additional parametric noise amplification further degrades the signal-to-noise ratio, limiting the effectiveness of nonlinear compensation [81].

Regarding compensation schemes to be compared, besides the BPS method, which will be used for the CPR stage, we also included a phase-only compensation scheme outlined in [77]. In this algorithm, the author performed a grid search on phase (GSP) to find the best suitable fitting for the residual dithering-induced phase modulation ϕ_m . Two grids of 10×10 were applied for the amplitude and phase values. While the searching range of phase is $[0, 2\pi]$ rad, the range for the amplitude is $[0, 0.2]$ rad. The algorithm returns the optimal values which minimise the squared distance between the corrected symbol and the nearest constellation

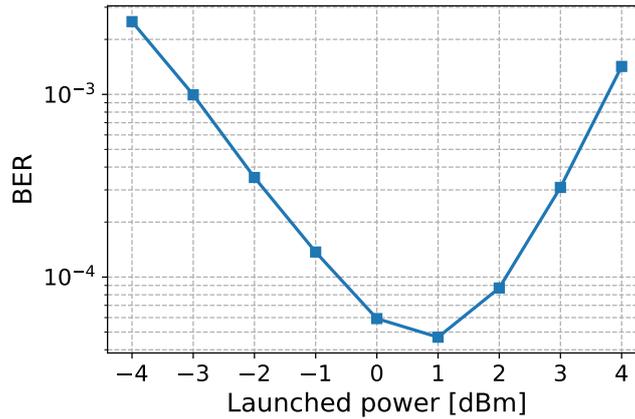


Figure 4.9: Launched power optimisation in mid-link OPC transmission.

point. This method is supported by a stage of BPS, so we will refer it as BPS-GSP in the later discussion. Our proposed algorithm will be referred as BPS-DDC.

4.4.1 Performance over transmission distance

Under the optimum launched power of 1 dBm, the performance comparison in terms of BER versus transmission distance between testing compensation schemes is shown in Fig. 4.10. In this analysis, the phase mismatch is set as $\delta\theta = 6^\circ$. We note that the number of spans N is always an even number because of the symmetry of the mid-link OPC transmission, with its value runs within the range between 4 and 20 in this simulation.

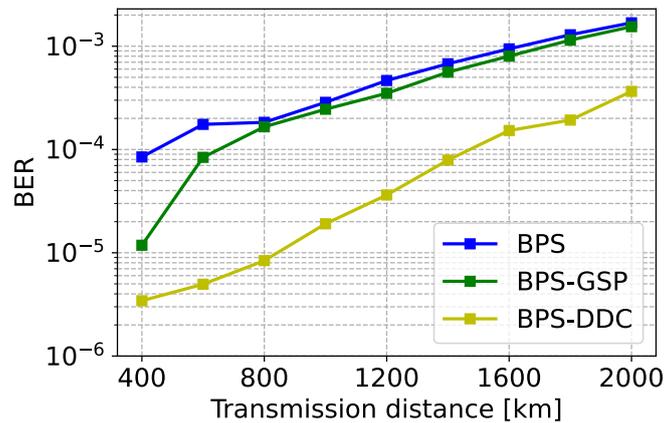


Figure 4.10: Performance of OPC transmission system with compensation schemes.

We observe that the BPS-GSP method achieves a significant BER improvement at short transmission distances, but the enhancement is quickly reduced as the number of spans in-

creases. It can be explained by the second term in Eq. (4.9) becomes more pronounced as more channel memory due to higher transmission distance. Compensation in only phase domain is thus insufficient for signal recovery task, and even the interaction with fibre dispersion causes an additional complex distortion, which makes the fitting on phase time series even more tricky and more prone to errors. Conversely, by including the fibre dispersion into the calculation, our proposed BPS-DDC can take into account of the interaction of residual dithering-induced phase distortion with it, and maintain a consistent BER improvement across considered transmission distance range. At the maximum transmission distance, we achieved a bit less than one order of BER improvement over other phase-only compensation schemes.

4.4.2 Impact of phase mismatch

It is worth testing the impact of phase mismatch $\delta\theta$, which obviously contributes to the overall system performance. Figure 4.11 makes comparison between different compensation schemes under the variation of the pump phase mismatch $\delta\theta$ after transmissions of 12 spans.

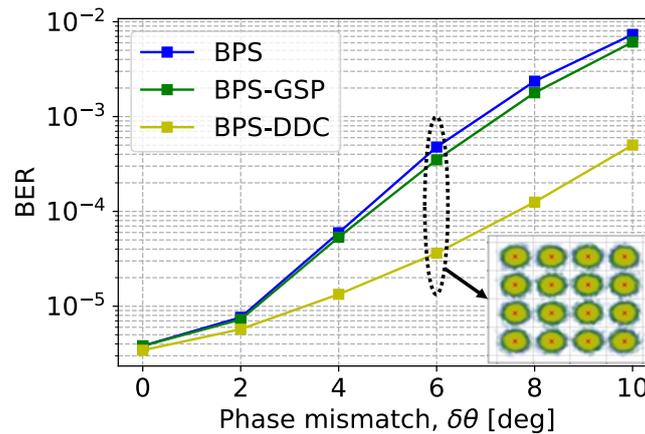


Figure 4.11: Performance of OPC transmission system with compensation schemes considering different phase mismatches.

The differences between schemes are hard to distinguish for up to the phase mismatch $\delta\theta = 2^\circ$, but become more obvious as the phase mismatch becomes larger. We can see that the BPS-GSP fails to offer a substantial BER improvement regardless the amount of phase mismatch at this long transmission distance. In contrast, our proposed BPS-DDC can provide a significant BER improvement for the phase mismatch which is even larger than 6° as seen in the previous analysis. The graph also highlights that the enhancement is still meaningful even when the phase mismatch is up to 10° . We also included in Fig. 4.11 the constellation diagrams

of the final symbols corresponding to different compensation schemes at the phase mismatch $\delta\theta = 6^\circ$. The colours of the points are matched with the colours of the BER curves. The corrected symbols with BPS-DDC (yellow points) are suppressed in both amplitude and phase compared to with those from other schemes (blue and green points). That is the evidence to support the capability of mitigating simultaneously phase and amplitude of our proposed DDC scheme.

4.5 Conclusion

In this chapter, we investigated the transmission with mid-link OPC and discussed where the residual dithering-induced phase distortion comes from and how it degrades the signal constellation. The main focus of this chapter is conversion mechanism which creates phase-to-phase and phase-to-amplitude at the received signal after the second fibre segment. The effect is caused by the interaction of residual dithering-induced phase distortion with channel dispersion, and becomes more significant when the transmission distance increases. To mitigate these distortions, we have developed a two-stage DSP compensation scheme, with assistance from BPS method. By estimate the residual dithering-induced phase distortion through linear regression algorithm, we could estimate and effectively remove both phase and amplitude distortion induced by imperfect pump-phase modulation from our transmitted signal and achieved a significant BER improvement over across wide range of number of spans for 28 Gbaud 16-QAM signal transmission.

Chapter 5

Phase and amplitude distortion compensation in optical parametric amplified systems

In this chapter, we extend our investigation into phase and amplitude distortion caused by pump dithering, focusing now on optical transmissions employing FOPAs. While these systems encounter challenges similar to those seen in OPC systems, the dithering-induced distortions in FOPA system are more complex and difficult to estimate. This complexity comes from the fact that distortions are introduced at multiples cascaded stages, with each contributing to the overall dithering-induced distortion. Moreover, the interaction with fibre dispersion accumulates over cascaded stages, further complicating the distortion pattern. In this chapter, we propose two novel online dithering compensation schemes: a multi-branch approach [3] and a single-branch method [5]. Both are designed for simultaneously mitigating of phase and amplitude distortion, with the latter is an improved version of the former. The single-branch compensation scheme is suitable for both single- and dual-pump FOPA configurations, as well as different pump-phase modulation schemes. The method and results presented within this chapter are also included in the CLEO-PR paper [7] and the Optics Express journal article [10].

5.1 Fibre-optical parametric amplified transmission

In this study, we examine a FOPA transmission link with N spans, as depicted in Fig. 5.1. Each span consists of 100 km of SMF, resulting in 20 dB of signal loss, which is subsequently compensated by a following FOPA.

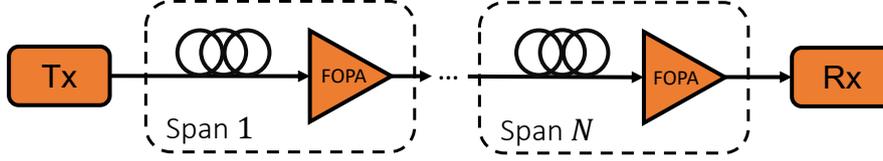


Figure 5.1: Schematic diagram of a transmission system with N cascaded FOPA stages.

The considered FOPA model incorporates both single-pump and dual-pump configurations, assuming no pump depletion. The FOPA complex signal gain for the model without pump-phase modulation is provided in Eq. (2.26). We used the same parameters listed in Tables 2.1 and 2.2 for the FOPA numerical model in this chapter. As discussed in Chapter 2, achieving a 25 dB FOPA gain, which is necessary to compensate for the 20 dB signal loss and an additional 5 dB insertion loss, requires an increase in the SBS threshold via pump-phase modulation. Therefore, to accurately model FOPA within the transmission link in Fig. 5.1, a new calculation for FOPA gain under the influence of pump-phase modulation is needed, which will be discussed in the next section.

5.2 Influence of pump-phase modulation

Assume the electrical driving signal within pump-phase modulation schemes is a combination of multiple N_t tones, as discussed in Chapter 2, given the definition $\varphi_t = \partial\varphi(t)/\partial t$ as the first derivative of the pump phase, we can write it as

$$\varphi_t = \sum_{i=1}^{N_t} A_{m_i} \omega_{m_i} \cos(\omega_{m_i} t + \xi_{m_i}), \quad (5.1)$$

where the selection of frequencies ω_{m_i} is detailed in section 2.4.2, while the amplitude A_{m_i} and phase ξ_{m_i} can be optimised as outlined in section 2.5. The phase modulation introduces a time-dependent instantaneous phase mismatch, $\delta\beta(t)$, which is added to the linear wavevector mismatch $\Delta\beta$ and subsequently to the total wavevector mismatch κ in Eqs. (2.15), (2.30), (2.36) and (2.37). This additional term is a function of φ_t and in the single-pump scheme, it is

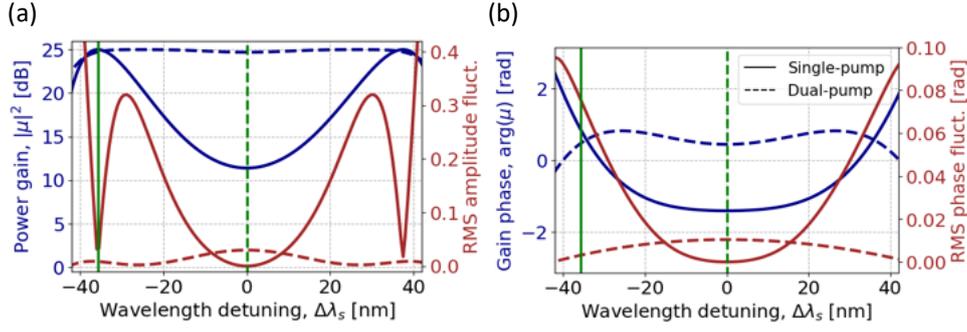


Figure 5.2: (a) The amplitude and (b) phase spectral responses of the FOPA complex gain, along with the corresponding dithering-induced RMS amplitude and phase fluctuations as a function of signal wavelength detuning.

expressed as [82, 83]

$$\delta\beta(t) = \beta^{(2)}\varphi_t^2 - \beta^{(3)}\left(\varphi_t\Delta\omega_s^2 + \frac{1}{3}\varphi_t^3\right) + \frac{1}{12}\beta^{(4)}\left(\varphi_t^4 + 6\varphi_t^2\Delta\omega_s^2\right). \quad (5.2)$$

This term for the dual-pump scheme can be expressed as [84]

$$\delta\beta(t) = \frac{1}{2}\beta^{(3)}(\Delta\omega_s^2 - \Delta\omega_p^2)(\varphi_{1,t} + \varphi_{2,t}), \quad (5.3)$$

where $\varphi_{1,t}$ and $\varphi_{2,t}$ are the first derivatives of the two pump phases corresponding to the pumps P_1 and P_2 . The pump phase modulation thus modulates the FOPA complex gain, originally given in Eq. (2.26), as

$$\tilde{\mu}_s(t) = \left[\cosh[g(t)L] + j\frac{\kappa(t)}{2g(t)} \sinh[g(t)L] \right] e^{[j(2\gamma P_0 - \frac{\kappa(t)}{2})]L}, \quad (5.4)$$

where $\kappa(t) = \kappa + \delta\beta(t)$ and the time-dependent $g(t)$ is calculated from Eq. (2.17) with the new $\kappa(t)$. Equation (5.4) simplifies to Eq. (2.26) when $\delta\beta(t) = 0$, which occurs when $\varphi_t = 0$ in the single-pump scheme (i.e. no pump-phase modulation) or $\varphi_{1,t} = -\varphi_{2,t}$ in the dual-pump scheme (i.e. the two pumps are ideally in phase opposition). In this analysis, RF amplifier is linear. Because the total wavevector mismatch κ is modulated by the pump-phase modulation (in single-pump scheme) or by residual dithering (in the dual-pump scheme), the FOPA gain modulation exhibits temporal variations in both amplitude and phase. The magnitude of these distortions depends on the signal frequency detuning $\Delta\omega_s$, as shown in Eqs. (5.2) and (5.3).

Quantifying the gain fluctuations is essential to select the optimal operating region for FOPA. To achieve this, we define the root-mean-square (RMS) amplitude fluctuation σ_{amp} and the RMS phase fluctuation σ_{phase} of the discrete time-varying FOPA complex gain $\tilde{\mu}_s[i]$ as

$$\sigma_{\text{amp}} = \sqrt{\sum_{i=1}^{N_S} (|\tilde{\mu}_s[i]| - |\mu_s|)^2}, \quad (5.5)$$

$$\sigma_{\text{phase}} = \sqrt{\sum_{i=1}^{N_S} (\angle\{\tilde{\mu}_s[i]\} - \angle\{\mu_s\})^2}, \quad (5.6)$$

where N_S is the number of samples and μ_s represents the FOPA complex gain without pump-phase modulation, calculated from Eq. (2.26). These dithering-induced fluctuations, plotted against signal wavelength detuning $\Delta\lambda_s$, are shown in Fig. 5.2. The fluctuations in this graph are caused by the four-tone phase modulation scheme. The graph also displays the power gain and phase spectral responses of the FOPA complex gain versus the signal wavelength detuning, calculated from the central wavelength $\Delta\lambda_s = \lambda_s - \lambda_c$ (in the dual-pump scheme) or from the pump wavelength $\Delta\lambda_s = \lambda_s - \lambda_p$ (in the single-pump scheme), in the absence of pump phase modulation. The RMS fluctuations in the dual-pump scheme are generally smaller than that in the single-pump scheme because of the relatively small residual dithering signal due to the imperfect tone adjustment transferred to the signal instead of the entire dithering signal contributed by one pump in the single-pump scheme. In detail, the magnitude of the first derivative of pump phase φ_t participating in modulating the FOPA gain is smaller in the dual-pump scheme than the single-pump one.

The maximum RMS fluctuations in both amplitude and phase in the dual-pump scheme can be seen at the zero wavelength detuning, i.e. $\Delta\lambda_s = 0$ nm, while that is where the RMS fluctuations in the single-pump scheme reach their minimum points. A local minimum of the RMS amplitude fluctuation in the single-pump scheme can be seen at $\Delta\lambda_s = -35.7$ nm, where the optimal gain is achieved. It is an interesting property because when we operate FOPA at this wavelength for maximum gain, only the phase distortion prevails, and the primary impact on the output signal mainly originates from the phase fluctuation of the FOPA gain. The RMS phase and amplitude fluctuations in the single-pump scheme then experience rapid growths as the wavelength detuning goes beyond this value. Figure 5.2 illustrates the selected operating wavelengths for both the single-pump and dual-pump schemes, including two notable wavelength detuning values mentioned above. The solid lines represent the optimal operating wavelength for the single-pump scheme that yields 25 dB power gain, and we will also investigate the dual-pump scheme at this wavelength. The dashed lines denotes another operating

wavelength for the dual-pump scheme, where the RMS fluctuations are highest, to test the capability of the proposed compensation scheme to phase and amplitude suppression.

5.3 Phase and amplitude distortion compensation schemes

In this section, we will propose DDC algorithms, which compensate for the dithering-induced phase distortion and also the phase-to-phase and phase-to-amplitude distortions due to the effect outlined in section 4.2.2. We will begin with the examination of the baseband equivalent model of the FOPA transmission system to see how the cascaded stages of phase distortion contribute differently than the single-stage in Chapter 4. We will then introduce a compensation scheme built from the legacy of the algorithm designed for mid-link OPC systems, followed by an advanced compensation scheme with a reduced complexity.

5.3.1 Operating principle

Similar to the algorithm designed for OPC systems, the proposed DDC methods here were developed through reverse engineering the transmission link, taking into account the precise interaction dynamics between dithering-induced phase fluctuations and the fibre chromatic dispersion. We first examine the equivalent baseband model, as depicted in Fig. 5.3, where the ASE noise is not shown for the sake of simplicity.

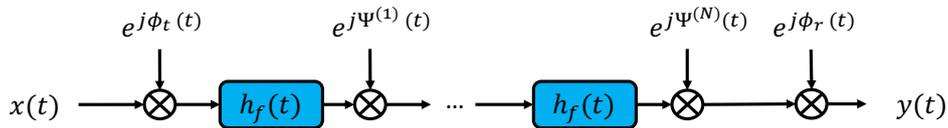


Figure 5.3: The baseband equivalent model of FOPA transmission.

The relationship between the input $x(t)$ and output $y(t)$ signal waveforms can be expressed as

$$y(t) = \left\{ \left[\left(x(t) e^{j\phi_t(t)} * h_f(t) \right) e^{j\Psi^{(1)}(t)} * \dots * h_f(t) \right] e^{j\Psi^{(N)}(t)} \right\} e^{j\phi_r(t)}, \quad (5.7)$$

where $*$ denotes the convolution operation. Since the effect of fibre nonlinearity is not considered in this analysis, each fibre span was modelled by a linear filter characterised by the impulse response of $h_f(t)$. The phase noise components, $\phi_t(t)$ and $\phi_r(t)$, originated from the laser sources at the transmitter and receiver, respectively, due to their nonzero linewidths.

Similar to the transmission with OPC, they were modelled as Wiener processes with increments that follow a Gaussian distribution [78]. At the n -th span, the total phase distortion is given by $\Psi^{(n)}(t) = \phi_p^{(n)}(t) + \phi^{(n)}(t)$, where $\phi_p^{(n)}(t)$ represents the phase noise of the pump laser, modelled in the same way as $\phi_t(t)$ and $\phi_r(t)$. The term $\phi^{(n)}(t)$ accounts for the dithering-induced phase distortion, which is caused by fluctuations in the complex gain of the FOPA driven by pump-phase modulation, as discussed in Chapter 2. By inverting the signal propagation, an approximate estimate of the transmitted signal $\hat{x}(t)$ is derived from the received signal $y(t)$ as

$$\hat{x}(t) = \left\{ \left[\left(y(t) e^{-j[\phi_r(t) + \Psi^{(N)}(t)]} * h_e(t) \right) \dots e^{-j\Psi^{(1)}(t)} \right] * h_e(t) \right\} e^{-j\phi_t(t)}, \quad (5.8)$$

where $h_e(t)$ represents the inverse response of each fibre span, satisfying the condition $h_e(t) * h_f(t) = \delta(t)$, where $\delta(\cdot)$ denotes the Dirac delta function. It is important to emphasise that phase variations due to non-zero linewidth of laser sources occur over time scales that are much longer than those associated with dithering-induced phase distortion. Therefore, we can neglect the interaction between laser phase noise and fibre dispersion. This allows us to focus exclusively on the interaction between dithering-induced phase distortion and chromatic dispersion in our analysis. As a result, Eq. (5.8) simplifies to

$$\hat{x}(t) = e^{-j\delta\phi(t)} \left\{ \left[\left(y(t) e^{-j\phi^{(N)}(t)} * h_e(t) \right) \dots e^{-j\phi^{(1)}(t)} \right] * h_e(t) \right\}, \quad (5.9)$$

where $\delta\phi(t) = \sum_n \phi_p^{(n)}(t) + \phi_t(t) + \phi_r(t)$ represents the accumulated phase noise contributed by all laser sources along the transmission link. By applying the small-signal approximation $e^{i\alpha} \simeq 1 + i\alpha$ to the dithering-induced phase fluctuations $\phi^{(n)}(t)$, where $n = 1, 2, \dots, N$, we can further simplify Eq. (5.9) to

$$\hat{x}(t) \simeq y'(t) - \sum_{n=1}^N \left\{ \left[\left(y'(t) * h_f^{(n)}(t) \right) j\phi^{(n)}(t) \right] * h_e^{(n)}(t) \right\}, \quad (5.10)$$

where $h_f^{(n)}(t)$ and $h_e^{(n)}(t)$ refer to the impulse response and inverse impulse response, respectively, of an equivalent linear filter that accumulates the effects of n spans. In detail, $h_f^{(n)}(t)$ can be expressed as the convolution of n individual filters, i.e. $h_f^{(n)}(t) = h_f(t) * \dots * h_f(t)$. A similar expression applies for the inverse impulse response $h_e^{(n)}(t)$. The first term, $y'(t)$, represents the signal after being compensated for both laser phase noise and chromatic dispersion, i.e.

$$y'(t) = \left[y(t) * h_e^{(N)}(t) \right] e^{-j\delta\phi(t)}. \quad (5.11)$$

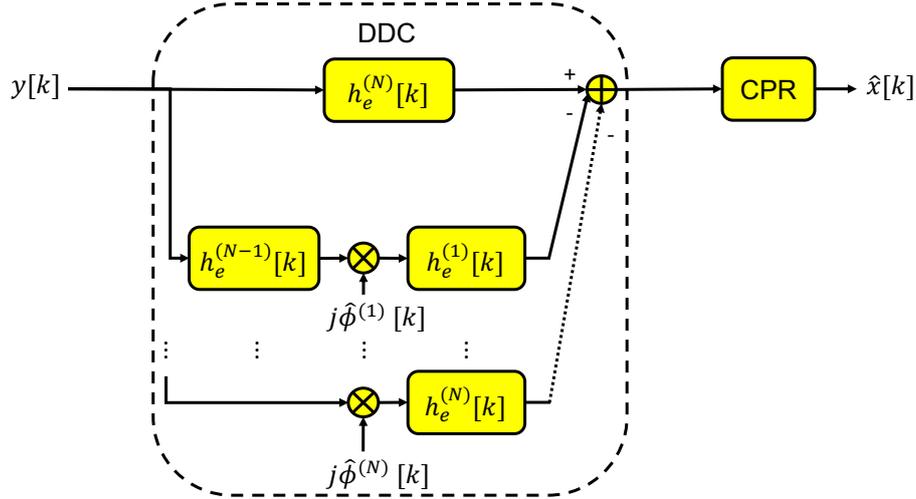


Figure 5.4: Proposed multi-branch dithering-induced distortion compensation scheme.

Eq. (5.10) reveals that the conventional DSP output, denoted by $y'(t)$, deviates from the transmitted signal due to the presence of the second term, where the dithering-induced phase distortion $\phi^{(n)}(t)$ plays a significant role. We can also observe that this distortion becomes increasingly pronounced as the number of fibre spans increases. This observation suggests the need to develop a scheme for estimating and compensating for this distortion in the received signal.

5.3.2 Multi-branch compensation scheme

This scheme will focus on the estimation of the dithering-induced phase distortions $\phi^{(n)}(t)$ values, and we use them to recreate the resulting complex distortion, which is the second term in Eq. (5.10). We can rewrite Eq. (5.10) as

$$\hat{x}(t) \simeq e^{-j\delta\phi(t)} \left(y(t) * h_e^{(N)}(t) - \sum_{n=1}^N \left\{ \left[\left(y(t) * h_e^{(N-n)}(t) \right) j\phi^{(n)}(t) \right] * h_e^{(n)}(t) \right\} \right). \quad (5.12)$$

We proposed the compensation scheme based on this equation, with the principle shown in Fig. 5.4. The CPR algorithm employed here is the BPS method, known for its efficiency in hardware implementation [66]. The compensation scheme is constructed by $N + 1$ branches, with N lower branches corresponding to N terms within the second term in Eq. (5.12). Given the FIR filter impulse responses $h_e^{(n)}[k]$ can be estimated by [52], we are only required to estimate the phase distortion terms $\phi^{(n)}[k]$. It can be done by learning the coefficient vectors

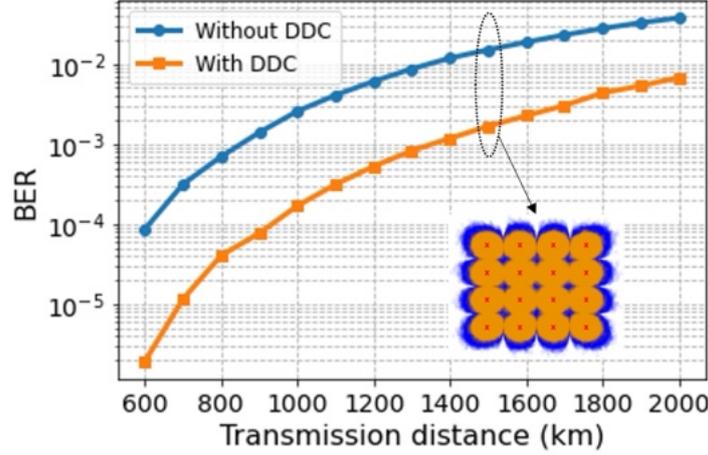


Figure 5.5: Performance comparison between the conventional carrier phase recovery algorithm and the proposed compensation algorithm. The inset illustrate the constellation diagrams after 1500 km.

$\theta^{(n)}$ in the linear regression $\phi^{(n)}[k] = (\theta^{(n)})^T \mathbf{B}^{(n)}[k]$, where the feature vector $\mathbf{B}^{(n)}[k] \in \mathbb{R}^{2N_t}$, defined as

$$\mathbf{B}^{(n)}[k] = [\sin[\omega_{m_1} k], \cos[\omega_{m_1} k], \dots, \sin[\omega_{m_{N_t}} k], \cos[\omega_{m_{N_t}} k]]^T, \quad (5.13)$$

where ω_{m_i} are the dithering frequencies which are assumed to be known. We now can express the output of each branch and subsequently the predicted symbol $\hat{x}[k]$ with respect to $\theta^{(n)}$. Given the reference transmitted symbol $x[k]$ within the training phase, we can use the complex LMS algorithm [85] to solve for $\theta^{(n)}$, which defines the dithering-induced phase distortion $\phi^{(n)}[k]$. In detail, we evaluated the error contributed by each branch, and update each $\theta^{(n)}$ accordingly based on that. We can then produce the predicted symbol $\hat{x}[k]$ using the scheme in Fig. 5.4.

The performance of this multi-branch DDC algorithm is compared against the scheme with a single stage of BPS algorithm, as shown in Fig. 5.5. We achieved a consistent BER improvement of around one order of magnitude compared to the scheme without DDC. The constellation diagrams obtained at 1500 km are also included to show the capability of suppression of both phase and amplitude distortion. However, the number of branches depends on the number of FOPA cascaded stages makes this algorithm hard to scale as we have more fibre spans.

5.3.3 Single-branch compensation scheme

In this section, we plan to reduce the computational cost with a single branch of compensation. The complete DSP chain, incorporating the proposed single-branch DDC algorithm, is illustrated in Fig. 5.6. The received signal $y[k]$ at time instant k is obtained after matched filtering and downsampling, which are omitted here for simplicity. The signal $y'[k]$ refers to the output of a conventional DSP chain, which includes CDC and CPR [52],

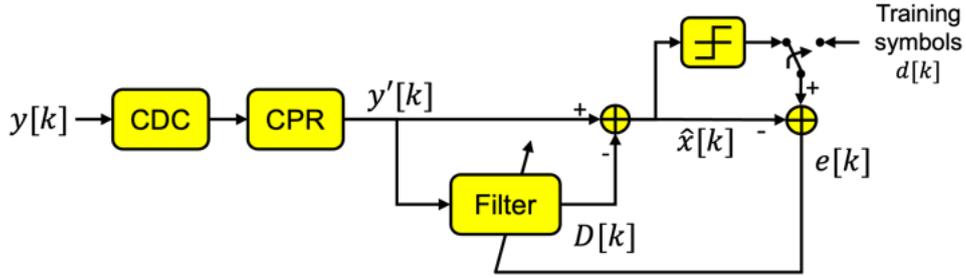


Figure 5.6: Dithering-induced distortion compensation in the DSP chain.

In this stage, we aimed to estimate the second term in Eq. (5.10) as the discrete-time dithering-induced complex distortion $D[k]$, i.e.

$$D[k] = \sum_{n=1}^N \left\{ \left[\left(y'[k] * h_f^{(n)}[k] \right) j\phi^{(n)}[k] \right] * h_e^{(n)}[k] \right\} \quad (5.14)$$

The channel impulse response $h_f^{(n)}[k]$ can be expressed as a sum of scaled and shifted Dirac delta functions, i.e.

$$h_f^{(n)}[k] = \sum_{m=-\infty}^{\infty} h_{f,m}^{(n)} \delta[k - m], \quad (5.15)$$

where $h_{f,m}^{(n)}$ represents the m -th tap coefficient of the filter $h_f^{(n)}[k]$. Consequently, the convolution $y'[k] * h_f^{(n)}[k]$ can be expressed as

$$y'[k] * h_f^{(n)}[k] = \sum_{m=-\infty}^{\infty} y'[k - m] h_{f,m}^{(n)}. \quad (5.16)$$

By applying the same procedure to the inverse impulse response $h_e^{(n)}[k]$, we can finally represent the dithering-induced complex distortion $D[k]$ as

$$D[k] = j \sum_{m,m'=-M/2}^{M/2} y'[k - m - m'] \sum_{n=1}^N \left(h_{f,m-m'}^{(n)} \phi^{(n)}[k - m'] h_{e,m'}^{(n)} \right), \quad (5.17)$$

where $M/2$ represents the filter delay. Since all impulse responses considered here are

assumed to be finite, we can constrain the memory of this function to $M + 1$, which depends on the number of spans N , to reduce the algorithm's computational complexity. The term within the bracket in Eq. (5.17), which is $h_{f,m-m'}^{(n)}\phi^{(n)}[k - m']h_{e,m'}^{(n)}$, is a function of dithering-induced phase distortion $\phi^{(n)}[k - m']$, which shares the same sinusoidal waveform as $\phi^{(n)}[k]$. Equation (5.17) indicates that we can estimate the dithering-induced complex distortion $D[k]$ by passing the conventional DSP output $y'[k]$ through an adaptive time-varying digital filter. This filter's tap coefficients, $h_{f,m-m'}^{(n)}\phi^{(n)}[k - m']h_{e,m'}^{(n)}$, are functions of the $\phi^{(n)}[k]$ terms.

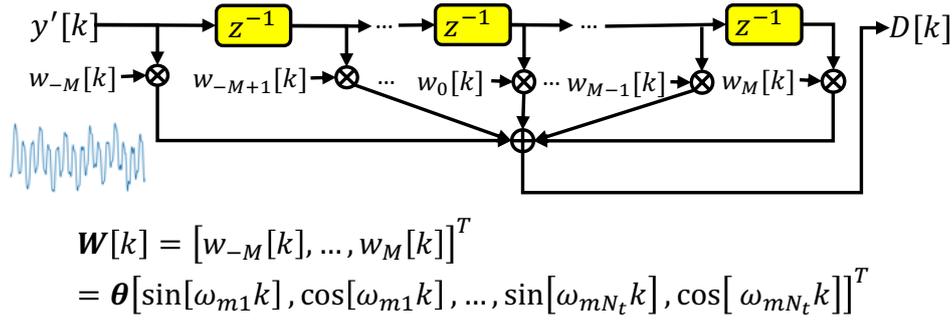


Figure 5.7: Dithering-induced distortion estimation filter.

We thus propose a filter structure to estimate the dithering-induced complex distortion $D[k]$ as shown in Fig. 5.7. The tap coefficients $w_{-M}[k]$, $w_{-M+1}[k]$, ..., $w_0[k]$, ..., $w_{M-1}[k]$, $w_M[k]$ can be learned using a linear regression approach. These coefficients are functions of the first derivative of the pump phase φ_t , as defined in Eq. (5.1). They are predicted by fitting a parametric model based on the known modulation frequencies f_{m_j} or ω_{m_j} as described in Eq. (5.1). We thus can define the feature vector $\mathbf{B}[k]$ similar to that in Eq. (5.13). By defining the coefficient vector $\boldsymbol{\theta}[k] \in \mathbb{C}^{(2M+1) \times 2N_t}$ such that $\mathbf{W}[k] = \boldsymbol{\theta}[k]\mathbf{B}[k]$, we can determine the vector of time-varying filter taps $\mathbf{W}[k] = [w_{-M}[k], \dots, w_M[k]]^T \in \mathbb{C}^{(2M+1) \times 1}$. The dithering-induced complex distortion $D[k]$ is then estimated as $D[k] = \mathbf{W}^T[k]\mathbf{Y}'[k]$, where $\mathbf{Y}'[k] = [y'[k - M], \dots, y'[k], \dots, y'[k + M]] \in \mathbb{C}^{(2M+1) \times 1}$ is a signal block of $y'[k]$ of size $(2M + 1)$. The transmitted signal at time instant k can be recovered as $\hat{x}[k] = y'[k] - D[k] = y'[k] - \mathbf{B}^T[k]\boldsymbol{\theta}^T[k]\mathbf{Y}'[k]$. The objective is now to update the coefficient vector $\boldsymbol{\theta}[k]$ to minimise the difference between $\hat{x}[k]$ and the reference symbol $x_{\text{ref}}[k]$. This can be achieved using the complex LMS algorithm [85], with the feature vector $\mathbf{B}[k]$ and the signal block $\mathbf{Y}'[k]$ are engineered at the beginning of each update cycle. The complete flow chart detailing the steps within each cycle is shown in Fig. 5.8.

The complex LMS algorithm updates the coefficient vector $\boldsymbol{\theta}[k]$ using the complex error $e[k]$,

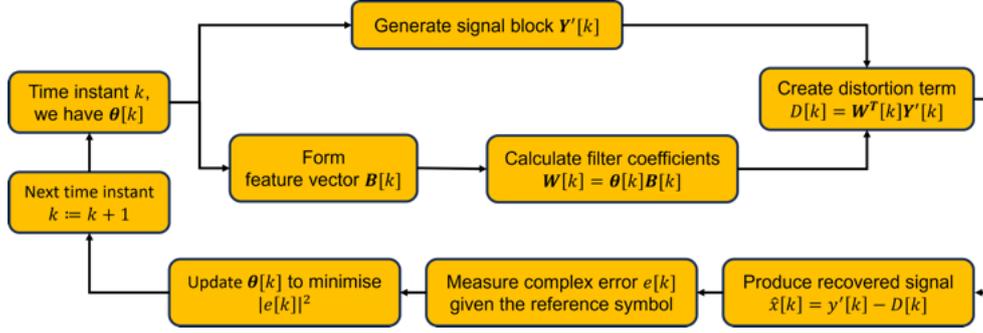


Figure 5.8: Flow diagram of dithering-induced distortion compensation algorithm.

which is the difference between the reference symbol and the recovered symbol. This update happened at the symbol speed. During the training phase, the reference symbol is provided as the training symbol $d[k]$, while outside the training phase, it is generated by a decision circuit based on the known symbol alphabets, as depicted in Fig. 5.6.

5.4 Compensation results

We conducted numerical simulations of the transmission of a single-polarisation 16-QAM signal across up to $N = 12$ fibre spans. The transmitted signal is pulse-shaped using a RRC filter with a roll-off factor of 0.1 and is upsampled to 8 samples per symbol. The details of the transmission and FOPA parameters are provided in Table 5.1. To prevent the coherent accumulation of dithering-induced distortion across multiple FOPA stages, we introduced a random time shift in the pump-phase modulation at each FOPA stage. In practice, each FOPA operates independently, often separated by hundreds of kilometres, meaning that the distortions introduced at different FOPA stages do not remain phase-aligned but instead exhibit random relative phase differences. That is why, to replicate this effect in our simulations, we applied a random time shift, uniformly distributed within the range $[0, 2\pi]$ radians, at each FOPA stage to create a more realistic transmission environment. The system's performance was evaluated using the Q^2 -factor, calculated as

$$Q^2[\text{dB}] = 20 \log(\sqrt{2} \text{erfc}^{-1}(2 \times \text{BER})), \quad (5.18)$$

where BER represents the ratio between the directly-counted error bits and the total bits of 10×2^{16} symbols under testing.

Parameters	Values
Modulation	16-QAM
Baud rate	28 [GBaud]
Span length	100 [km]
Loss	0.2 [dB/km]
Dispersion	17 [ps/(nm.km)]
Nonlinearity	1.2 [$W^{-1}km^{-1}$]
Tx/Rx laser linewidths	50 [kHz]
FOPA laser linewidths	30 [kHz]
Noise figure	4.5 [dB]

Table 5.1: FOPA transmission numerical simulation parameters

5.4.1 Performance over number of spans

Figure 5.9 shows the performance comparison between the schemes without and with our proposed DDC algorithm, measured in terms of Q^2 -factor versus the number of fibre spans. The comparison was conducted across three different scenarios: the FOPA single-pump scheme, and two FOPA dual-pump schemes - one with the same operating wavelength as the single-pump scheme ($\Delta\lambda_s = -35.7$ nm) and the other with zero wavelength detuning ($\Delta\lambda_s = 0$ nm). The single-pump FOPA is operated only at $\Delta\lambda_s = -35.7$ nm because, unlike dual-pump configuration, which provides a flat gain spectrum, the single-pump scheme achieves its maximum power gain only at this specific wavelength. Operating it at any other wavelength would not provide sufficient compensation for signal attenuation and insertion loss, which total 25 dB. For the dual-pump FOPA, in addition to testing at $\Delta\lambda_s = -35.7$ nm, we also evaluate the compensation performance at $\Delta\lambda_s = 0$ nm. This allows us to verify the effectiveness of our proposed algorithm across different operating wavelengths, leveraging the flat gain spectrum of the dual-pump FOPA.

A four-tone pump-phase modulation was employed, with a base frequency of $\nu_{m1} = 60$ MHz, resulting in a frequency set of [60, 180, 540, 1620] MHz. In the non-DDC schemes, only CDC and CPR blocks were applied, with the latter using the BPS algorithm with a test-phase resolution of $B = 32$ phase angles [66]. The filter lengths for both the BPS and the DDC will be optimised, as discussed later.

The FOPA dual-pump scheme at $\Delta\lambda_s = -35.7$ nm represented by the red curves clearly perform better than the other scenarios. As shown in Fig. 5.2, operating FOPA at this wavelength exhibits the smallest fluctuations in both amplitude and phase among the pumping schemes considered. Our analysis indicated that these curves are very close to the theoretical limit in which transmission does not have any penalties from pump-phase modulation.

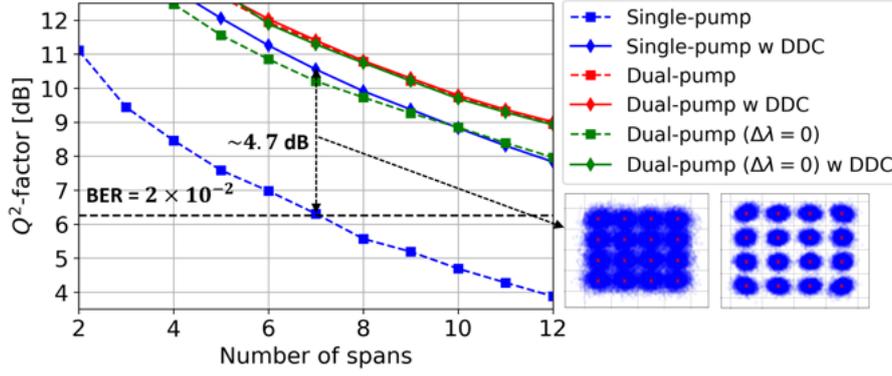


Figure 5.9: The performance of the proposed compensation scheme in terms of Q^2 -factor versus number of spans.

Given that the distortion impact in this scenario is minimal, there is limited potential for further enhancement by our proposed DDC algorithm. Therefore, the difference between the curves without and with DDC is not significant. The performance when operating the FOPA at $\Delta\lambda_s = 0$ nm is depicted by the green curves. The increased amplitude and phase fluctuations observed in Fig. 5.2 manifest as noticeable performance degradation compared to the previous scenario. In this case, our proposed DDC algorithm demonstrated its effectiveness by raising the Q^2 -factor to the level comparable to those achieved at the optimal wavelength. In the single-pump scheme, the significantly higher phase fluctuation, coupled with their interaction with fibre dispersion, results in a substantial drop in system performance. Nevertheless, our DDC algorithm consistently provides a significant improvement in Q^2 -factor as the number of fibre spans increases, achieving an approximate 4.7-dB Q^2 improvement at a BER level of 2×10^{-2} , after 7 spans. The insets in Fig. 5.9 show the constellation diagrams that corresponds to this improvement at 700 km.

5.4.2 Amplitude and phase suppression capability

Fig. 5.10 illustrates the accumulated RMS fluctuations in both phases and amplitude of the received symbols relative to the transmitted symbols, plotted as functions of the number of spans. The RMS fluctuations in amplitude, Δ_{amp} , and in phase, Δ_{phase} , were calculated as

$$\Delta_{\text{amp}} = \sqrt{\frac{1}{N_s} \sum (|\hat{x}| - |x_{\text{ref}}|)^2}, \quad (5.19)$$

$$\Delta_{\text{phase}} = \sqrt{\frac{1}{N_s} \sum (\angle \hat{x} - \angle x_{\text{ref}})^2}, \quad (5.20)$$

where $N_s = 2^{16}$ is the number of symbols and the calculations were taken and averaged over 10 batches. Recall that \hat{x} represents the recovered symbol, while x_{ref} denotes the reference symbol. In the non-DDC scenario, \hat{x} is obtained after processing through only conventional DSP blocks, which include CDC and CPR.

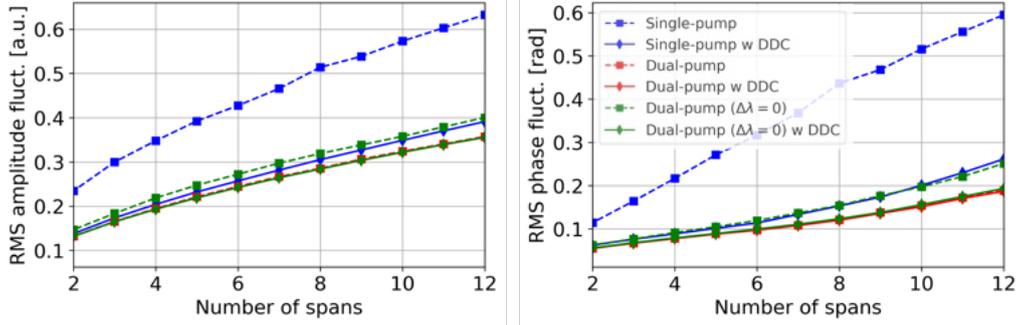


Figure 5.10: The evolution of the RMS phase and amplitude fluctuations.

The results shown in Fig. 5.10 closely mirror the performance trends observed in the Q^2 -factor analysis. The RMS fluctuation levels for the FOPA at $\Delta\lambda_s = 0$ were effectively reduced to levels similar to those of the FOPA dual-pump scheme operating at the optimal wavelength of $\Delta\lambda_s = -35.7$ nm. Although the single-pump scheme exhibited very high fluctuation levels with steep slopes as the number of spans increases, significant improvements were still achieved. With the application of the DDC algorithm, the curves experienced much gentler slopes and notably lower RMS fluctuations. This behaviour confirmed the effectiveness of our proposed DDC algorithm in simultaneously mitigating fluctuations in both signal amplitude and phase.

5.4.3 Tolerance to dithering frequency deviations

In our algorithm, constructing the feature vector $\mathbf{B}[k]$ from the given dithering modulation frequencies is a crucial step. To assess the resilience of our DDC algorithm to deviations in these frequencies from their expected values, we evaluated its performance under varying conditions. In detail, the dithering modulation frequencies were not fixed but instead followed a normal distribution with mean values matching the expected frequency set, i.e. [60, 180, 540, 1620] MHz, and a standard deviation denoted as σ_f . Figure 5.11 demonstrates the Q^2 improvement achieved by our proposed DDC algorithm over conventional DSP after 7 fibre spans, plotted against the standard deviation of the dithering modulation frequencies, σ_f .

We observed that there is no noticeable reduction in Q^2 -factor enhancement for both the

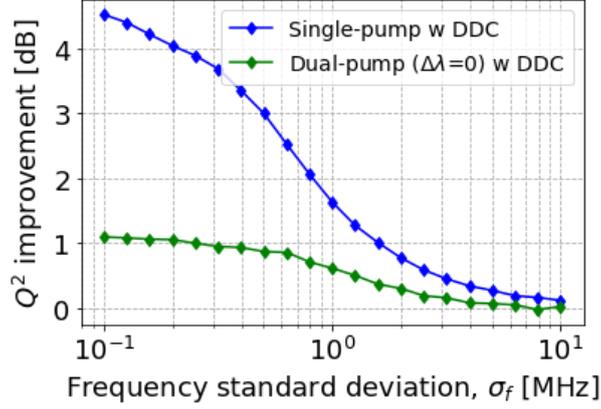


Figure 5.11: Performance improvement due to DDC algorithm versus the dithering frequency variation.

single-pump scheme and the dual-pump scheme at $\Delta\lambda_s = 0$ when $\sigma_f = 0.1$ MHz. The dual-pump scheme at the optimal wavelength of $\Delta\lambda_s = -35.7$ nm was excluded from this analysis since it offers minimal room for improvement. However, as σ_f increases, a predictable decline in the performance of both schemes was evident. At $\sigma_f = 1$ MHz, our DDC method still achieved approximately half the improvement seen under the ideal condition when $\sigma_f = 0$. This indicated that maintaining accurate RF frequency generation in electronic components is important for preserving a decent system performance. In practice, the accuracy of these devices typically ensures that frequency variations remain within 1 MHz, which supports the effectiveness of our DDC algorithm under realistic operating conditions.

5.4.4 Base frequency

We already know that the RMS fluctuations in amplitude and phase depend on the first derivative of the pump-phase modulation, φ_t , as defined in Eq. (5.1). The modulation frequencies ω_{mi} and consequently the broadened bandwidth of the pump spectrum due to phase modulation play a crucial role in determining the extent of dithering-induced distortion. For a given pump-modulation scheme, while the number of frequencies is fixed and their relative locations are selected based on the principles discussed in Chapter 2, the base frequency ν_{m1} is the key parameter which determines the broadened bandwidth of the pump spectrum. In this analysis, we assessed the performance of our DDC algorithm, under a four-tone pump-modulation scheme, focusing on how it varies with the base frequency ν_{m1} . The results are plotted in Fig. 5.12 for both the single-pump scheme and the dual-pump scheme with $\Delta\lambda_s = 0$, after 7 and 12 spans.

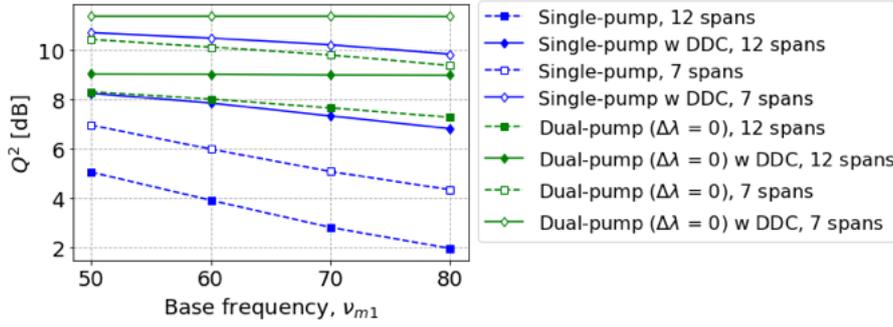


Figure 5.12: Q^2 -factor performance as a function of the base frequency ν_{m1} .

In the single-pump scheme, we observe a sharp decline in the Q^2 -factor performance curves without the DDC algorithm as the base frequency ν_{m1} increases from 50 to 80 MHz, resulting in a total decrease of around 3 dB within this region. This performance degradation is attributed to the higher RMS amplitude and phase fluctuations caused by the larger first derivative term φ_t , which is directly related to the base frequency ν_{m1} . The DDC algorithm demonstrated its effectiveness by mitigating this decline, resulting in more gradual curves corresponding to both considered transmission lengths, with a maximum degradation of only around 1 dB. In the dual-pump scheme, the DDC algorithm nearly eliminates these decline trends, making the Q^2 -factor performance largely independent of the base frequency ν_{m1} .

5.4.5 Filter optimisation process

The averaging filter length, denoted by $2K + 1$, and the number of filter taps, denoted by $2M + 1$, are critical parameters that shape the responses of the BPS and the DDC algorithm respectively. Figure 5.13 demonstrates the process of selecting the optimal filter lengths for both BPS and DDC, plotted for different values of base frequency ν_{m1} . While increasing the number of filter taps generally enhances the Q^2 -factor performance, this improvement comes at the cost of higher computational complexity. Therefore, the optimal filter lengths for both algorithms were chosen as the smallest values beyond which further increases return no substantial performance gain.

In the BPS graph shown in Fig. 5.13, the rapid increase in Q^2 -factor performance highlights the necessity of extending the BPS filter length to address the phase cycle slip issue. The cycle slips occur when estimated phase time series by the algorithm experiences a discontinuity by a multiple of $\pi/2$, due to incorrect symbol detection by the decision-directed circuit. This is

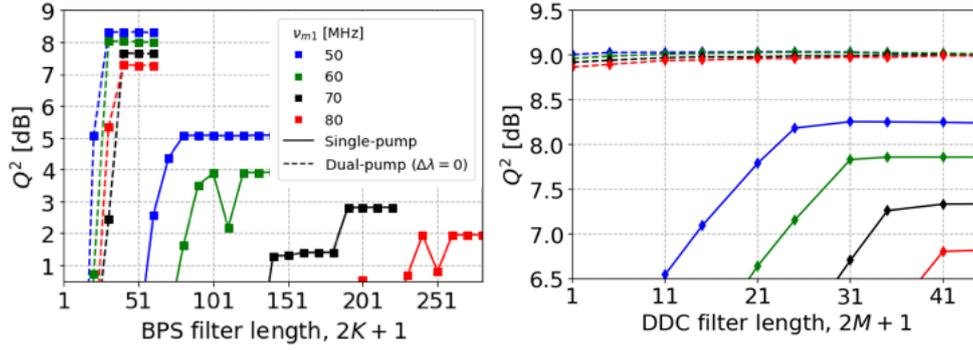


Figure 5.13: The filter optimisation for BPS and DDC algorithms.

often influenced by high-speed phase variations, such as dithering-induced phase distortion. To mitigate this effect, a larger block size is required since higher BPS filter length ensures that the output symbol decision is more heavily influenced by neighbouring symbols, thereby reducing the chance of overreacting to sudden phase changes. We also observe that a longer BPS filter length is required when either a higher base frequency ν_{m1} is used or when single-pump scheme is employed. This is due to the increased RMS fluctuations in the phase of the FOPA output signal in both scenarios, which in turn makes the decision-directed circuit more prone to incorrect detection. A longer filter length is thus necessary in these cases to stabilise the phase estimation.

For the DDC algorithm, a similar trend is observed where a longer DDC filter length generally provides better compensation capability. This is because the filter needs to account for sufficient memory to capture the interactions between the dithering-induced phase distortion and fibre dispersion, as described in Eq. (5.17). However, like the BPS algorithm, this improvement plateaus once a sufficient number of neighbouring symbols are included. The optimal DDC filter length scales linearly with the base frequency ν_{m1} due to the larger phase fluctuations in the FOPA gain, which was discussed in the previous subsection. It results in stronger inter-symbol interference, thereby necessitating a higher-order filter for effective compensation.

5.5 Conclusion

We have presented a new online DSP algorithm to simultaneously mitigate phase and amplitude distortions caused by pump-phase modulation and its interaction with the chromatic dispersion in transmissions using multiple cascaded FOPAs. Integrated at the of end of the conventional

DSP chain, which includes CDC and CPR, the algorithm effectively estimates and removes the complex distortion accumulated during transmission. Our numerical simulations have demonstrated its effectiveness through the improvements in Q^2 -factor performance across the different FOPA pump schemes, including an approximate 4.7-dB gain for 28-GBaud 16-QAM transmission with a single-pump scheme after 7 spans. This also indicates a substantial increase of transmission reach at a given Q^2 -factor level. Furthermore, the performance improvement over conventional DSP achieved for a dual-pump FOPA operating around the centre of its gain bandwidth indicates that the algorithm can be advantageously deployed in wavelength-division multiplexing scenarios. The proposed scheme is important for future FOPA transmission designs, where pump-phase modulation is essential for achieving the required gain. In the next chapter, we will explore a new approach to this problem: a non-parametric model for distortion estimate.

Chapter 6

Kernel-based compensation in optical parametric amplified systems

In this chapter, we will investigate the dithering-induced distortion compensation problem using kernel-methods. This represents a step forward in advancing our proposed algorithm from the previous chapter, employing a non-parametric approach that avoids explicit engineering of the dithering-induced distortion's feature vector. By allowing the algorithm to learn the data pattern itself, we help it generalise better to unseen data. Another objective we aim to achieve is avoiding the two-stage approach and seeking a one-stage compensation scheme that can also compensate for laser phase noise. We will propose three kernel-based compensation algorithms within this chapter: one for phase-only compensation, published in the CSNDSP paper [6], and two others with complex-valued processing capability for phase and amplitude distortion compensation, published in the ECOC paper [8] and IPC paper [9].

We will begin this chapter by continuing the discussion of kernel methods in section 3.4. There are two issues we need to address in applying kernel methods to our dithering compensation problem. Firstly, the phase and amplitude distortions in the received signal result not only from sinusoidal pump dithering but also from laser phase noise. Since we aim for a single block that can compensate for both, the time series representing the phase distortion cannot be captured by a single periodic kernel or the Gaussian kernel. The second issue with kernel-based learning is the growth of the dictionary, which adds complexity when calculating the kernel matrix. An online algorithm cannot handle this computational cost, so a proper method to prune the data dictionary must be introduced. We will discuss the solution for the first problem in section

6.1, where a custom kernel is created from fundamental kernel functions. The second problem will be addressed in section 6.2, where a sliding-window approach makes the algorithm suitable for online operation.

6.1 Combining kernels

In many real-world scenarios, the underlying function we aim to model often exhibits multiple types of structures, making it difficult to capture its behaviour using a single kernel. Each fundamental kernel, tailored for a specific pattern type, may fail to accurately model complex relationships. In these cases, combining different kernels provides a more flexible and expressive approach, allowing a richer variety of characteristics to be captured [86].

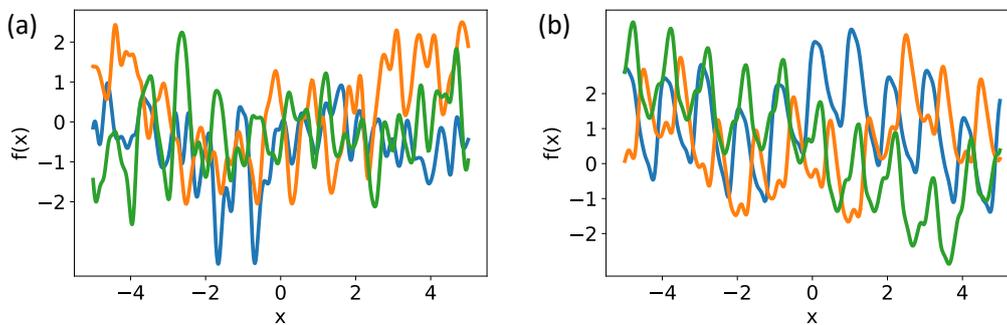


Figure 6.1: Example functions sampled from GP prior when using a combined kernel from Gaussian and periodic kernels by (a) multiplying them and (b) adding them.

There are two primary ways to combine kernels: addition and multiplication [87], and from which more complex kernels can be created. We consider the combination of two kernels defined in section 3.4.4, Gaussian kernel and the periodic kernel, to demonstrate the purpose of each operation.

Multiplying kernels models the interaction between different function behaviours, working as an AND operation [87]. In this case, the combined kernel returns a high value only when both contributing kernels yield high values simultaneously. For example, if two input points \mathbf{x} and \mathbf{x}' are in the same phase of oscillation (as defined by the periodic kernel) but are sufficiently far apart (as measured by the Gaussian kernel), the combined kernel will not return a high value. Therefore, the values of $f(\mathbf{x})$ and $f(\mathbf{x}')$ will not be close, even though they should be similar when using the periodic kernel alone. The result is that the functions sampled from the GP defined by the combined kernel are periodic, but with smoothly varying amplitudes depending

on the lengthscale of the Gaussian kernel. This behaviour can be seen in the example functions shown in Fig. 6.1(a). Multiplying the Gaussian kernel and periodic kernel is particularly useful for modelling functions with a periodic pattern influenced by underlying trends.

In contrast, adding kernels works like an OR operation [87], allowing us to capture pattern types independently. In this case, there is no interaction between the function behaviours, with each contributing separately to the overall pattern. For example, when we add a Gaussian kernel with a periodic kernel and sample functions from the corresponding GP, the result can be seen in Fig. 6.1(b). The sampled functions exhibit periodic pattern with constant amplitude, a characteristic inherited from the periodic kernel. Meanwhile, the Gaussian kernel contributes to the smooth variations in the slopes of these periodic functions.

In summary, combining kernels by multiplying or adding can be used to explicitly encode structural assumptions about complicated underlying functions. This method is particularly useful when prior knowledge is available regarding the phenomena being modelled. Further refinement of this approach could involve automatic kernel construction, where a suitable model is discovered through a search process [88]. However, this advanced topic lies beyond the scope of this thesis.

6.2 Sliding-window kernel-based method

In this section, we will discuss why we need to develop a sparsification strategy for kernel method learning with large datasets. We will then introduce the online sparsification approach we have chosen for further analysis, which can reduce the computation burden of the original kernel method.

We recall that if N is the number of data points, the kernel matrix \mathbf{K} has a size of $N \times N$. As N increases, which is the case in optical receivers where data arrives sequentially, the kernel method faces scalability issues. The computational and memory costs are $O(N^3)$ and $O(N^2)$, respectively, and they grow significantly. Memory resources become limited for storing new data, and computational time can exceed the incoming data interval [89]. Sparsification is needed to reduce both costs by discarding less relevant data points from the dictionary. Moreover, it helps avoid overfitting to the training data. By considering only the most relevant points, we enable the algorithm to generalise better to future data [90].

In online mode, we receive the input \mathbf{x}_n at time step n and are required to make the predicted

output \hat{y}_n corresponding to that input. The learning is then evaluated by an error quantity $y_n - \hat{y}_n$, where y_n is the true output. The algorithm updates its solution to minimise the cost function, which depends on the error, and uses the learned solution for the next available data. When the kernel method operates in online mode, the solution α of the kernel ridge regression in Eq. (3.26) is updated at each time step n as

$$\alpha_n = (\mathbf{K}_n + \lambda \mathbf{I})^{-1} \mathbf{y}_n, \quad (6.1)$$

where \mathbf{K}_n is the kernel matrix \mathbf{K} at time step n , and \mathbf{y}_n is the corresponding desired target. A naive approach to kernel learning involves the growth in size of the kernel matrix \mathbf{K}_n , as illustrated in Fig. 6.2(a). This approach faces challenges in terms of memory and computational costs, as discussed above.

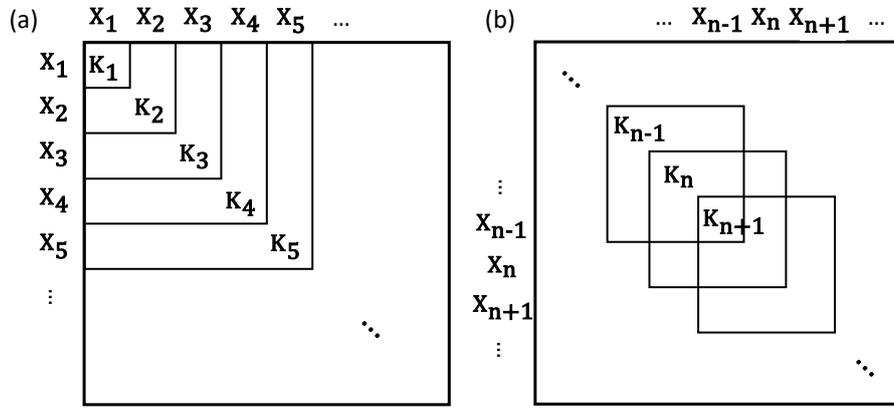


Figure 6.2: (a) Kernel matrices \mathbf{K}_n with growing sizes. (b) Kernel matrices \mathbf{K}_n with fixed sizes in sliding-window approach.

A standard approach to solving this problem is to set the number of data points considered, or the dictionary, N in Eq. (3.20) a finite and fixed value as the algorithm runs. This can be achieved by setting growth and pruning criteria. The former decides whether a new data point \mathbf{x} is admitted into the dictionary, while the latter provides a benchmark to discard less informative data points. One simple criterion for growth is the coherence criterion [91], where the similarities between the new point \mathbf{x} and existing points \mathbf{x}_i in the dictionary are measured. The maximum value among these is then compared to a coherence threshold to ensure that the new point added is not too similar to the existing points. A more complex growth criterion, known as approximate linear dependency (ALD), is outlined in [92]. It checks whether the new basis vector $\theta(\mathbf{x})$ in the feature space is linearly independent or nearly independent of the existing basis vectors. Regarding pruning approaches, a criterion based on the function

representation $f(\mathbf{x}) = \sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x})$ was proposed. In detail, any existing data point \mathbf{x}_i with the smallest coefficient magnitude $|\alpha_i|$ can be discarded [93], as it corresponds to the basis vector that contributes least to the function representation. Among other approaches, a simple and straightforward method is to discard the oldest data points. This is part of the sliding-window approach [94], where the algorithm admits a new available point and discards the oldest one. The evolution of kernel matrices \mathbf{K}_n in this approach is demonstrated in Fig. 6.2(b). This approach is useful in scenarios where the statistical properties of the data change over time, such as variations in the amplitudes and phases of sinusoidal waves. In such cases, older data points may become less relevant to the current state of the system, making the sliding-window approach a suitable choice for maintaining an adaptive model. A more complicated pruning approach, known as least squares error or greedy error pruning, can be found in [95]. In this method, the point to discard is the one whose removal from the dictionary results in the smallest increase in the squared error $|y_n - \hat{y}_n|^2$, which corresponds to the point that contributes least to the overall prediction.

In the scope of this thesis, we opt for the sliding-window approach due to its compatibility with evolving data characteristics and its reduced complexity. In the next sections, we will delve into the application of the kernel method for phase and amplitude compensation in optically parametric amplified links.

6.3 Kernel-based phase compensation using real combined kernel

In this chapter, we consider a FOPA link, as illustrated in Fig. 6.3. In each stage, we assume a 20-dB linear loss, equivalent to 100 km of SMF, which is then compensated by a single-pump FOPA. Our FOPA operates at the optimal wavelength, where it yields a power gain of 25 dB to account for an additional 5 dB of device insertion loss. The FOPA parameters used in this chapter differ slightly from the previous model, with the full list provided in Table 6.1. A four-tone pump-phase modulation scheme ($N_t = 4$) with a base frequency $f_1 = 100$ MHz was used, resulting in a set of RF tones of [0.1, 0.3, 0.9, 2.7] GHz. Since the single-pump FOPA operates at the optimal wavelength, as discussed in the previous chapter, the phase distortion dominates over amplitude distortion at the FOPA output. In the FOPA link, this dithering-induced phase distortion is mixed with the laser phase noise caused by nonzero laser linewidth, making

the total phase distortion the primary contributor to the system degradation. We thus aim to develop a phase compensation scheme which can simultaneously mitigate both types of phase distortion.

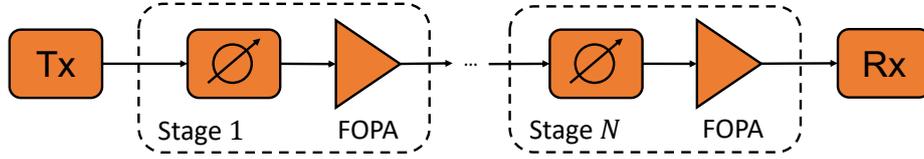


Figure 6.3: FOPA link with multiple cascaded stages.

Parameters	Values
λ_0	1562.9 [nm]
λ_p	1563.7 [nm]
$\beta^{(3)}$	1.2×10^{-41} [s^3m^{-1}]
$\beta^{(4)}$	-2.85×10^{-55} [s^4m^{-1}]
γ	10 [$\text{W}^{-1}\text{km}^{-1}$]
L	200 [m]
P	1.785 [W]

Table 6.1: Single-pump scheme parameters

In this section, we first focus on a single-stage phase-only compensation scheme designed to simultaneously mitigate both laser phase noise and dithering-induced phase distortion.

6.3.1 Compensation scheme and kernel selection

We consider the high-level diagram of the proposed compensation scheme, as shown in Fig. 6.4. At the symbol index n , the total phase distortion of the received symbol r_n can be written as $\phi_n = \delta\phi_n + \varphi_n$, where $\delta\phi_n$ and φ_n are the cumulative random laser phase noise modelled as a Wiener process and the dithering-induced phase distortion. The core of this compensation scheme is a sliding-window kernel recursive least square (SWKRLS)-based phase estimation block, which aims to predict the total phase distortion $\hat{\phi}_n$ as accurately as possible, approximating ϕ_n . This estimated phase distortion is then used to rotate the received symbol r_n to produce the corrected symbol $z_n = r_n e^{-j\hat{\phi}_n}$. A decision circuit is employed to return the decision-directed symbol, $[z_n]_D$ corresponding to z_n . The angle difference between the symbol before phase rotation r_n and the decision-directed symbol $[z_n]_D$, denoted as $\tilde{\phi}_n = \angle\{y_n\} - \angle\{[z_n]_D\}$, serves as a phase distortion teacher for training the SWKRLS-based phase estimation algorithm in free-running mode. During the training phase, the algorithm uses reference transmitted symbols to compute the phase distortion teacher until it reaches convergence.

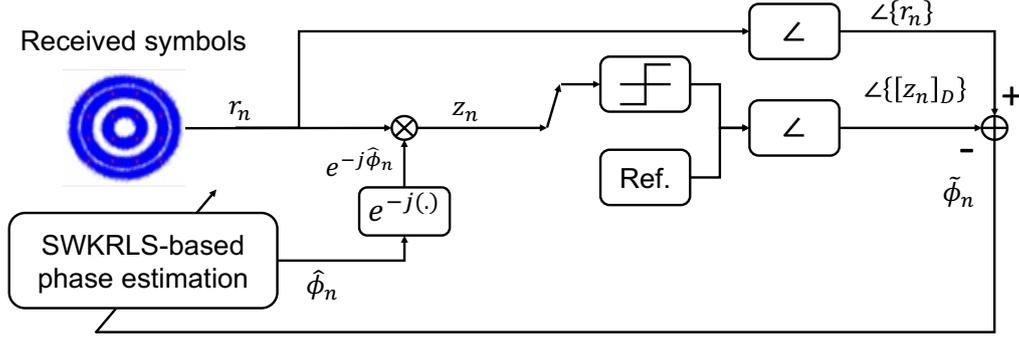


Figure 6.4: Block diagram of the proposed kernel-based phase recovery scheme.

The SWKRLS-based estimator applies kernel method to estimate the dithering-induced phase distortion, thereby selection of an appropriate kernel function is crucial for the estimation accuracy. As we learned from Chapter 3, this choice should be made based on the distinctive characteristics or underlying patterns of the phase distortion time series. Our prior understanding of the phase distortion reveals that the total phase distortion ϕ_n can be modelled as a combination of a periodic pattern, arising from the impact of pump-phase modulation, and a long-term variation trend, primarily due to laser phase noise. Based on the discussion in section 6.1, since these two pattern types contribute independently to the total phase distortion, adding a periodic kernel and a Gaussian kernel can allow us effectively capture these features.

The periodic pattern, which comes from the oscillations at the frequencies of the RF tones used within the FOPA's dithering scheme, can be captured by a kernel function which is summation of exponential sine squared kernels (defined in section 3.4.4) corresponding to each frequency, i.e.

$$k_{\text{Per}}(t_m, t_n) = \sum_{i=1}^{N_t} \exp\left(-\frac{2 \sin^2(\pi|t_m - t_n|/p_i)}{l_{\text{Per}}^2}\right), \quad (6.2)$$

where the data points are the time values t_n at symbol index n . The hyperparameters p_i ($i = 1, 2, \dots, N_t$) and l_{Per} represent the kernel's periods and lengthscale respectively. We assumed the number of dithering tones $N_t = 4$ for the considering scenario, corresponding to the four-tone pump-phase modulation. The periods p_i can be calculated from the RF modulating tone frequencies as $p_i = 1/f_i$, with the latter can be extracted at the receiver.

The long-term variation trend, which arises from the effect of laser phase noise, can be modelled by using a Gaussian kernel, which was defined in Eq. (3.35) of section 3.4.4. Finally, to capture both above characteristics of the phase distortion ϕ_n , we proposed a customised

kernel, which is a summation of the defined periodic kernel and the Gaussian kernel, i.e.

$$k_{\text{total}}(t_m, t_n) = k_{\text{Per}}(t_m, t_n) + k_{\text{RBF}}(t_m, t_n). \quad (6.3)$$

We applied the SWKRLS algorithm [94], which uses the sliding-window approach discussed in the previous section to constraint the dictionary size. This can help our algorithm to operate in an online setting. Our customised kernel $k_{\text{total}}(\cdot, \cdot)$ is selected as the kernel function for the SWKRLS algorithm.

We will quickly review the kernel learning strategy in this paragraph, with the detailed discussion can be found in section 3.4. The kernel method aims to minimise a least-squares cost function, i.e. $|\mathbf{y} - \hat{\mathbf{y}}|^2$, which is calculated between the true output \mathbf{y} and the predicted one $\hat{\mathbf{y}}$. By using a valid kernel function $k(\cdot, \cdot)$, the algorithm maps the input data within the dictionary, denoted as \mathbf{X} , into a higher-dimensional space as $\dot{\mathbf{X}}$. We used the notation \cdot here instead of $\phi(\cdot)$ as defined in section 3.4 to avoid confusion with the phase distortion ϕ_n . In that feature space, the predicted output $\hat{\mathbf{y}}$ can be linearly represented as $\hat{\mathbf{y}} = \dot{\mathbf{X}}\dot{\mathbf{w}}$. According to Representer Theorem, the weight $\dot{\mathbf{w}}$ can be represented with the basis vectors defined by rows of $\dot{\mathbf{X}}$, i.e. $\dot{\mathbf{w}} = \dot{\mathbf{X}}^T \boldsymbol{\alpha}$. Given the definition of kernel matrix $\mathbf{K} = \dot{\mathbf{X}}\dot{\mathbf{X}}^T$, the least-squares cost function can be rewritten as $|\mathbf{y} - \mathbf{K}\boldsymbol{\alpha}|^2$. The kernel method aims to solve for $\boldsymbol{\alpha}$ using the training set (\mathbf{X}, \mathbf{y}) , and then makes prediction for a new test point \mathbf{x}_* as $y_* = \sum_i \alpha_i k(\cdot, \mathbf{x}_*)$ using known data inputs within dictionary. In this problem, we selected the kernel matrix based on our customised kernel, i.e. $\mathbf{K}(i, j) = k_{\text{total}}(t_i, t_j)$.

In the SWKRLS algorithm, and in this problem, we are given the input-output pairs of time values and measured phase distortion $\{(t_1, \tilde{\phi}_1), (t_2, \tilde{\phi}_2), \dots\}$ that come sequentially as the transmission runs. A sliding window of size N is applied to limit the dictionary to only the last N pairs. Therefore, an observation input matrix $\mathbf{X}_n = [t_n, t_{n-1}, \dots, t_{n-N+1}]^T \in \mathbb{R}^N$ and an observation output vector $\mathbf{y}_n = [\tilde{\phi}_n, \tilde{\phi}_{n-1}, \dots, \tilde{\phi}_{n-N+1}]^T \in \mathbb{R}^N$ are formed at the symbol index n . A regularised kernel matrix at this time, $\mathbf{K}_n = \dot{\mathbf{X}}_n \dot{\mathbf{X}}_n^T + \lambda \mathbf{I}$, is calculated, where \mathbf{I} is the identity matrix. The matrix \mathbf{K} has previously been used without regularisation, but from now on, it will be considered as the regularised one for simplicity. The solution $\boldsymbol{\alpha}_n$ is obtained through the corresponding inverse kernel matrix \mathbf{K}_n^{-1} , i.e.

$$\boldsymbol{\alpha}_n = \mathbf{K}_n^{-1} \mathbf{y}_n. \quad (6.4)$$

However, the computation of \mathbf{K}_n and \mathbf{K}_n^{-1} are not done explicitly because of the high computational cost. The SWKRLS algorithm instead updates them recursively using the previous

one-step matrices \mathbf{K}_{n-1} and \mathbf{K}_{n-1}^{-1} , i.e. [94]

$$\mathbf{K}_n = \begin{bmatrix} \bar{\mathbf{K}}_{n-1} & \mathbf{b}_n \\ \mathbf{b}_n^T & c_n \end{bmatrix}, \quad (6.5)$$

where $\mathbf{b}_n = [k(t_{n-N+1}, t_n), \dots, k(t_{n-1}, t_n)]^T \in \mathbb{R}^{N-1}$ is the vector of kernels calculated between the last $N - 1$ inputs and the new input t_n , and $c_n = k(t_n, t_n) + \lambda$. The matrix $\bar{\mathbf{K}}_{n-1}$ is the created by removing from \mathbf{K}_{n-1} the first row and column. The inverse kernel matrix \mathbf{K}_n^{-1} is also updated accordingly as

$$\mathbf{K}_n^{-1} = \begin{bmatrix} \bar{\mathbf{K}}_{n-1}^{-1} - \bar{\mathbf{K}}_{n-1}^{-1} \mathbf{b}_n \mathbf{d}_n^T & \mathbf{d}_n \\ \mathbf{d}_n^T & e_n \end{bmatrix}, \quad (6.6)$$

where $e_n = (c_n - \mathbf{b}_n^T \bar{\mathbf{K}}_{n-1}^{-1} \mathbf{b}_n)^{-1}$ and $\mathbf{d}_n = -\bar{\mathbf{K}}_{n-1}^{-1} \mathbf{b}_n e_n$. After finding the solution α_n using Eq. (6.4), the SWKRLS algorithm produce the predicted phase for the next symbol at symbol index of $(n + 1)$ as [94]

$$\hat{\phi}_{n+1} = \sum_{i=1}^N \alpha_i \kappa(t_{n+1-i}, t_{n+1}), \quad (6.7)$$

by calculating kernel values between the points within the dictionary and the new point, $t_n + 1$. This predicted phase value will be used to rotate the symbol r_{n+1} and the cycle will be repeated.

6.3.2 Compensation results

To evaluate the effectiveness of the proposed algorithm, we performed numerical simulations of a single-polarisation 16-QAM transmission at 28-GBaud. The transmitted signal is pulse-shaped using a RRC filter with a roll-off factor of 0.1. The link parameters are the same as listed in Table 5.1, with the FOPA parameters are taken from Table 6.1. At each FOPA stage, we also introduced a random time shift to the sinusoidal RF tones within pump-phase modulation scheme. The selected evaluation metric is BER measured over 50 batches of 2^{16} symbols.

We started with our analysis by optimising the kernel hyperparameters, which are defined in Eqs. (3.35) and (6.2). Since the periods p_i within (6.2) can be extracted from the modulating RF frequencies, we are required to optimise two lengthscale hyperparameters l_{Per} and l_{RBF} of the periodic kernel and Gaussian kernel respectively. This can be achieved by maximising the log marginal likelihood as defined in section 3.4.5 of a Gaussian process [73]. It was taken on the training data set which is available in the pre-setting phase. We plotted the calculated normalised marginal likelihood as the heat map in Fig. 6.5, where larger values closer to one refers to models fitting the data with high confidence. The pre-setting phase lasts 100 symbols and the

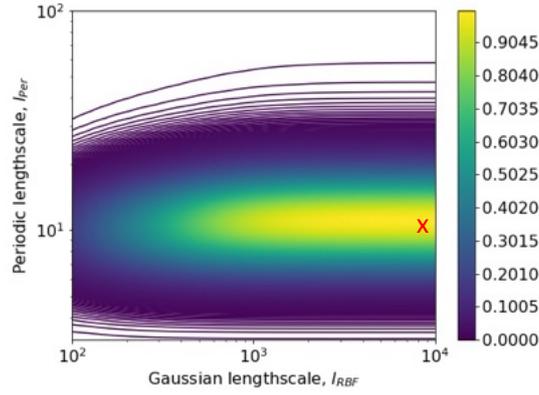


Figure 6.5: Hyperparameter optimisation with log marginal likelihood for kernel-based phase recovery.

regularisation parameter is selected as $\lambda = 0.1$ within this optimisation procedure. Examining Fig. 6.5 suggests us to select the optimum values for the periodic kernel's lengthscale $l_{\text{per}} = 10$ and the Gaussian kernel's lengthscale $l_{\text{RBF}} = 10^4$. To demonstrate our selection, we plotted in Fig. 6.6 the comparison between the fitted data and the training data over the pre-setting phase. We observed a good matching between these two time series, thereby reflecting the accuracy of hyper-parameters selection. The SWKRLS algorithm applied here with the dictionary size $N = 100$.

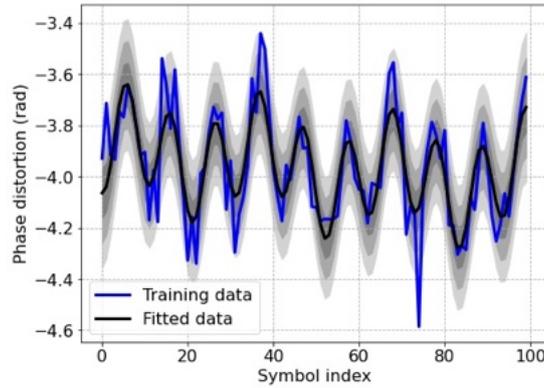


Figure 6.6: Data fitting within the training set.

The performance comparison between our proposed SWKRLS-based phase compensation scheme and the LMS-based phase recovery [65] was made. We plotted the BER curves of two schemes versus the number of FOFA stages in Fig. 6.7. The compensation capability of the conventional CPR method rapidly degrades as number of FOFA stages increases. It is an expected behaviour because the method is designed for slower laser phase noise, which is typical at kHz level. It thus cannot track and effectively the dithering-induced phase distur-

tion, which is much faster and accumulated across FOPA stages. Increasing the capability of tracking high-speed phase variation by increasing the learning rate of LMS-based algorithm may potentially make the recovery scheme more vulnerable to phase cycle slips. Our proposed SWKRLS-based scheme, which utilises a careful selection of a customised kernel, can significantly outperforms the LMS-based phase recovery scheme within one stage of compensation. We observed at least one order of magnitude BER improvements across various number of FOPA stages.

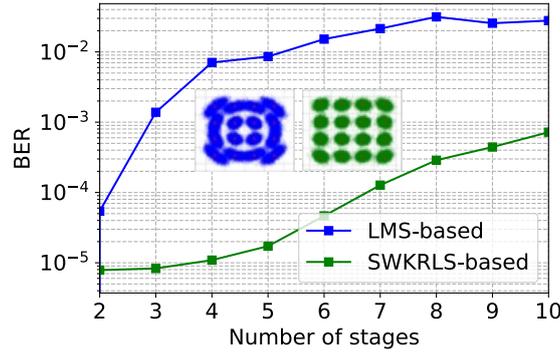


Figure 6.7: Performance result comparison for kernel-based phase compensation versus conventional phase recovery.

This result is the first step with kernel-based methods although it only offers the phase distortion compensation capability. In the next section, we will advance the method to make it operate in the complex domain. By achieving that, we can enable the simultaneous dithering-induced phase and amplitude distortion compensation capability.

6.4 Kernel-based phase and amplitude compensation using complex kernel

We aim to design a phase and amplitude distortion compensation scheme based on kernel methods in this section. Since only dithering-induced phase distortion prevails when we operate FOPA at the optimal wavelength, we are required to shift the operating wavelength to a different one where the dithering-induced amplitude distortion is more pronouncing. By using the FOPA model with parameters listed in Table 6.1, we plotted in Fig. 6.8 its amplitude and phase spectral responses along with the corresponding RMS fluctuations in phase and amplitude. We selected the operating region with the wavelength detuning $\Delta\lambda_s = -20$ nm, where the dithering-induced

amplitude distortion reaches its peak.

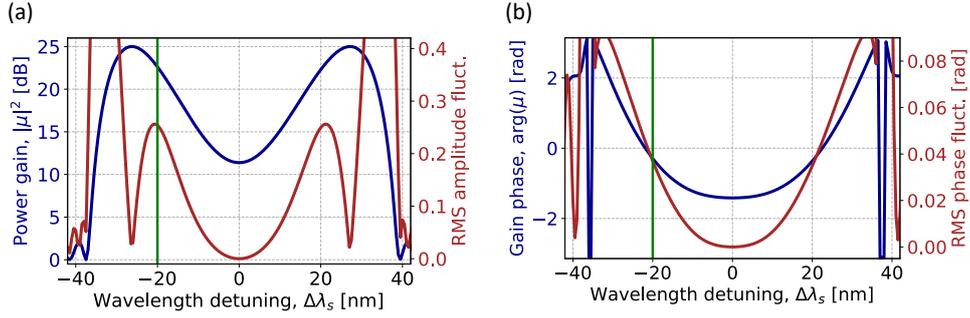


Figure 6.8: (a) The amplitude and (b) phase spectral responses of the FOPA complex gain and the corresponding dithering-induced RMS amplitude and phase fluctuations versus the signal wavelength detuning.

The FOPA link considered in this scenario is the same as that in the previous section, which is plotted in Fig. 6.3. The signal at FOPA output in this scenarios exhibits both phase and amplitude distortions, which are also accumulated along the link.

6.4.1 Compensation scheme

A straightforward approach to apply kernel methods in the complex domain is to use the complex versions of widely-known real kernels. For example, the complex version of the real Gaussian kernel, which is defined in Eq. (3.35), is given as

$$\kappa(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^L \frac{(x'_i - x_i^*)^2}{2\sigma^2}\right), \quad (6.8)$$

where the data points $\mathbf{x}, \mathbf{x}' \in \mathbb{C}^L$ and the superscript $*$ denotes the complex conjugation. We will use this kernel function for the complex-valued (CV)-SWKRLS algorithm, which is a part of the compensation diagram shown in Fig. 6.9.

Instead of predicting the phase distortion, our proposed CV-SWKRLS-based distortion estimator aims to produce a CV distortion $\hat{d}_n \in \mathbb{C}$ at the symbol index n . The total complex distortion is technically removed from the received symbol r_n by dividing it by the estimated distortion \hat{d}_n to obtain $z_n = r_n / \hat{d}_n$. The rotated symbol z_n is then passed through a decision circuit which produces a decision-directed symbol $[z_n]_D$. A decision-driven distortion is then calculated by taking the ratio between the received symbol r_n and the decision-directed symbol $[z_n]_D$, i.e. $\tilde{d}_n = r_n / [z_n]_D$. This decision-driven distortion is then fed back to the CV-SWKRLS-based estimation block to train the algorithm and also to become one of the inputs

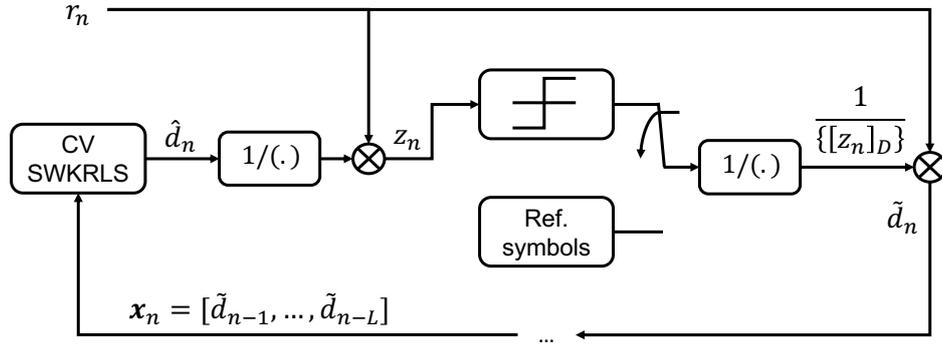


Figure 6.9: Block diagram of the proposed complex-valued kernel-based recovery scheme.

for future distortion's prediction. The scheme uses the exact transmitted symbols in the role of decision-directed symbols to calculate the decision-driven distortion before the convergence is reached.

The problem formulation for CV-SWKRLS in the online setting can be described as following: it attempts to make one-step ahead prediction of the complex distortion \hat{d}_n based on a vector of last L decision-driven values, i.e. $\mathbf{x}_n = [\tilde{d}_{n-1}, \dots, \tilde{d}_{n-L}]^T$. In the sliding-window approach, the dictionary size of N is fixed by only accounting for the last N input-output pairs. The input observation matrix is defined as $\mathbf{X}_n = [\mathbf{x}_n, \dots, \mathbf{x}_{n-N+1}]^T$ and the output observation vector as $\mathbf{y}_n = [\tilde{d}_n, \dots, \tilde{d}_{n-N+1}]^T$. The kernel function is used here is the complex Gaussian kernel in Eq. 6.8. The procedure to solve for the coefficient vector α_n is defined in the previous section and in [94]. After obtaining the updated solution α_n at the end of n -th cycle, the CV-SWKRLS-based estimator can make prediction of the total distortion \hat{d}_{n+1} at the beginning of $(n + 1)$ -th cycle using the expression

$$\hat{d}_{n+1} = \sum_{i=1}^N \alpha_i \kappa(\mathbf{x}_{n+1-i}, \mathbf{x}_{n+1}). \quad (6.9)$$

The calculation is made using the kernel values measured between the data points \mathbf{x}_i within the dictionary and the new input vector \mathbf{x} . The estimated distortion \hat{d}_{n+1} is then used to recover the distorted symbol r_{n+1} as the cycle continues.

6.4.2 Compensation results

The complex Gaussian kernel function in Eq. 6.8 has only one hyperparameter, which is the lengthscale σ . The CV-SWKRLS algorithm's hyperparameters include the block length L , the dictionary size N and the regularisation parameter λ . For the visualisation purpose, we can only perform the optimisation process for maximum three hyperparameters. We realised that the regularisation parameter λ can be set fixed, and let the kernel width σ govern the overfitting problem. The optimisation was performed for 20 batches of 2^{16} symbols received at the end of 10 FOPA stages, and we used BER as the evaluation metric. The optimisation results taken for block length L , dictionary size N and kernel width σ are shown in Fig. 6.10. Our decision which omits the regularisation parameter λ can also be support by this graph. Based on definition of Gaussian kernel function in Eq. (6.8), a smaller value of kernel width σ makes the kernel function narrower, and the algorithm only considers the nearby data points as being similar. It results in the algorithm more sensitive to noise and overfit, while a greater value of σ can help reduce it. This is the reason why the BER level starts becoming meaningful as the kernel width σ exceeds a certain threshold. For the optimal point, which yields the lower BER, we observed it at $N = 50$, $L = 50$, $\sigma = 10^{1.2}$, which is marked by the blue circle. This set of hyperparameters will be used for further performance evaluation.

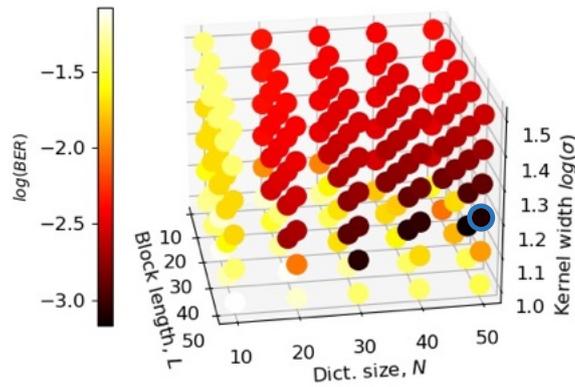


Figure 6.10: Hyperparameter optimisation for the kernel method using complex kernel function.

The performance comparison, in terms of BER, between our proposed algorithm and the conventional CPR algorithm, LMS-based phase recovery [65], was performed for 100×2^{16}

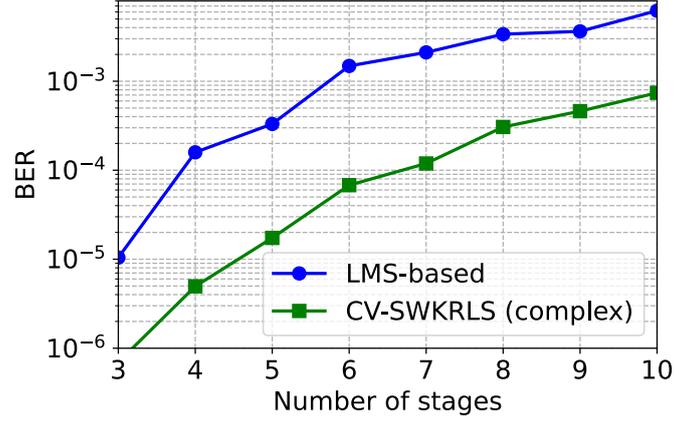


Figure 6.11: Performance comparison between complex-valued kernel-based compensation using complex kernel function and the conventional phase recovery algorithm.

symbols. Two BER curves are shown in Fig. 6.11, with a substantial BER improvement of our proposed algorithm over the conventional CPR algorithm. The same reason as that in the previous section can be used to explain this difference. The conventional CPR algorithm is designed for the relatively slow laser phase noise and does not perform well to the high-frequency dithering-induced phase distortion. We also included the constellation diagrams between two schemes after 10 stages to show the effectiveness of our proposed scheme.

6.5 Kernel-based phase and amplitude compensation using complexification of RKHS

In this section, we apply a technique allowing us to exploit the popular real kernel functions instead of using the complex versions. We will then make a comparison between two techniques at the end of the section.

6.5.1 Learning with complexification of Hilbert space

In this section, we considered another way to operate kernel methods in the complex domain, which is the concept of "complexification" of real RKHS [96]. In this case, the complex input data $\mathbf{x} \in \mathbb{C}^L$ is represented as $\mathbf{x} = \mathbf{x}_r + j\mathbf{x}_i$, where the real and imaginary parts $\mathbf{x}_r, \mathbf{x}_i \in \mathbb{R}^L$ respectively are mapped to a "complexified" RKHS based on the rule

$$\hat{\Phi}(\mathbf{x}) = \Phi(\mathbf{x}_r, \mathbf{x}_i) + j\Phi(\mathbf{x}_r, \mathbf{x}_i) \quad (6.10)$$

where $\Phi(\mathbf{x}_r, \mathbf{x}_i) = \kappa(\cdot, (\mathbf{x}_r, \mathbf{x}_i))$ is the feature map with a chosen real kernel. It thus allows us to use a real version of Gaussian kernel, defined in Eq. (3.35), for complex-valued data processing.

The compensation scheme with CV-SWKRLS-based distortion estimation block is the same as in Fig. 6.9. The CV-SWKRLS-based algorithm also aims to predict the one-step ahead complex distortion \hat{d}_n based last L decision-driven distortions, defined in the vector $\mathbf{x}_n = [\tilde{d}_{n-1}, \dots, \tilde{d}_{n-L}]$. However, in this case, instead of using directly the vector \mathbf{x}_n , we converted it into a real-valued composite representation $\hat{\mathbf{x}}_n \in \mathbb{R}^{2L}$ defined as

$$\hat{\mathbf{x}}_n = [\Re(\mathbf{x}_n), \Im(\mathbf{x}_n)]. \quad (6.11)$$

In the online setting of the CV-SWKRLS algorithm, we defined the observation input matrix from the composite versions of \mathbf{x} as $\hat{\mathbf{X}}_n = [\hat{\mathbf{x}}_n, \dots, \hat{\mathbf{x}}_{n-N+1}]^T$, and the observation output vector $\mathbf{y}_n = [\tilde{d}_n, \dots, \tilde{d}_{n-N+1}]^T$. The procedure to solve for the coefficient vector α_n is defined in section 6.3. The prediction for the distortion at $(n+1)$ -th cycle is also performed based on the composite representations as

$$\hat{d}_{n+1} = \sum_{i=1}^N \alpha_i \kappa(\hat{\mathbf{x}}_{n+1-i}, \hat{\mathbf{x}}_{n+1}). \quad (6.12)$$

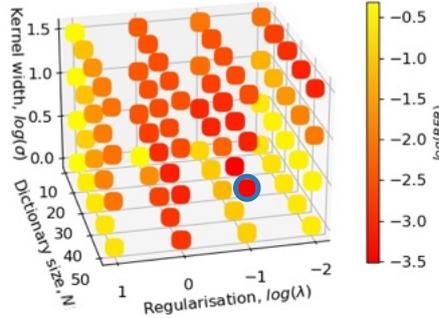


Figure 6.12: Hyperparameter optimisation for kernel-based method using complexification of RKHS.

6.5.2 Compensation results

Following the similar optimisation procedure, we carried a grid search over a space with three hyperparameters: dictionary size N , kernel width σ and regularisation parameter λ . We set the block length as $L = 20$. The optimisation was taken for 50 batches of 2^{16} symbols received

after 10 FOPA stages with BER being the evaluation metric. The optimisation result was plotted Fig. 6.12, where we observed the optimal operating point, which is denoted by the blue circle, at this set of hyper-parameters: $N = 50$, $\sigma = 10^{0.5}$, $\lambda = 10^{-1}$.

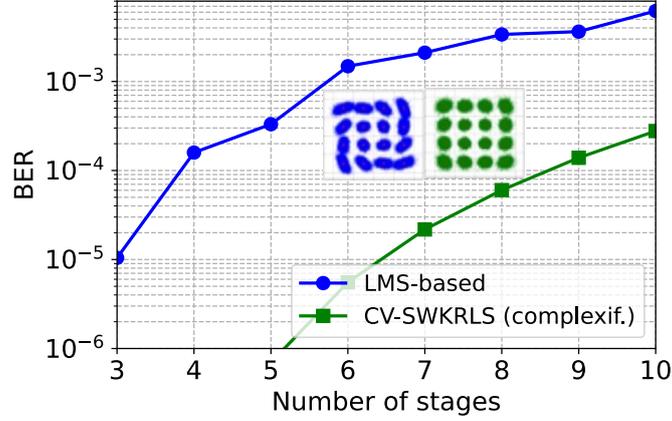


Figure 6.13: Performance comparison between complex-valued kernel-based compensation using complexification of RKHS and the conventional phase recovery algorithm.

We demonstrated the performance of our proposed CV-SWKRLS-based compensation scheme with the RKHS complexification concept in Fig. 6.13. It is compared against the conventional CPR scheme with LMS-based algorithm [65]. Our CV-SWKRLS-based scheme, as expected, significantly outperforms the conventional CPR method by achieving at least one order of magnitude of BER improvement across various number of spans. The capability of simultaneous phase and amplitude distortion compensation of our proposed algorithm can be evidenced in Fig. 6.14. The figure demonstrates the evolution of RMS fluctuations for both phase and amplitude as a function of the number of cascaded FOPA stages. The calculations of RMS for amplitude and phase can be found in Eqs. (5.19) and (5.20), respectively. In both quantities, our proposed algorithm shows improved suppression capability compared to the conventional CPR scheme.

We have discussed two methods to operate the kernel method in the complex domain: either using a directly complex kernel function or through the complexification of RKHS. We observe a substantial BER improvement with the second method over the first, as illustrated in Fig. 6.15. This can be explained by the fact that complexifying a real RKHS is more effective when the real and imaginary parts of the data are independent and exhibit same properties in terms of similarity [97], as is the case for our considered FOPA link. Although the received signal exhibits both amplitude and phase distortions, phase distortion remains dominant, as

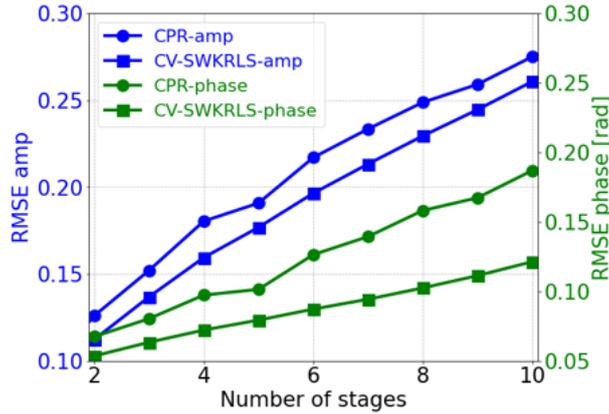


Figure 6.14: Amplitude and phase suppression capability of kernel-based compensation using complexification of RKHS compared to the conventional phase recovery algorithm.

evident from the constellation diagrams of the conventional CPR algorithm. By treating the real and imaginary components separately, the complexification of real RKHS enables the model to better capture their independent characteristics. This allows for more effective suppression of phase distortion compared to amplitude distortion, resulting in a more circular constellation diagram. For the first method, the use of pure complex kernel function has notable drawbacks, as discussed in [98]. In detail, the complex Gaussian kernel introduces challenges such as non-stationary and inconsistencies in similarity measurement, the real parts are measured in terms of similarity and the imaginary parts are measured in terms of dissimilarity. These issues highlight the limitations of purely complex kernels in our scenario.

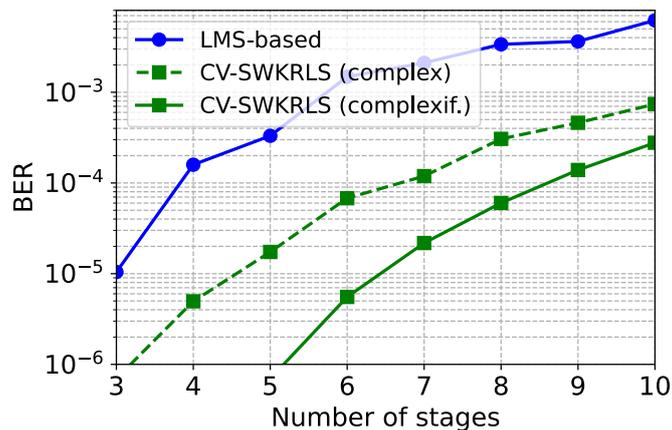


Figure 6.15: Performance comparison between two complex-valued kernel-based compensation schemes and the conventional phase recovery algorithm.

6.6 Conclusion

In this chapter, we have presented algorithms that use a kernel-based approach to simultaneously mitigate phase and amplitude distortions caused by pump-phase modulation in the transmission link with multiple cascaded FOPA stages. A single-stage compensation scheme using kernel methods was proposed for phase distortion mitigation, which outperformed the conventional phase recovery algorithm. We also discussed the operation of kernel methods in the complex domain, enabling the simultaneous compensation of phase and amplitude distortion. We also compared two approaches related to complex-valued processing, with the method using the complexification of RKHS providing better improvement than the one using pure complex kernels. While the results presented in this work focus on a single wavelength, we believe that the proposed method is scalable and can be applied to WDM systems with multiple wavelengths. The compensation approach is wavelength-specific, but it can be extended across a wide range of wavelengths by adjusting the parameters for each channel accordingly. In a WDM system, the compensation process can be executed independently and in parallel for each wavelength channel, making it feasible to implement the method in systems with many channels.

Chapter 7

Conclusions and future work

The thesis has introduced several DSP algorithms designed to mitigate the adverse effects of pump-phase modulation while preserving the efficacy of SBS suppression. This work lays the groundwork for the integration of fibre-optical parametric devices into modern optical communications. With effective SBS suppression supported by DSP algorithms, the FOPA can offer exceptionally high gain with low noise. Moreover, its wide gain bandwidth makes it a promising solution for future dense wavelength-division multiplexing (DWDM) systems. The proposed DSP algorithm enable the OPC to remain a highly effective all-optical compensation technique for fibre impairments, including chromatic dispersion and fibre nonlinearity. We will now review key findings we have discussed within this thesis before exploring potential future directions.

7.1 Key findings

Section 2.5 in Chapter 2 introduced an optimisation approach for the amplitude and phase parameters of the RF tones within pump-phase modulation schemes. The goal was to achieve a broadened pump spectrum with evenly distributed frequency peaks, which is important to effectively suppress the SBS effect. This desired spectrum was set as the target for our learning process. Using a least-squares cost function, regression on the RF tones was implemented in Tensorflow with batch gradient descent.

Chapter 4 studied the mid-link OPC transmission systems and examined how both phase and amplitude distortions develop, driven by the residual dithering-induced phase distortion and the memory effect of channel fibre dispersion. We proposed a linear regression approach to estimate this phase distortion based on a feature vector constructed from sinusoidal terms

with detected modulating frequencies. This enabled us to recreate the phase-to-phase and phase-to-amplitude distortion conversion in the time domain and reverse it via a two-stage compensation scheme.

In Chapter 5, we advanced the DDC scheme by addressing multiple sources of dithering phase distortion, particularly in transmission systems with cascaded FOPA stages. The accumulation of phase distortion across stages, along with interactions between accumulated phase distortion and fibre dispersion across different segments, significantly increases the problem's complexity. We began with a multi-branch approach that estimated the dithering phase distortion in each FOPA stage, compensating for corresponding phase and amplitude distortions. We then improved the complexity of the algorithm by developing a single-branch compensation approach. In this approach, an adaptive time-varying FIR filter was employed to compensate for both amplitude and phase distortions, as evidenced by a reduction in their respective RMS values. An intensive study was conducted to demonstrate the robustness of the algorithm, including scenarios involving random modulating frequency variations and increases in the base frequency increase.

Chapter 6 introduced a novel approach using kernel methods, a non-parametric machine learning technique, to mitigate the compensation problem. Kernel methods, known for their ability to automatically discover underlying patterns in data, were used to enhance our compensation algorithm without the need for explicit engineering of feature vectors. Operating in an online setting, we can effectively simplify the compensation scheme into a single stage, which also accounts for the laser phase noise. The first scheme for phase compensation was proposed by selecting a custom kernel function that captured the mathematical characteristics of the phase distortion time series. Two additional algorithms were developed to process signals in the complex domain, which is essential for simultaneous mitigation of phase and amplitude distortions. Applying kernel methods in the complex domain is an emerging research area, as they have traditionally been used for real-valued data. Two approaches were introduced: using pure complex kernels and the complexification of real RKHS, both of which demonstrated the effectiveness in BER improvements.

7.2 Future work

In addition to FOPA, phase-sensitive amplifier (PSA) is a powerful amplification technique and a crucial tool for advanced communication systems. PSA is known for its ability to enhance sensitivity by amplifying the signal while suppressing noise [99]. This property makes it valuable in applications that require high signal fidelity. The integration of DSP algorithms is important to preserve the benefits of PSA in noiseless operation. Research into DSP algorithms for PSA systems is thus promising to unlock new possibilities for PSA applications.

Regarding kernel methods, widely linear kernels [100] are interesting extensions of linearity in statistical modelling by incorporating both linear and nonlinear components. This allows them to effectively learn complex data structures and support both real-valued and complex-valued data. Additionally, the generalised kernel learning framework for complex-valued data, introduced in [97], offers promising adaptability to our problem, providing greater flexibility depending on the distortion characteristics. These approaches present exciting opportunities to further enhance the kernel-based phase and amplitude compensation scheme.

We also aim to extend the scope of the kernel-based dithering compensation method further by exploring its potential in mitigation of other channel impairments, such as fibre nonlinearity and chromatic dispersion. Kernel methods are universal approximators according to Representer Theorem, and with a rich selection of kernel functions, they offer the possibility to address impairments like channel memory and nonlinearity.

In conclusion, this thesis has shown promising results for the integration of fibre-optical parametric devices into modern optical networks. The proposed DSP algorithms significantly reduce penalties due to pump-phase modulation, easing the need for complex sub-system designs to counter SBS effect. This research opens up exciting future directions, both in optical network applications and in the further improvement of the algorithms.

List of References

- [1] L. H. Nguyen, S. Boscolo, A. D. Ellis, and S. Sygletos, "Machine learning enabled compensation of phase-to-amplitude distortion in opc systems," in *2023 Conference on Lasers and Electro-Optics (CLEO)*, pp. 1–2, 2023.
- [2] L. H. Nguyen, T. T. Nguyen, A. D. Ellis, S. Sygletos, and S. Boscolo, "Digital signal processing for optical phase conjugation assisted coherent systems," in *2023 23rd International Conference on Transparent Optical Networks (ICTON)*, pp. 1–4, 2023.
- [3] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Machine learning enabled digital compensation of phase-to-amplitude distortion in fibre-optical parametric amplifier based transmission links," *EPJ Web Conf.*, vol. 287, p. 13022, 2023.
- [4] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Design of optical transmission link with cascaded fibre-optical parametric amplifiers," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, pp. 1–3, 2024.
- [5] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Digital signal processing for fibre-optical parametrically amplified transmission links," in *2024 24th International Conference on Transparent Optical Networks (ICTON)*, pp. 1–4, 2024.
- [6] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Online kernel-based phase recovery for parametrically amplified optical transmission," in *2024 14th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pp. 95–98, 2024.
- [7] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Online phase and amplitude distortion compensation in fopa transmission systems," in *2024 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, pp. 1–2, 2024.
- [8] L. H. Nguyen, S. Boscolo, and S. Sygletos, "Complex-valued kernel-based phase and amplitude distortion compensation in parametrically amplified optical links," in *2024 50th European Conference on Optical Communication (ECOC)*, 2024.
- [9] L. H. Nguyen, S. Boscolo, and S. Sygletos., "Complex-valued kernels for compensation of pump dithering induced distortions in fopa links," in *2024 IEEE Photonics Conference (IPC)*, pp. 1–2, 2024.
- [10] L. H. Nguyen, S. Boscolo, and S. Sygletos., "Online digital compensation of pump dithering induced phase and amplitude distortions in transmission links with cascaded fibre-optical parametric amplifiers," *Opt. Express*, vol. 32, pp. 13467–13477, Apr 2024.

- [11] M. Bastamova, V. Gordienko, S. Sygletos, M. Tan, A. Donodin, L. H. Nguyen, F. Bessin, S. Boscolo, N. Doran, and A. D. Ellis, "Performance evaluation of a polarisation insensitive Mach-Zehnder fiber parametric amplifier with 38 channel transmission," in *2024 50th European Conference on Optical Communication (ECOC)*, 2024.
- [12] M. E. Marhic, *Fiber Optical Parametric Amplifiers, Oscillators and Related Devices*. Cambridge University Press, 2nd ed., 2007.
- [13] G. P. Agrawal, *Nonlinear Fiber Optics*. Elsevier Science, 4th ed., 2006.
- [14] M. Hirano, T. Nakanishi, T. Okuno, and M. Onishi, "Silica-based highly nonlinear fibers and their application," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 15, no. 1, pp. 103–113, 2009.
- [15] M. E. Marhic, P. A. Andrekson, P. Petropoulos, S. Radic, C. Peucheret, and M. Zazayerifar, "Fiber optical parametric amplifiers in optical communication systems," *Laser & Photonics Reviews*, vol. 9, pp. 50–74, January 2015. Review article.
- [16] M. E. Marhic, N. Kagi, T.-K. Chiang, and L. G. Kazovsky, "Broadband fiber optical parametric amplifiers," *Opt. Lett.*, vol. 21, no. 8, pp. 573–575, 1996.
- [17] T. Torounidis, P. A. Andrekson, and B. E. Olsson, "Fiber-optical parametric amplifier with 70-dB gain," *IEEE Photonics Technology Letters*, vol. 18, no. 10, pp. 1194–1196, 2006.
- [18] M. Vasilyev, "Distributed phase-sensitive amplification," *Opt. Express*, vol. 13, no. 19, pp. 7563–7571, 2005.
- [19] A. Yariv, "Phase conjugate optics and real-time holography," *IEEE Journal of Quantum Electronics*, vol. 14, no. 9, pp. 650–660, 1978.
- [20] A. Yariv, D. Fekete, and D. M. Pepper, "Compensation for channel dispersion by nonlinear optical phase conjugation," *Opt. Lett.*, vol. 4, pp. 52–54, Feb 1979.
- [21] S. Watanabe, T. Naito, and T. Chikama, "Compensation of chromatic dispersion in a single-mode fiber by optical phase conjugation," *IEEE Photonics Technology Letters*, vol. 5, no. 1, pp. 92–95, 1993.
- [22] A. D. Ellis, M. Tan, M. A. Iqbal, M. A. Z. Al-Khateeb, V. Gordienko, G. S. Mondaca, S. Fabbrì, M. F. C. Stephens, M. E. McCarthy, A. Perentos, I. D. Phillips, D. Lavery, G. Liga, R. Maher, P. Harper, N. Doran, S. K. Turitsyn, S. Sygletos, and P. Bayvel, "4 Tb/s transmission reach enhancement using 10×400 Gb/s super-channels and polarization insensitive dual band optical phase conjugation," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1717–1723, 2016.
- [23] 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 *2016 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2016.
- [24] S. Yoshima, Y. Sun, Z. Liu, K. R. H. Bottrill, F. Parmigiani, D. J. Richardson, and P. Petropoulos, "Mitigation of nonlinear effects on WDM QAM signals enabled by optical phase conjugation with efficient bandwidth utilization," *J. Lightwave Technol.*, vol. 35, no. 4, pp. 971–978, 2017.

- [25] J. Wang, Y. Du, C. Liang, Z. Li, and J. Fang, "Performance evaluation of highly nonlinear fiber (HNLF) based optical phase conjugation (OPC) in long haul transmission of 640 Gbps 16-QAM CO-OFDM," *Photonics*, vol. 8, no. 2, 2021.
- [26] E. Ip, "Nonlinear compensation using backpropagation for polarization-multiplexed transmission," *Journal of Lightwave Technology*, vol. 28, no. 6, pp. 939–951, 2010.
- [27] D. Rafique, J. Zhao, and A. D. Ellis, "Digital back-propagation for spectrally efficient WDM 112 Gbit/s PM M-ary QAM transmission," *Opt. Express*, vol. 19, no. 6, pp. 5219–5224, 2011.
- [28] A. Napoli, Z. Maalej, V. A. J. M. Sleiffer, M. Kuschnerov, D. Rafique, E. Timmers, B. Spinnler, T. Rahman, L. D. Coelho, and N. Hanik, "Reduced complexity digital back-propagation methods for optical communication systems," *Journal of Lightwave Technology*, vol. 32, no. 7, pp. 1351–1362, 2014.
- [29] A. D. Ellis, M. E. McCarthy, M. A. Z. A. Khateeb, M. Sorokina, and N. J. Doran, "Performance limits in optical communications due to fiber nonlinearity," *Adv. Opt. Photon.*, vol. 9, no. 3, pp. 429–503, 2017.
- [30] K. Inoue, "Four-wave mixing in an optical fiber in the zero-dispersion wavelength region," *Journal of Lightwave Technology*, vol. 10, no. 11, pp. 1553–1561, 1992.
- [31] R. Jopson, A. Gnauck, and R. Derosier, "Compensation of fibre chromatic dispersion by spectral inversion," *Electronics Letters*, vol. 29, pp. 576–578(2), April 1993.
- [32] G. S. He, "Optical phase conjugation: principles, techniques, and applications," *Progress in Quantum Electronics*, vol. 26, no. 3, pp. 131–191, 2002.
- [33] R. W. Boyd, *Nonlinear Optics*. Elsevier Science, 2020.
- [34] V. Gordienko, A. D. Szabo, M. F. C. Stephens, V. Vassiliev, C. B. Gaur, and N. J. Doran, "Limits of broadband fiber optic parametric devices due to stimulated Brillouin scattering," *Optical Fiber Technology*, vol. 66, p. 102646, 2021.
- [35] T. Kurashima, T. Horiguchi, and M. Tateda, "Thermal effects of Brillouin gain spectra in single-mode fibers," *IEEE Photonics Technology Letters*, vol. 2, no. 10, pp. 718–720, 1990.
- [36] Y. Imai and N. Shimada, "Dependence of stimulated Brillouin scattering on temperature distribution in polarization-maintaining fibers," *IEEE Photonics Technology Letters*, vol. 5, no. 11, pp. 1335–1337, 1993.
- [37] J. Hansryd, F. Dross, M. Westlund, P. A. Andrekson, and S. N. Knudsen, "Increase of the SBS threshold in a short highly nonlinear fiber by applying a temperature distribution," *Journal of Lightwave Technology*, vol. 19, no. 11, pp. 1691–1697, 2001.
- [38] N. Yoshizawa and T. Imai, "Stimulated Brillouin scattering suppression by means of applying strain distribution to fiber with cabling," *Journal of Lightwave Technology*, vol. 11, no. 10, pp. 1518–1522, 1993.
- [39] R. Engelbrecht, "Analysis of SBS gain shaping and threshold increase by arbitrary strain distributions," *Journal of Lightwave Technology*, vol. 32, no. 9, pp. 1689–1700, 2014.

- [40] K. Shiraki, M. Ohashi, and M. Tateda, "Suppression of stimulated Brillouin scattering in a fibre by changing the core radius," *Electronics Letters*, vol. 31, pp. 668–669, 1995.
- [41] M. Ohashi and M. Tateda, "Design of strain-free-fiber with nonuniform dopant concentration for stimulated Brillouin scattering suppression," *Journal of Lightwave Technology*, vol. 11, no. 12, pp. 1941–1945, 1993.
- [42] K. Shiraki, M. Ohashi, and M. Tateda, "Performance of strain-free stimulated Brillouin scattering suppression fiber," *Journal of Lightwave Technology*, vol. 14, no. 4, pp. 549–554, 1996.
- [43] D. Cotter, "Suppression of stimulated Brillouin scattering during transmission of high-power narrowband laser light in monomode fibre," *Electronics Letters*, vol. 18, no. 15, pp. 638–640, 1982.
- [44] B. Anderson, C. Robin, A. Flores, and I. Dajani, *Experimental study of SBS suppression via white noise phase modulation*, vol. 8961 of *SPIE LASE*. SPIE, 2014.
- [45] A. Flores, C. Robin, A. Lanari, and I. Dajani, "Pseudo-random binary sequence phase modulation for narrow linewidth, kilowatt, monolithic fiber amplifiers," *Optics Express*, vol. 22, no. 15, pp. 17735–17744, 2014.
- [46] J. B. Coles, B. P. P. Kuo, N. Alic, S. Moro, C. S. Bres, J. M. C. Boggio, P. A. Andrekson, M. Karlsson, and S. Radic, "Bandwidth-efficient phase modulation techniques for stimulated Brillouin scattering suppression in fiber optic parametric amplifiers," *Optics Express*, vol. 18, no. 17, pp. 18138–18150, 2010.
- [47] Y. Liu, Z. Lv, Y. Dong, and Q. Li, "Research on stimulated Brillouin scattering suppression based on multi-frequency phase modulation," *Chinese Optics Letters*, vol. 7, no. 1, pp. 29–31, 2009.
- [48] J. Mauro, S. Raghavan, and A. Ruffin, "Enhanced stimulated Brillouin scattering threshold through phase control of multitone phase modulation," *Optical Engineering*, vol. 49, no. 10, p. 100501, 2010.
- [49] Y. Dong, Z. Lu, Q. Li, and Y. Liu, "Broadband Brillouin slow light based on multifrequency phase modulation in optical fibers," *Journal of the Optical Society of America B*, vol. 25, no. 12, pp. C109–C115, 2008.
- [50] S. K. Korotky, "Multifrequency lightwave source using phase modulation for suppressing stimulated Brillouin scattering in optical fibers," U.S. Patent 5,566,381, Oct. 15, 1996.
- [51] T. Okoshi and K. Kikuchi, *Coherent Optical Fiber Communications*. Springer Dordrecht, 1988.
- [52] S. J. Savory, "Digital filters for coherent optical receivers," *Optics Express*, vol. 16, no. 2, pp. 804–817, 2008.
- [53] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *Journal of Lightwave Technology*, vol. 34, no. 1, pp. 157–179, 2016.
- [54] K. Kikuchi, "History of coherent optical communication and challenges for the future," in *2008 Digest of the IEEE/LEOS Summer Topical Meetings*, pp. 107–108, 2008.

- [55] M. S. Faruk and S. J. Savory, "Digital signal processing for coherent transceivers employing multilevel formats," *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1125–1141, 2017.
- [56] S. J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164–1179, 2010.
- [57] R. Kudo, T. Kobayashi, K. Ishihara, Y. Takatori, A. Sano, and Y. Miyamoto, "Coherent optical single carrier transmission using overlap frequency domain equalization for long-haul optical systems," *J. Lightwave Technol.*, vol. 27, pp. 3721–3728, Aug 2009.
- [58] J. C. Geyer, C. R. S. Fludger, T. Duthel, C. Schulien, and B. Schmauss, "Efficient frequency domain chromatic dispersion compensation in a coherent polmux QPSK-receiver," in *2010 Conference on Optical Fiber Communication (OFC/NFOEC), collocated National Fiber Optic Engineers Conference*, pp. 1–3, 2010.
- [59] T. Xu, G. Jacobsen, S. Popov, M. Forzati, J. Mårtensson, M. Mussolin, J. Li, K. Wang, Y. Zhang, and A. T. Friberg, "Frequency-domain chromatic dispersion equalization using overlap-add methods in coherent optical system," *Journal of Optical Communications*, vol. 32, no. 2, pp. 131–135, 2011.
- [60] T. Xu, G. Jacobsen, S. Popov, J. Li, E. Vanin, K. Wang, A. T. Friberg, and Y. Zhang, "Chromatic dispersion compensation in coherent transmission system using digital filters," *Opt. Express*, vol. 18, pp. 16243–16257, Jul 2010.
- [61] A. Eghbali, H. Johansson, O. Gustafsson, and S. J. Savory, "Optimal least-squares FIR digital filters for compensation of chromatic dispersion in digital coherent optical receivers," *Journal of Lightwave Technology*, vol. 32, no. 8, pp. 1449–1456, 2014.
- [62] A. J. Viterbi and A. M. Viterbi, "Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission," *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 543–551, 1983.
- [63] E. Ip and J. M. Kahn, "Feedforward carrier recovery for coherent optical communications," *Journal of Lightwave Technology*, vol. 25, no. 9, pp. 2675–2692, 2007.
- [64] Y. Mori, C. Zhang, K. Igarashi, K. Katoh, and K. Kikuchi, "Unrepeated 200-km transmission of 40-Gbit/s 16-QAM signals using digital coherent receiver," *Optics Express*, vol. 17, no. 3, pp. 1435–1441, 2009.
- [65] I. Fatadin, D. Ives, and S. J. Savory, "Blind equalization and carrier phase recovery in a 16-QAM optical coherent system," *Journal of Lightwave Technology*, vol. 27, no. 15, pp. 3042–3049, 2009.
- [66] T. Pfau, S. Hoffmann, and R. Noe, "Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for m-QAM constellations," *Journal of Lightwave Technology*, vol. 27, no. 8, pp. 989–999, 2009.
- [67] S. S. Haykin, *Adaptive Filter Theory*. Pearson Education, 2002.
- [68] T. Xu, G. Jacobsen, S. Popov, J. Li, T. Liu, Y. Zhang, and P. Bayvel, "Analytical investigations on carrier phase recovery in dispersion-unmanaged n-PSK coherent optical communication systems," *Photonics*, vol. 3, no. 4, 2016.

- [69] L. Debnath and P. Mikusiński, *Introduction to Hilbert Spaces with Applications*. Elsevier Academic Press, 3rd ed., 2005.
- [70] J. L. Rojo-Alvarez, M. Martinez-Ramon, J. Munoz-Mari, and G. Camps-Valls, *Digital Signal Processing with Kernel Methods*. Wiley-IEEE Press, 1st ed., 2018.
- [71] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, no. 1, pp. 82–95, 1971.
- [72] S. v. Vaerenbergh, *Kernel Methods for Nonlinear Identification, Equalization and Separation of Signals*. PhD thesis, Universidad de Cantabria, 2010.
- [73] C. E. Rasmussen and K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [74] D. Mackay, "Introduction to Gaussian processes," in *NATO ASI Series F Computer and Systems Sciences*, pp. 133–166, 1998.
- [75] M. A. Z. Al-Khateeb, M. Tan, M. A. Iqbal, A. Ali, M. E. McCarthy, P. Harper, and A. D. Ellis, "Experimental demonstration of 72% reach enhancement of 3.6Tbps optical transmission system using mid-link optical phase conjugation," *Opt. Express*, vol. 26, pp. 23960–23968, Sep 2018.
- [76] A. A. I. Ali, M. Al-Khateeb, T. Zhang, F. Ferreira, and A. Ellis, "Enhanced nonlinearity compensation efficiency of optical phase conjugation system," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2019.
- [77] T. T. Nguyen, S. Boscolo, A. A. I. Ali, S. Sygletos, S. Takasaka, R. Sugizaki, and A. D. Ellis, "Digital compensation of imperfect pump counter-phasing induced phase distortion in optical phase conjugation of high-order qam," *Optics Express*, vol. 29, no. 11, pp. 17464–17475, 2021.
- [78] M. Lax, "Classical noise. v. noise in self-sustained oscillators," *Phys. Rev.*, vol. 160, pp. 290–307, Aug 1967.
- [79] J. Pedro and S. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 4, pp. 1150–1163, 2005.
- [80] J. Wang and K. Petermann, "Small signal analysis for dispersive optical fiber communication systems," *Journal of Lightwave Technology*, vol. 10, no. 1, pp. 96–100, 1992.
- [81] A. D. Ellis, M. E. McCarthy, M. A. Z. Al-Khateeb, and S. Sygletos, "Capacity limits of systems employing multiple optical phase conjugators," *Opt. Express*, vol. 23, pp. 20381–20393, Aug 2015.
- [82] A. Mussot, A. Durecu-Legrand, E. Lantz, C. Simonneau, D. Bayart, H. Maillotte, and T. Sylvestre, "Impact of pump phase modulation on the gain of fiber optical parametric amplifier," *IEEE Photonics Technology Letters*, vol. 16, no. 5, pp. 1289–1291, 2004.
- [83] M. Gao, C. Jiang, W. Hu, J. Wang, and H. Ren, "The effects of pump phase modulation on noise characteristics of fiber optical parametric amplifiers," *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, vol. 40, no. 3, pp. 431–436, 2006.

- [84] A. Vedadi, A. Mussot, E. Lantz, H. Maillotte, and T. Sylvestre, "Theoretical study of gain distortions in dual-pump fiber optical parametric amplifiers," *Optics Communications*, vol. 267, no. 1, pp. 244–252, 2006.
- [85] B. Widrow, J. McCool, and M. Ball, "The complex LMS algorithm," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 719–720, 1975.
- [86] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1–16, 2018.
- [87] D. Duvenaud, "The kernel cookbook," *Advice on Covariance functions*, pp. 12–39, 2015.
- [88] J. Lloyd, D. Duvenaud, R. Grosse, J. Tenenbaum, and Z. Ghahramani, "Automatic construction and natural-language description of nonparametric regression models," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, Jun. 2014.
- [89] J. Kivinen, A. Smola, and R. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2165–2176, 2004.
- [90] J. Platt, "A resource-allocating network for function interpolation," *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.
- [91] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.
- [92] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.
- [93] J. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1, pp. 85–105, 2002.
- [94] S. Van Vaerenbergh, J. Via, and I. Santamaria, "A sliding-window kernel RLS algorithm and its application to nonlinear channel identification," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5, pp. V–V, 2006.
- [95] B. de Kruif and T. de Vries, "Pruning error minimization in least squares support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 696–702, 2003.
- [96] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 964–978, 2011.
- [97] R. Boloix-Tortosa, J. J. Murillo-Fuentes, and S. A. Tsaftaris, "The generalized complex kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5213–5222, 2019.
- [98] R. Boloix-Tortosa, J. J. Murillo-Fuentes, I. Santos, and F. Pérez-Cruz, "Widely linear complex-valued kernel methods for regression," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5240–5248, 2017.

- [99] S. Olsson, H. Eliasson, E. Astra, M. Karlsson, and P. Andrekson, “Long-haul optical transmission link using low-noise phase-sensitive amplifiers,” *Nature Communications*, vol. 9, 06 2018.
- [100] F. A. Tobar, A. Kuh, and D. P. Mandic, “A novel augmented complex valued kernel LMS,” in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 473–476, 2012.