

**Digital signal processing for fiber optic communication systems
New approaches**

Bajaj, V.

DOI

[10.4233/uuid:c2031b74-bb40-498a-8572-3bc31795fb29](https://doi.org/10.4233/uuid:c2031b74-bb40-498a-8572-3bc31795fb29)

Publication date

2025

Document Version

Final published version

Citation (APA)

Bajaj, V. (2025). *Digital signal processing for fiber optic communication systems: New approaches*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:c2031b74-bb40-498a-8572-3bc31795fb29>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

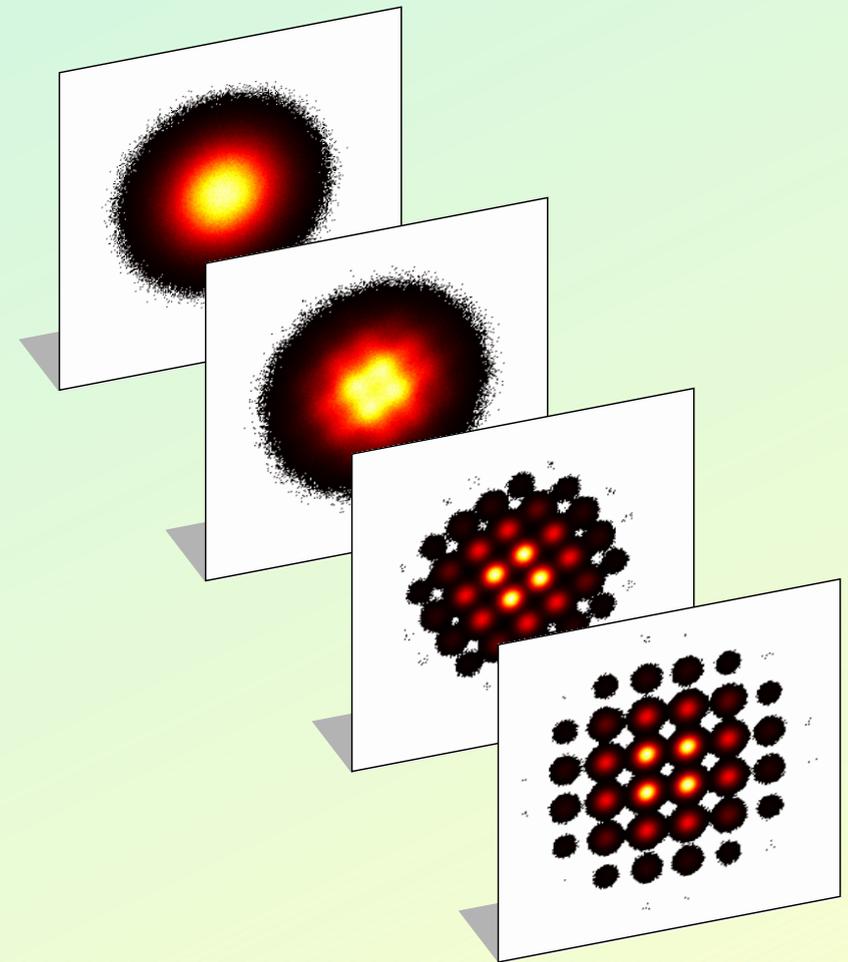
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

DIGITAL SIGNAL PROCESSING FOR FIBER OPTIC COMMUNICATION SYSTEMS

NEW APPROACHES



DIGITAL SIGNAL PROCESSING FOR FIBER OPTIC COMMUNICATION SYSTEMS: NEW APPROACHES

Vinod Bajaj

Vinod Bajaj

DIGITAL SIGNAL PROCESSING FOR FIBER OPTIC COMMUNICATION SYSTEMS

NEW APPROACHES

DIGITAL SIGNAL PROCESSING FOR FIBER OPTIC COMMUNICATION SYSTEMS

NEW APPROACHES

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Friday 25 April 2025 at 12:30 o' clock

by

Vinod BAJAJ

Master of Technology, Indian Institute of Technology Madras, India
born in Jodhpur, India.

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. M. Verhaegen,	Delft University of Technology, promoter
Dr. Ing. R. Van de Plas,	Delft University of Technology, promoter
Prof. Dr.-Ing. S. Wahls,	Karlsruhe Institute of Technology, Germany, promoter

Independent members:

Prof. G. J. T. Leus,	Delft University of Technology
Prof. Dr. D. Zibar,	Technical University of Denmark, Denmark
Dr. C. Okonkwo,	Technical University of Eindhoven
Dr. É. Awwad,	Télécom Paris, France
Prof. dr. ir. B. De Schutter,	Delft University of Technology, reserve member

Other members:

Dr. V. Aref,	Nokia Solutions and Networks GmbH & Co. KG Germany
--------------	-------------------------------------------------------

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766115.



Keywords: digital signal processing, fiber-optic communication, nonlinear Fourier transform, deep neural networks, polarization demultiplexing, probabilistic constellation shaping

Printed by: Drukkerij Haveka

Cover by: Constellation diagrams of PCS 64-QAM signal at various stages of polarization demultiplexing using FDJD-pr-MMA from Chapter 6, shown for one polarization. From back to front: mixed state, after removing cross mixing, after removing transmitter correlation and after pr-MMA.

Copyright © 2025 by V. Bajaj

ISBN 978-94-6518-036-6

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

CONTENTS

Summary	ix
Samenvatting	xiii
Acknowledgments	xvii
1 Introduction	1
1.1 Brief History of Fiber-Optic Communication Systems	2
1.2 Coherent Fiber-Optic Communication Systems	8
1.3 Digital Signal Processing for Coherent Fiber-Optic Communication Systems	10
1.3.1 Transmitter DSP (Tx-DSP)	10
1.3.2 Receiver DSP (Rx-DSP)	13
1.4 Advanced topics in DSP of Coherent Fiber-Optic Communication Systems	19
1.5 Nonlinear Fourier transform for Fiber-Optic Communication System	21
1.5.1 Soliton communication	23
1.6 Deep learning for DSP in FOCS	27
1.7 Outline of the Dissertation	31
2 Exact NFDm Transmission in the Presence of Fiber-Loss	35
2.1 Introduction	36
2.2 Basics of NFDm	37
2.3 Dispersion Decreasing Fiber and modified NFT	39
2.4 Simulation Setup and Results	41
2.5 Conclusion	49
3 Deep Neural Network-Based Digital Pre-distortion for High Baudrate Optical Coherent Transmission	51
3.1 Introduction	52
3.2 High Baud Rate Coherent Optical Transmitter	53
3.3 Review of Various DPD Techniques	55
3.3.1 Direct vs. Indirect Learning Architecture	55
3.3.2 Linear DPD	56
3.3.3 Volterra series-based DPD	57
3.4 Neural Network-based DPD	57
3.4.1 Soft-DAC Activation Unit	59
3.4.2 Initialization and Training of the NN-based DPDs	60

3.5	Results	60
3.5.1	Required Filter Length	61
3.5.2	Training Procedure.	62
3.5.3	DAC Voltage Variation	62
3.5.4	Verification of Pattern Independence	66
3.5.5	Evaluation in the Fiber Transmission Scenario	66
3.6	Computational complexity and pruning	67
3.7	Conclusion	69
4	54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion	71
4.1	Introduction	72
4.2	Proposed Neural Network-Based Joint DPD	72
4.3	Field Trial Experimental Setup	74
4.4	Results	74
4.5	Conclusion	75
5	Performance Analysis of Recurrent Neural Network-based Digital Pre-Distortion for Optical Coherent Transmission	77
5.1	Introduction	77
5.2	RNN-based DPDs	78
5.3	Training and Hyper-parameter Search	80
5.4	Results	80
5.5	Conclusion	82
6	Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations	83
6.1	Introduction	85
6.2	Brief Review of Polarization Mode Dispersion.	86
6.3	JD Assisted pr-MMA (JD-pr-MMA)	87
6.4	Proposed FDJD-pr-MMA for Convolutive Mixing	90
6.4.1	FDJD Algorithm	91
6.5	Additional Processing Blocks	93
6.5.1	Ambiguity Removal	93
6.5.2	Reconstruction of the Normalized PMD Channel Response	96
6.5.3	Candidate Phase Response Search	96
6.6	Simulation Setup	96
6.7	Results	97
6.7.1	JD-pr-MMA (RSOP channel and low PMD).	97
6.7.2	FDJD-pr-MMA (medium PMD)	100
6.8	Discussion	103
6.8.1	Pulse shaping	103
6.8.2	Effect of phase noise and frequency offset	104
6.8.3	Effect of transmit filters	104
6.8.4	Computational complexity.	105
6.9	Conclusion	105

7	Conclusions and Recommendations	107
7.1	Conclusions.	108
7.2	Recommendations for Future Work.	111
	List of Publications	135
7.3	Journal articles	135
7.4	Conference proceedings	135
	Curriculum Vitae	137

SUMMARY

The global internet protocol (IP) traffic is rising exponentially due to the increased number of bandwidth-intensive services like video-on-demand and cloud computing. To meet the demands of growing traffic, the data rates of fiber optic communication systems (FOCSs) need to be increased. In this regard, digital signal processing (DSP), which already plays a powerful role in the modern FOCSs, is being explored. Increasing the net data rate of FOCSs requires compensation for nonlinear impairments that can arise from the Kerr effect during the propagation of signal through fiber as well as from the non-ideal responses of transceiver hardware components. Furthermore, cutbacks in net data rates due to training overheads, i.e., the non-information carrying part of transmitted data consumed by DSP algorithms; need to be reduced. In this dissertation, we propose novel DSP approaches to address these problems of great practical interest.

The Kerr nonlinear effects add phase shifts to the signal, which are dependent on its instantaneous power. These phase distortions occur simultaneously with the dispersion effect of the fiber, which spreads signal pulses in time. The interplay is complicated and makes compensation of distortions challenging. The nonlinear Fourier transform (NFT), which offers immunity from the distortions of the Kerr effect, received great interest in recent years. The lossless nonlinear Schrödinger equation (NLSE), which models signal propagation in an ideal lossless optical fiber, belongs to a class of nonlinear partial differential equations known as integrable equations. These integrable equations can be solved exactly by NFT. Similar to the Fourier transform that translates a linear dispersive propagation in the time domain into phase delays in the signal spectrum, the NFT translates the nonlinear evolution of the signal governed by the lossless NLSE into trivial multiplications in the nonlinear Fourier spectrum of the signal. The NFT is exact for lossless fiber channels. In the presence of loss, the integrability property is violated. In lossy propagation, signal power reduces as it propagates. This in turn reduces the strength of the nonlinear effects along the length of the fiber. As practical fibers are lossy, the path-average approximation is often used to apply NFT on lossy fiber channels. In this approximation, the variation in the Kerr nonlinear effects due to the reduction in signal power is accounted as the variations in the Kerr-nonlinearity parameter of the fiber. Then, by approximating the varying Kerr-nonlinearity parameter with its average value over a span, a lossless fiber model is obtained. This approximation has errors associated with it which sacrifices the performance. We developed a NFT-based transmission system that is exact even in the presence of fiber loss. The proposed design eliminates errors due to loss, thus improving performance over the design that uses path-average approximation.

Digital pre-distortion (DPD) is a technique that is often used in the transmitter DSP to pre-compensate the transmitter distortions. Usually, a linear DPD is applied that corrects for the limited bandwidth of the components. In this approach, the variation in amplitude of the modulating signal is kept small to avoid excitation of nonlinear re-

gions of components. This results in a reduced signal power and signal-to-noise ratio (SNR). Larger variations in amplitude can improve SNR but may require a nonlinear DPD to pre-compensate the increased nonlinear distortions. Nonlinear DPDs based on the Volterra series are commonly used in radio-frequency amplifiers and coherent optical transmitters. Another type of DPD is based on neural networks (NNs), which received great attention recently, although its use in optical coherent transmitters remains limited to specific components or simulated models. Additionally, the considered systems operated at low symbol rates and did not use very high-order formats. However, above 100 G symbol rates and in very higher-order modulation scenarios, which is where next-generation transceivers operate, distortions due to large memory effects and stronger nonlinearity become more challenging. We developed a NN-based DPD for a state-of-the-art 128 G symbol rate coherent optical transmitter which provided a significant improvement in SNR. Furthermore, it enabled us to achieve a record 1.61 Tbps net data rate over an 80 km transmission link. The NN-DPD was compared with other DPDs, and its computational complexity was assessed and reduced through pruning. The proposed NN-DPD architecture was derived from feed-forward and convolutional NNs. To further reduce the complexity, we explored recurrent architectures. Recurrent NNs (RNNs) are known to have the capability to capture long memory in time-series data. In FOCSS, RNNs were shown to have better performance in fiber-nonlinearity equalization and end-to-end learning of transceivers. We investigated three popular bidirectional RNNs and found that DPDs based on them used more computations than our FFNN-based DPD to achieve similar performance. Finally, we also implemented IQ cross-talk compensation in our FFNN-based DPD. IQ crosstalk is the interference between two tributaries (i.e., its real and imaginary parts) of the signal that modulate on the carrier. We evaluated this DPD in an experiment over a field-deployed fiber of 48 km. A total data rate of 54.5 terabits per second was demonstrated by using 35 wavelengths. We also demonstrated that the NN-DPD is stable over time at its peak performance.

We also developed two non-data-aided (blind) algorithms for polarization demultiplexing of probabilistically shaped signals. The conventional blind algorithms used for demultiplexing are only suitable for uniformly distributed signals, i.e. when the transmission probability of each symbol is equal. Because of probabilistic shaping, signal distribution tends to be Gaussian. It is known that independently and identically distributed (iid) Gaussian signals, when mixed, cannot be separated from their mixture. To circumvent this problem, appropriate modifications in the conventional algorithms were proposed in the literature. We developed an alternative approach based on the fact that if Gaussian signals exhibit temporal correlations, it is possible to separate them from their mixture. We add temporal correlations to the probabilistically shaped signals at the transmitter, which are subsequently leveraged at the receiver to separate the polarizations. We demonstrated that our proposed two blind algorithms are effective in compensating for polarization impairments.

In summary, DSP is considered a flexible way to increase the data rates of modern fiber-optic communication systems by compensating for nonlinear distortions coming from Kerr effects and transceiver hardware. Moreover, the use of blind algorithms can minimize the reduction in net data rates by avoiding training overheads. For the Kerr nonlinearity, nonlinear Fourier transform (NFT)-based transmission systems are pro-

posed. However, they face practical challenges such as propagation losses of fibers and transceiver distortions. A method presented in this dissertation makes the NFT-based systems exact even in the presence of fiber loss, removing one of the practical challenges. The proposed design avoids any approximations and brings a comparative advantage in performance. To address another practical challenge imposed by the transceiver distortions, we developed a neural network (NN)-based DPD technique. The proposed NN-DPD compensates for distortions of a high symbol rate transmitter and significantly improves the signal integrity. We thus showed that hardware nonlinearity compensation matters in next-generation transceivers. Finally, we proposed two novel non-data-aided algorithms that are effective in mitigating polarization impairments in probabilistically shaped signals. Our algorithms demonstrate a non-conventional way to tackle the blind-demultiplexing problem. We believe it paves a new way to develop algorithms for DSP of fiber-optic communication systems.

SAMENVATTING

Het wereldwijde internet protocol (IP) verkeer neemt exponentieel toe als gevolg van het toegenomen aantal bandbreedte-intensieve diensten zoals video-on-demand en cloud computing. Om aan de eisen van het groeiende verkeer te voldoen, moeten de datasnelheden van glasvezelcommunicatiesystemen “Fiber-optic communication systems (FOCSs)” worden verhoogd. In dit opzicht wordt digitale signaalverwerking “Digital signal processing (DSP)”, die al een krachtige rol speelt in de moderne FOCSs, onderzocht. Het verhogen van de netto datasnelheid van FOCSs vereist compensatie voor niet-lineaire beperkingen die kunnen voortvloeien uit het Kerr-effect tijdens de voortplanting van signalen door glasvezel evenals door de niet-ideale reacties van hardwarecomponenten van de zondontvanger. Verder bezuinigingen op de netto datasnelheden als gevolg van overheadkosten voor training, zoals voor het niet-informatiedragende deel van de verzonden data die worden verbruikt door DSP-algoritmen, moeten worden verminderd. In dit proefschrift stellen we nieuwe DSP-benaderingen voor om deze problemen van groot praktisch belang aan te pakken.

De niet-lineaire effecten van Kerr voegen faseverschuivingen toe aan het signaal, die afhankelijk zijn van het momentane vermogen ervan. Deze fasevormingen treden gelijktijdig op met het dispersie-effect van de vezel, die signaalpulsen in de tijd verspreidt. Het samenspel is ingewikkeld en maakt het compenseren van vervormingen lastig. De niet-lineaire Fourier-transformatie “nonlinear Fourier transform (NFT)”, die immuniteit biedt tegen de vervormingen van het Kerr-effect, kreeg de afgelopen jaren grote belangstelling. De verliesloze niet-lineaire Schrödingervergelijking “nonlinear Schrödinger equation (NLSE)”, die de signaalvoortplanting in een ideale verliesvrije optische vezel modelleert, behoort tot een klasse van niet-lineaire partiële differentiaalvergelijkingen bekend als integreerbare vergelijkingen. Deze integreerbare vergelijkingen kunnen exact worden opgelost met NFT. Vergelijkbaar met de Fourier-transformatie die een lineaire dispersieve voortplanting in het tijdsdomein vertaalt in fasevertragingen in het signaalspectrum, de NFT vertaalt de niet-lineaire evolutie van het signaal dat wordt bestuurd door de verliesloze NLSE in triviale vermenigvuldigingen in het niet-lineaire Fourier-spectrum van het signaal. De NFT is exact voor verliesloze glasvezelkanalen. Bij verlies wordt de integreerbaarheidseigenschap geschonden. Bij verliesgevende voortplanting neemt het signaalvermogen af naarmate het zich voortplant. Dit vermindert op zijn beurt de sterkte van de niet-lineaire effecten langs de lengte van de vezel. Omdat praktische vezels verlieslatend zijn, wordt vaak de padgemiddelde benadering gebruikt om NFT toe te passen op verliesgevende vezelkanalen. In deze benadering wordt de variatie in de niet-lineaire Kerr-effecten als gevolg van de vermindering van het signaalvermogen beschouwd als de variaties in de Kerr-niet-lineariteitsparameter van de vezel. Door vervolgens de variërende Kerr-niet-lineariteitsparameter te benaderen met zijn gemiddelde waarde over een periode, wordt een verliesvrij vezelmodel verkregen. Deze benadering brengt fouten met zich mee die nadelig zijn voor de prestaties. We

hebben een op NFT gebaseerd transmissiesysteem ontwikkeld dat zelfs bij vezelverlies nauwkeurig is. Het voorgestelde ontwerp elimineert fouten als gevolg van verlies, waardoor de prestaties worden verbeterd ten opzichte van het ontwerp dat padgemiddelde benadering gebruikt.

Digitale pre-distortion “digital pre-distortion (DPD)” is een techniek die vaak wordt gebruikt in de zender-DSP om de zender vervormingen vooraf te compenseren. Meestal wordt een lineaire DPD toegepast die corrigeert voor de beperkte bandbreedte van de componenten. Bij deze benadering wordt de variatie in amplitude van het modulende signaal klein gehouden om excitatie van niet-lineaire gebieden van componenten te vermijden. Dit resulteert in een verminderd signaalvermogen en een verminderde signaal-ruisverhouding “signal to noise ratio (SNR)”. Grotere variaties in amplitude kunnen de SNR verbeteren, maar vereisen mogelijk een niet-lineaire DPD om de toegenomen niet-lineaire vervormingen vooraf te compenseren. Niet-lineaire DPDs gebaseerd op de Volterra-serie worden vaak gebruikt in radiofrequentieversterkers en coherente optische zenders. Een ander type DPD is gebaseerd op neurale netwerken “neural networks (NNs)”, die de laatste tijd veel aandacht hebben gekregen, hoewel het gebruik ervan in optische coherente zenders beperkt blijft tot specifieke componenten of gesimuleerde modellen. Bovendien werkten de onderzochte systemen met lage symboolsnelheden en gebruikten ze geen formaten van zeer hoge orde. Boven de 100 G symboolsnelheden en in modulatiescenario's van zeer hogere orde, waar transceivers van de volgende generatie werken, worden vervormingen als gevolg van grote geheugenefecten en sterkere niet-lineariteit echter een grotere uitdaging. We hebben een op NN gebaseerde DPD ontwikkeld voor een ultramoderne coherente optische zender met 128 G symboolsnelheid die een aanzienlijke verbetering in SNR opleverde. Bovendien hebben we hierdoor een recordsnelheid van 1,61 Tbps netto data kunnen bereiken over een transmissieverbinding van 80 km. De NN-DPD werd vergeleken met andere DPDs en de rekencomplexiteit ervan werd geëvalueerd en verminderd door middel van vereenvoudiging. De voorgestelde NN-DPD-architectuur is afgeleid van “feed-forward neural networks (FFNNs)” en convolutionele NNs. Om de complexiteit verder te verminderen, hebben we terugkerende architecturen onderzocht. Het is bekend dat terugkerende NNs “recurrent neural networks (RNNs)” het vermogen hebben om lang geheugen vast te leggen in tijdreeksgegevens. In FOCSs bleek dat RNNs betere prestaties leverden bij het egaliseren van niet-lineariteit van vezels en het end-to-end leren van zenderontvangers. We onderzochten drie populaire bidirectionele RNNs en ontdekten dat daarop gebaseerde DPDs meer berekeningen gebruikten dan onze op FFNN gebaseerde DPD om vergelijkbare prestaties te bereiken. Ten slotte hebben we ook IQ-overspraakcompensatie geïmplementeerd in onze op FFNN gebaseerde DPD. IQ-overspraak is de interferentie tussen twee zijtakken (dat wil zeggen de reële en denkbeeldige delen ervan) van het signaal die op de draaggolf moduleren. We hebben deze DPD geëvalueerd in een experiment over een in het veld opgestelde vezel van 48 km. Een totale datasnelheid van 54,5 terabit per seconde werd aangetoond door gebruik te maken van 35 golf lengten. We hebben ook aangetoond dat de NN-DPD in de loop van de tijd stabiel is op zijn topprestaties.

We hebben ook twee niet-data-ondersteunde (blinde) algoritmen ontwikkeld voor de polarisatie-demultiplexing van probabilistisch gevormde signalen. De conventionele blinde algoritmen die worden gebruikt voor demultiplexing zijn alleen geschikt voor

uniform verdeelde signalen, dat wil zeggen wanneer de transmissiewaarschijnlijkheid van elk symbool gelijk is. Door de probabilistische vormgeving neigt de signaalverdeling naar Gaussisch. Het is bekend dat onafhankelijk en identiek verdeelde “independent and identically distributed (iid)” Gaussische signalen, wanneer ze worden samengevoegd, niet meer van elkaar kunnen worden gescheiden. Om dit probleem te omzeilen, werden in de literatuur passende wijzigingen in de conventionele algoritmen voorgesteld. We hebben een alternatieve benadering ontwikkeld, gebaseerd op het feit dat als Gaussische signalen temporele correlaties vertonen, het mogelijk is deze wel van elkaar te scheiden. We voegen temporele correlaties toe aan de probabilistisch gevormde signalen bij de zender, die vervolgens bij de ontvanger worden gebruikt om de polarisaties te scheiden. We hebben aangetoond dat onze voorgestelde twee blinde algoritmen effectief zijn in het compenseren van polarisatiestoornissen.

Samenvattend wordt DSP beschouwd als een flexibele manier om de datasnelheden van moderne glasvezelcommunicatiesystemen te verhogen door te compenseren voor niet-lineaire vervormingen die afkomstig zijn van Kerr-effecten en transceiverhardware. Bovendien kan het gebruik van blinde algoritmen de vermindering van de netto datasnelheden minimaliseren door trainingsoverheads te vermijden. Voor de niet-lineariteit van Kerr worden op niet-lineaire Fourier-transformatie “nonlinear Fourier transform (NFT)” gebaseerde transmissiesystemen voorgesteld. Ze worden echter geconfronteerd met praktische uitdagingen, zoals voortplantingsverliezen van vezels en transceiververvormingen. Een methode die in dit proefschrift wordt gepresenteerd, maakt de op NFT gebaseerde systemen exact, zelfs in de aanwezigheid van vezelverlies, waardoor een van de praktische uitdagingen wordt weggenomen. Het voorgestelde ontwerp vermijdt elke benadering en brengt een vergelijkbaar voordeel op het gebied van prestaties met zich mee. Om een andere praktische uitdaging aan te pakken die wordt opgelegd door de vervormingen van de transceiver, hebben we een op een neurale netwerk gebaseerde DPD-techniek ontwikkeld. De voorgestelde NN-DPD compenseert vervormingen van een zender met hoge symbolnelheid en verbetert de signaalintegriteit aanzienlijk. We hebben dus aangetoond dat niet-lineariteitscompensatie van hardware van belang is in transceivers van de volgende generatie. Ten slotte hebben we twee nieuwe, niet-data-ondersteunde algoritmen voorgesteld die effectief zijn in het verminderen van polarisatiestoornissen in probabilistisch gevormde signalen. Onze algoritmen demonstreren een niet-conventionele manier om het blind-demultiplexprobleem aan te pakken. Wij geloven dat dit een nieuwe manier opent om algoritmen te ontwikkelen voor DSP van glasvezelcommunicatiesystemen.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to all those who have offered their support and guidance throughout my PhD journey.

Firstly, sincere thanks to my academic supervisor, Sander Wahls and industrial supervisor, Vahid Aref, for trusting me with this PhD project. Their research inspirations and constructive feedback have been invaluable. I am grateful for their readiness to help, often at a short notice. I deeply admire Sander's kindness and scientific temperament, qualities that I tried to embody throughout my PhD. I am also grateful to Vahid for hosting me at NBL. His technical insights and thoughtful suggestions have enriched my experience and understanding. Thank you both for your unwavering support. My sincere gratitude also extends to my promoters Raf Van de Plas and Michel Verhaegen for their encouragement and constructive feedback. I am also deeply grateful to the FONTE grant for funding three years of my PhD, and to the entire FONTE team for their valuable discussions and feedback. Special thanks go to Sergie K Turitsyn, for his guidance and encouragement, and to Christiane D Saad, for her help with bureaucratic matters. Furthermore, I appreciate Fred Buchali and Mathieu Chagon for contributing to my work at NBL. It has been a wonderful experience working with all of you, and I am profoundly thankful for the opportunity.

I also extend my thanks to DCSC colleagues for creating a fun environment. Special thanks to Srinivas, Pascal and Peter for fruitful knowledge sharing on topics of NFT and optical communication. Thanks to other colleagues Sina, Abhimanyu, Roger, Jacques, Markus, Yu Chen, Hamed, Leonoor, Tushar, Ali, and Shahzeb: thank you all for enjoyable talks during lunch and coffee breaks. Thanks to Marieke, Heleen, Francy, Erica, Sandra and other supporting staff for your assistance with practical matters.

I extend my gratitude to my friends Ankush, Nimish and Nandlal for creating a home-like environment and for helping me during my relocations. Thanks to Vladislav, Stenio, Fabiano and Hisham for the enjoyable moments in Stuttgart. I am grateful to my friends Lakshmi Narayanan, Amol, Rohan, Pankaj and Subhash for cheering me up from time to time. A heartfelt thank you to Lakshmi Narayanan for his inspiring and calming conversations.

Finally, my deepest gratitude to my family and my wife, Nivedita whose support cannot be expressed with words. Thank you Nivedita for your unconditional love and support during the ups-and-downs of this journey. Thank you for being there.

*Vinod Bajaj
Paris, July 2024*

1

INTRODUCTION

This chapter introduces applications of digital signal processing (DSP) algorithms for coherent fiber-optic communication systems (FOCSs). First, we review a brief history of FOCSs. Then, we describe typical architectures of coherent FOCSs. In the third section, we provide an overview of DSP algorithms used in modern coherent FOCSs. Afterwards, advanced DSP algorithms for the compensation of nonlinear impairments in FOCSs in order to increase the data rates are discussed. The outline of this dissertation is provided in the final section.

1.1. BRIEF HISTORY OF FIBER-OPTIC COMMUNICATION SYSTEMS

Telecommunication systems have become an essential infrastructure today. They carried around 46 terabytes per second of global internet protocol (IP) traffic, which is a proxy for data flows, over the whole world in 2017 [1]. Fiber-optic communication systems (FOCSs) are information highways transporting huge chunks of data at tremendous speed. The main advantage of FOCSs over other communication systems is the use of visible or infrared lightwaves as a carrier to transmit data. The frequency of carriers in FOCSs is around 100 terahertz, which is 10,000 times higher than that of typical microwave communication systems operating at 1 GHz. The use of such a high carrier frequency provides an enormous bandwidth for data transmission. In addition, optical fiber cables have much lower losses, are lighter, have a small form factor and are less susceptible to electromagnetic interference than other alternatives like coaxial cables. These differences made these systems, in a sense, superior to all other existing communication systems.

The history of FOCS is often categorized into several generations, where in each generation a remarkable growth in the information rate, which refers to the amount of data that can be transmitted per unit of time, is observed. Figure 1.1 shows the evolution of the optical fiber data rates over the years. A trend of almost 10 times increase in the information rate for every four years is observed. This growth is attained due to technical advances shown in the figure such as improved transmission fibers, Erbium-doped fiber amplifier (EDFA), and wavelength division multiplexing (WDM) technologies. Figure 1.2 summarizes novel technologies of each generation. The proposition of using dielectric waveguides as a medium that could guide optical energy can be considered as the starting point for the history of FOCSs [3]. Prior to that, it was clear that using lightwaves as carriers brings huge bandwidth, but free space was considered as a medium which is susceptible to weather conditions [4]. In dielectric waveguides, the refractive index of the cylindrical cross-section is decreased gradually or in a step along the radial direction. The parts of the cross-section with higher and lower refractive indices are called core and cladding, respectively. Such a refractive index profile enables total internal reflection of light when it is launched into these waveguides at a particular angle, allowing propagation along the length of the waveguides. However, the manufactured waveguides of that time were not suitable for signaling. They suffered from large losses in signal power. The realization of loss factors below 20 dB/km is considered a break-point for the widespread application of optical fibers in telecommunications because it allowed longer repeater spacing of more than 1 km, thus reducing cost [5], [6]. In 1977, live telephone traffic was carried by a fiber optic communication system for the first time, at a data rate of 1.54 Mb/s [7], [8]. During the period 1977-1979, many field demonstrations of FOCSs were carried out. These systems operated around 0.8 μm wavelength and carried about a 45 Mb/s data rate. At the transmitter, these systems modulate information on the intensity of the light by switching an optical source on or off in regular time intervals representing an input bit. The time interval is commonly referred to as symbol time duration or symbol time period. This type of information modulation is known as the on-off keying (OOK) format. At the receiver, a photo-diode is used that generates an electric current

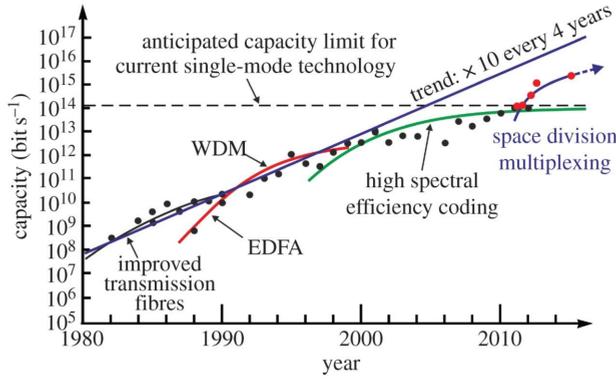


Figure 1.1: Evolution of information rates of fiber-optic communication systems. EDFA: Erbium-doped fiber amplifier, WDM: wavelength division multiplexing. The figure is taken from [2].

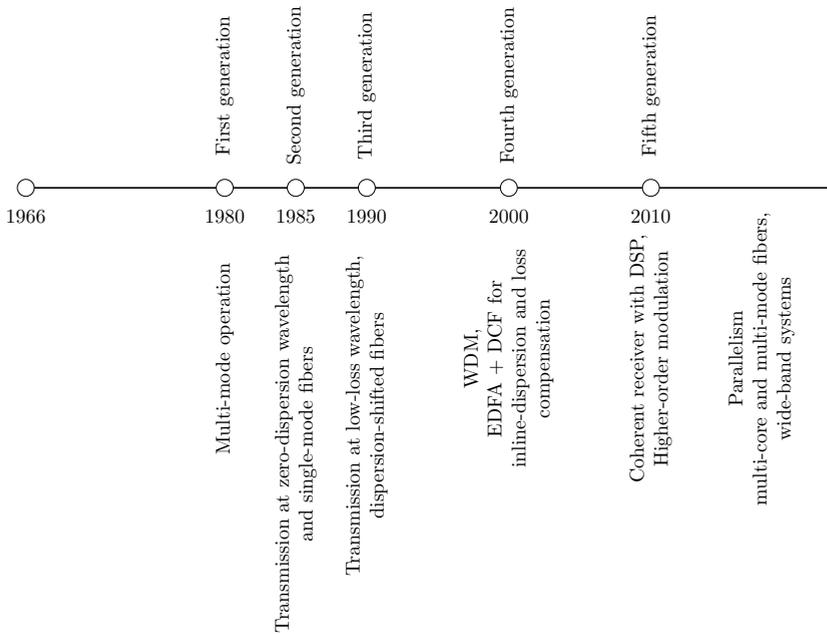


Figure 1.2: Evolution of fiber-optic communication systems in various generations and their corresponding features. WDM: wavelength division multiplexing, EDFA: Erbium-doped fiber amplifier, DCF: dispersion compensating fiber.

proportional to the instantaneous power of the received signal. A decision threshold is then applied to the photocurrent to detect the bits. These systems are fairly simple and known as intensity modulation and direct detection (IM-DD) systems. To compensate for the power losses during the propagation, a regenerator unit is placed at around every 10 km [9]. At each regenerator unit, the optical signal is regenerated electrically [7], i.e., it is detected and then modulated on an optical carrier for transmission over the next span. These systems were available commercially in 1980 and are often referred to as the first-generation of FOCSs.

The first-generation systems used multi-mode transmission technology. A mode refers to a specific solution to the wave equation, which is a second-order linear partial differential equation that describes the propagation of waves. An optical fiber, depending on its design parameters such as core radius and difference in refractive index of core and cladding, can support a finite number of guided modes at any frequency. The spatial distribution of an optical mode in a fiber does not change with propagation [9]. Each mode has a cut-off frequency above which it cannot propagate. Pulses exciting different modes of fiber propagate through different optical path lengths and reach the other end at different times leading to a cross-talk, referred to as inter-modal dispersion. In step-index fibers, the refractive index is mostly uniform in the core region with a sharp decrease at the core-cladding interface. This kind of refractive index profile causes large differences in optical path lengths of different modes which leads to high inter-modal dispersion. To minimize the inter-modal dispersion, graded-index fibers were used in the first generation. In these fibers, the refractive index decreases gradually in the core region in a way that the difference in optical path lengths of different modes is reduced. However, graded-index fibers cannot mitigate inter-modal dispersion completely in practice.

In the second generation, the signal transmission was restricted only to a single mode, also referred to as the fundamental mode of the fiber [10]. A single-mode step-index fiber is designed in such a way that all the higher-order modes are cut off at the operating frequency thus eliminating the inter-modal dispersion [9]. Let's denote \hat{x} , \hat{y} and \hat{z} as three orthogonal unit vectors in a 3 dimensional space in Cartesian coordinate system. The fundamental mode in a single-mode fiber is linearly polarized in either \hat{x} or \hat{y} direction assuming lightwave propagation in \hat{z} direction. This means a single-mode fiber (SMF) can support two orthogonal polarization modes. In an ideal SMF, each polarization mode can carry separate data without any interference but, in practice, manufacturing defects and environmental factors like stress make the behavior of a fiber-channel polarization dependent. This leads to an interference between the two orthogonal polarizations. This interference was not compensated at that time and therefore only single polarization was effectively used for data transmission.

In a single-mode condition, the propagation of a pulse in an optical fiber is approximated by the one-dimensional nonlinear Schrödinger equation (NLSE) [9]

$$\frac{\partial Q(\ell, t)}{\partial \ell} - \frac{\alpha}{2}Q(\ell, t) + j\frac{\beta_2}{2}\frac{\partial^2 Q(\ell, t)}{\partial t^2} - j\gamma|Q(\ell, t)|^2Q(\ell, t) = 0. \quad (1.1)$$

The term $Q(\ell, t)$ denotes the slowly varying complex envelope of the electric field $E(\ell, t) = \frac{1}{2}(Q(\ell, t)\exp(-j\omega_0 t) + Q^*(\ell, t)\exp(j\omega_0 t))$. Here, $j = \sqrt{-1}$ and $(.)^*$ denotes

complex-conjugate operation. The slowly varying envelope approximation is applied to separate the slowly changing variations in the field from the fast-changing variations at the optical frequency ω_0 . This approximation is valid when the spectrum of the field $\Delta\omega$ is much smaller relative to the carrier frequency ω_0 i.e., $\Delta\omega/\omega_0 \ll 1$. This is generally true for communication signals because $\omega_0 \sim 10^{15}$ per second and $\Delta\omega \sim 10^{10}$ per second. The pulse propagates at the group velocity v_g . In (1.1), t is the time in the frame of reference moving with the pulse at the group velocity. It is related to the absolute time t_0 , as $t_0 = t + \ell/v_g$ where ℓ is the propagation distance. The second term in the equation represents the optical power loss during the pulse propagation with α being the attenuation or loss parameter. The third term accounts for the chromatic dispersion that causes the pulse to spread in time as it propagates through the fiber. The group velocity dispersion parameter of fiber is denoted by $\beta_2 = \frac{d^2\beta(\omega)}{d\omega^2}$, where $\beta(\omega)$ is the phase constant of the wave. The dispersion coefficient β_2 varies with wavelength λ . The wavelength region where $\beta_2 > 0$, is called the normal dispersion region. In this region, the group velocity v_g increases with an increase in λ . On the other hand, in anomalous dispersion region i.e. $\beta_2 < 0$, the group velocity v_g decreases with an increase in λ . The fourth term accounts for the Kerr nonlinear effect [11], which adds phase shifts proportional to the instantaneous optical power of the pulse $|Q(\ell, t)|^2$. The nonlinear parameter of the fiber is denoted by γ . During the 1980s, FOCSs were operated in low-power regions that did not trigger the Kerr nonlinear effects. Thus, optical fiber behaved as a linear channel, i.e. (1.1) with $\gamma = 0$. After the mitigation of inter-modal dispersion, chromatic dispersion (the third term in (1.1)) became a main obstacle in increasing the transmission rates. It causes signal pulses to broaden beyond their symbol time duration and overlap with their neighboring pulses, an effect known as inter-symbol interference (ISI). It was recommended to use longer wavelengths from the infrared spectrum around $1.3 \mu\text{m}$, where the dispersion in glass fibers is minimal [12], [13]. A smaller dispersion meant lower ISI, which allowed faster data rates [14]. In the second generation, commercial systems that operated at up to 1.7 Gb/s bit rates and regenerator spacing of 50 km were available [9].

The regenerator spacing could not be increased further because of large attenuation ($\alpha \sim 0.5$ dB/km) of signal power and poor receiver sensitivity. Receiver sensitivity is the minimum optical power required in the received signal for successful detection. A receiver with poor sensitivity requires more optical power in the received signal, which limits the regenerator spacing. Although a special region around 1550 nm wavelength was previously known to have a low attenuation ($\alpha \sim 0.2$ dB/km) [13], dispersion was large at that wavelength, causing more ISI. Hence, it was necessary to operate at around 1310 nm where the lowest dispersion was observed. Later, dispersion-shifted fibers were designed that had low dispersion as well as low attenuation at around 1550 nm. These fibers allowed faster data rates and longer regenerator spacing. During this time, Erbium-doped fiber amplifiers (EDFAs) were developed [15], which could boost signal power optically, avoiding signal regeneration using optoelectronic conversion [16].

At the same time, receivers based on coherent detection schemes were also investigated because of their higher receiver sensitivity that could increase regenerator spacing [17]. The basic idea in these receivers is to combine the received signal with a continuous-wave optical field generated by a narrow-linewidth laser, called the local os-

cillator (LO). The received signal can be represented as $E_s = A_s \exp(-j\omega_0 t + \phi_s)$ and the LO signal as $E_{LO} = A_{LO} \exp(-j\omega_{LO} t + \phi_{LO})$. Here, A and ϕ denote the amplitude and the phase, respectively. The subscripts 's' and 'LO' represent signal and local oscillator, respectively. The photocurrent detected from the mixed signal is

$$I = R(|E_s + E_{LO}|^2), \quad (1.2)$$

where R is a proportionality constant. Since $|E_s|^2 = |A_s|^2$ and $|E_{LO}|^2 = |A_{LO}|^2$, we can expand (1.2) as

$$I = R \left(|A_s|^2 + |A_{LO}|^2 + 2\sqrt{|A_s|^2 \cdot |A_{LO}|^2} \cos(\omega_0 t - \omega_{LO} t + \phi_s - \phi_{LO}) \right). \quad (1.3)$$

If $A_s \ll A_{LO}$, the first term can be neglected. The second component (i.e. $R|A_{LO}|^2$) is a direct current component that can be filtered out. The resultant photocurrent will be $I = 2R\sqrt{|A_s|^2 \cdot |A_{LO}|^2}$, provided $\phi_s = \phi_{LO}$ and $\omega_0 = \omega_{LO}$. On the other hand, a direct detection receiver could only give photocurrent $I = R|A_s|^2$. The receiver sensitivity can be improved with coherent detection as it can detect a received signal with lower power by using a high-power LO. In 1990, third-generation systems were demonstrated that could operate at 2.5 Gb/s and transmit up to 2200 km using EDFAs and coherent detection [18]. However, coherent detection faced difficulties in locking the phase of LO with that of signal, i.e. in keeping $\phi_s = \phi_{LO}$. It wasn't a commercial success. Instead, wavelength division multiplexing (WDM) fueled the growth of data rates in the fourth generation.

WDM allows many data-carrying channels in parallel through a single fiber. In WDM, information is modulated on multiple wavelengths with each wavelength having its dedicated transmitter. All modulated lightwaves are multiplexed optically and sent through a single fiber. At the receiver side, these wavelengths are demultiplexed and the signal on each wavelength is processed with its receiver. The use of WDM technology increases the total data rates of fiber by a factor equal to the number of wavelengths. Transmission systems were demonstrated that used large bandwidth EDFAs and could amplify the entire WDM signal optically without needing regenerators [19], [20]. Long-distance transmissions without electrical regeneration led to a large accumulation of chromatic dispersion. Therefore, dispersion was managed periodically in the link by employing a special fiber with negative dispersion, also known as dispersion compensating fiber (DCF). DCFs undo the chromatic dispersion accumulated over the span of the standard SMF. Repeater modules, each consisting of DCF followed by an EDFA, replaced the regenerators. They were placed periodically in the long links after around every 80 km. In the fourth generation, it was possible to realize long-haul transmissions. In 1988, the first transatlantic fiber-optic cable TAT-8 came into service, whose data rate exceeded the combined throughput of all the existing transatlantic coaxial cables [21]. WDM technology provided a remarkable growth in data rates of FOCSS in the fourth generation [22], [23]. A data rate of terabits per second was possible over a single fiber by using the WDM technology [24].

The transmission systems used so far were linear systems because they did not take into account the Kerr nonlinearity effect [11]. However, nonlinear systems like soliton-based transmission systems were also investigated extensively [25]. Solitons are specially shaped pulses that do not spread during propagation because the linear disper-

sion effect is counterbalanced by the nonlinear Kerr effect. They promised a solution to the dispersion accumulation problem. At that time, one of the large-scale field trials demonstrated soliton-based transmission with about a 96 Gbits/s data rate over 5745 km [26]. However, soliton-based systems faced many practical challenges. For example, designing a soliton-based WDM system required dispersion compensation periodically along the length. Note that, periodic dispersion compensated links already existed for the linear systems that were relatively simpler to design. Another aspect is spectral efficiency (SE), which is the data rate carried by the signal per unit bandwidth in bits/s/Hz. Soliton systems offer a small SE because soliton pulses decay slowly and therefore need longer symbol time duration. On the other hand, the linear systems could use digital pulse shaping to minimize the time-frequency support of the signals. Therefore, linear systems were more attractive to telecommunication industries. More details on soliton-based systems are discussed later in 1.5.1. By the end of the fourth generation, it was well understood that to achieve a higher information rate, spectral efficiency needed to be improved.

Communication signals are often visualized by plotting the set of states of the optical field in the complex plane. These plots are known as constellation diagrams (see Figure 1.3) wherein the states are depicted by circles which are commonly called symbols. The on-off keying (OOK) signal was the most commonly used signal format so far. In a symbol time duration, OOK scheme transmit one of the two states of the optical field as shown in Figure 1.3(a) and thus can encode only one bit of information. Its maximum SE is 1 bit/symbol. On the other hand, a quadrature phase shift keying

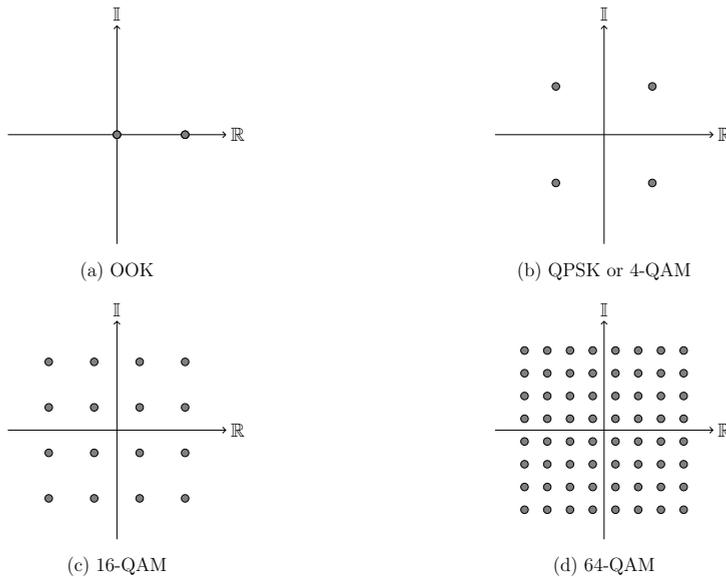


Figure 1.3: An example representation of various signals using the constellation diagrams in the complex plane. The signals shown are: on-off keying (OOK), quadrature phase shift keying (QPSK) also known as 4-quadrature amplitude modulation (QAM), 16-QAM and 64-QAM. Each circle represents a state in the optical field.

(QPSK) format (Figure 1.3(b)) has four states, which can represent On and Off states of two bits, while a 16-quadrature amplitude modulation (16-QAM) format (Figure 1.3(c)) has 16 states that can encode 4 bits. The maximum SE of 4-QAM (QPSK), 16-QAM and 64-QAM (Figure 1.3(b-d)) are 2, 4, and 8 bits/symbol, respectively. This type of multilevel amplitude and phase modulation was used in the fifth-generation systems wherein coherent detection retrieved the phase and amplitude of the received signal. We know that in coherent detection the phase of the LO laser needs to be locked with the phase of the transmit laser. During the 1980s, the locking was done using optical phase-locked loops (OPLLs) [17], [27], which were highly unstable [28]. The first coherent detection systems were demonstrated that did not use an OPLL [29], [30]. Instead, the phase locking was achieved using digital carrier phase recovery [31]. This type of receiver is known as the “digital coherent” receiver.

Digital coherent receivers use digital signal processing (DSP) to process the coherently detected signal. In addition to carrier recovery, DSP could also compensate for the polarization-related impairments from the optical fiber channel. This was interesting because now the information can be modulated on the two orthogonal polarizations and retrieved digitally at the receiver doubling the spectral efficiency. These receivers are called polarization diversity receivers [32]. Furthermore, with the magnitude and phase of the signal recovered at the receiver, other impairments like accumulated chromatic dispersion can also be compensated by digital equalization [33], [34]. This led to the development of application-specific integrated circuits (ASICs), which used fast analog to digital converters (ADCs) with DSP and supported optoelectronic hardware of coherent receivers [35]. A modern state-of-the-art system often uses these DSP techniques together with a polarization diversity configuration to maximize data rates through optical fibers.

In the current sixth generation, the growth in data rates is projected to come from massive parallelism by using ultra-wideband transmission and space-division multiplexing [36]. Investigations for commercial viability of going beyond the C-band (wavelength range of 1530 nm-1565 nm) to $C + L$ (1530nm-1625nm) and $S + C + L$ (1460nm-1625nm) bands are ongoing [37], [38]. In addition, researchers are exploring few-mode, multi-mode, and multi-core fibers transmission [39].

1.2. COHERENT FIBER-OPTIC COMMUNICATION SYSTEMS

Modern fiber-optic communication links of distances longer than 500 km use polarization-diverse coherent optical transceivers. An exemplary schematic of a single wavelength point-to-point coherent system is shown in Figure 1.4. Like other communication systems, it can be categorized into three parts: a transmitter, which encodes information on the optical carrier, a medium, through which the signal propagates (the optical fiber link), and a receiver that retrieves the information from the optical carrier.

A polarization-diverse coherent system uses phase, amplitude, as well as polarization of the optical carrier to transmit information. At the transmitter, incoming bits are processed using the DSP to synthesize a signal digitally. This signal is composed of four waveforms i.e. two pairs accounting for modulation in two orthogonal polarizations: horizontal (H) and vertical (V). Each pair modulates the real part i.e. in-phase denoted

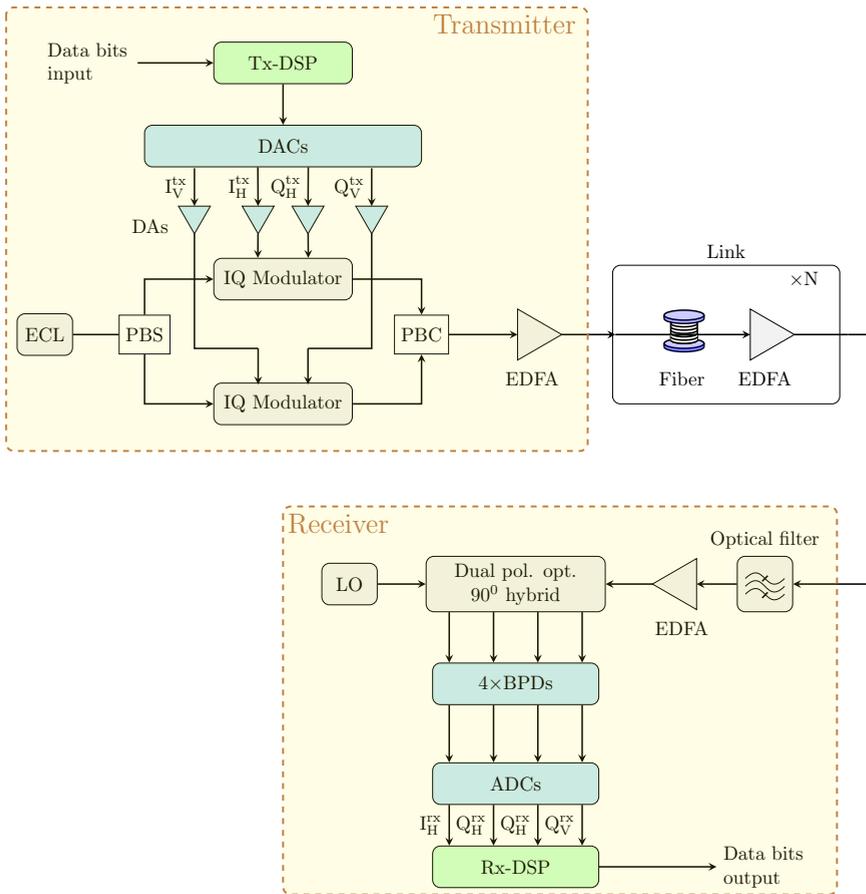


Figure 1.4: Schematic of a typical polarization-diverse coherent optical transmission system. [Tx: transmitter, DSP: digital signal processing, DACs: digital to analog converters, DAs: driver amplifiers, ECL: external cavity laser, PBS: polarization beam splitter, PBC: polarization beam combiner, EDFA: Erbium-doped fiber amplifier, LO: local oscillator, BPDs: balanced photo-detectors, ADCs: analog to digital converters, Rx: receiver]

by I and the imaginary part i.e. quadrature-phase denoted by Q of the electrical field in each polarization. The waveforms are loaded into the memory of digital-to-analog converters (DACs), which generate four electrical waveforms. The four waveforms are referred to as the I_H^{tx} , Q_H^{tx} , I_V^{tx} , and Q_V^{tx} tributaries in Figure 1.4. Four driver amplifiers (DAs) amplify these waveforms and provide them to the dual polarization (DP) IQ modulator. An optical carrier, generated from a narrow linewidth (spectral width) laser, is fed to the DP IQ modulator. In the modulator, the carrier is separated into two orthogonal polarizations, H & V, using a polarization beam splitter (PBS). Then, complex envelopes of each polarization of the carrier are separately modulated with corresponding I and Q waveforms. After modulation, both polarizations are combined using a polarization beam combiner (PBC) in the modulator. The output of the modulator is a DP modulated

optical signal. The signal is then amplified using an EDFA before it is injected into the optical fiber link.

The optical fiber link consists of several spans, where each span has around 80 km of fiber followed by a lumped amplification using EDFA. The gain of the EDFAs is set to compensate for the losses that occurred by propagation in the corresponding span. After amplification, the signal proceeds to the next span, until it reaches the receiver.

At the receiver, the signal is first filtered to remove out-of-band noise. Then, the signal is detected coherently by mixing it with a continuous wave laser (local oscillator) tuned at the same frequency as the optical carrier. As the signal is dual polarized, both the received signal and the local oscillators are split into two orthogonal polarizations and mixed in two separate optical 90° hybrids. At the output ports of the hybrids, four balanced photo-detectors (BPDs) detect the signal. Then, four analog-to-digital converters (ADCs) digitize the detected electrical signals, i.e., I_H^{rx} , Q_H^{rx} , I_V^{rx} , and Q_V^{rx} . These signals are processed with the receiver DSP in order to extract the transmitted information bits.

1.3. DIGITAL SIGNAL PROCESSING FOR COHERENT FIBER-OPTIC COMMUNICATION SYSTEMS

Digital signal processing offers flexibility in designing transmit signals. Furthermore, it can compensate for the impairments caused by transceiver hardware and propagation through fiber link. Although most of the DSP in coherent systems is similar to the DSP of wireless communication systems, modifications are required for optical fiber channels. Furthermore, the clock frequency of DSP in FOCSs needs to be much faster than that of wireless DSP as the signaling in FOCS takes place in the 10s of GHz range. Various DSP implementation techniques such as parallelization and pipelining are required to implement DSP algorithms for real-time data processing [40].

1.3.1. TRANSMITTER DSP (TX-DSP)

The transmitter DSP in a typical coherent fiber-optic communication system synthesizes the waveforms to be modulated on the optical carrier. It takes a bit-stream as input, applies several processing steps and provides four sequences I_H , Q_H , I_V , and Q_V , that can be loaded into the memory of DACs. The key steps of a typical transmitter DSP are shown in a block diagram in Figure 1.5 and explained below in detail.

FEC ENCODING

Forward error correction is a technique that allows to detect and correct errors in data transmission over noisy or unreliable communication channels. The FEC encoding stage

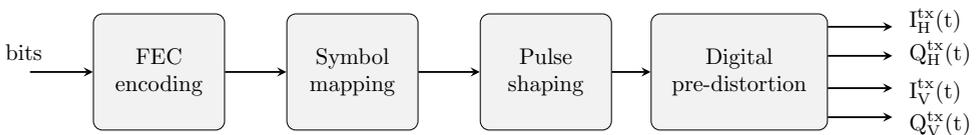


Figure 1.5: An example schematic of processing in transmitter DSP. [FEC: forward error correction]

adds redundant bits (overhead) on top of the input raw information bits. These redundant bits can be used at the receiver to detect and correct errors. A simple example of FEC encoding is to repeat each raw bit three times i.e., for each information bit there are two additional bits. During the FEC decoding at the receiver, decisions on all three bits are compared and the most common decision is chosen as the correct one. This way, if any one of the three bits is erroneous it will get corrected. In the FEC encoding, the additional bits are derived from the original bits. They do not add any new information. The additional bits are referred to as FEC overhead. The overhead of the simple FEC scheme explained above is $2/3$. In practice, the overhead is much lower. For example, Reed-Solomon FEC codes that are standardized for submarine links have $1/14$ overhead. It can bring down the bit error ratio (BER) i.e., the ratio of erroneous bits to total bits in a block, from 10^{-4} to 10^{-15} [41].

SYMBOL-MAPPING

After FEC encoding, groups of bits \mathbf{b}_k are mapped to complex-valued symbols a_k . Here, k denotes the symbol index. The symbols a_k are chosen from a set known as the modulation format, which is denoted by \mathbf{A} . The symbols a_k are often referred to as transmit symbols or ideal symbols. Usually, M-ary quadrature amplitude modulation (QAM) formats are used. Here, M represents the number of unique complex-valued symbols in set \mathbf{A} . M is chosen as some power of 2 so that a group of $\log_2 M$ information bits is mapped to a complex symbol, i.e. $\mathbf{b}_k = \{b_1, \dots, b_{\log_2 M}\}$. An example of a bit-mapping for a 16-QAM signal is shown in Figure 1.6. The mapping of bits is usually done using gray codes such that neighboring symbols differ in only one bit. This helps in minimizing bit errors over an additive white Gaussian noise (AWGN) channel. In a noisy AWGN channel, a symbol is often wrongly detected to be its close neighbor. In that case, though a symbol error is made, only 1 bit out of $\log_2 M$ will be erroneous.

If the symbols a_k are chosen uniformly from a M-QAM format then the signals are

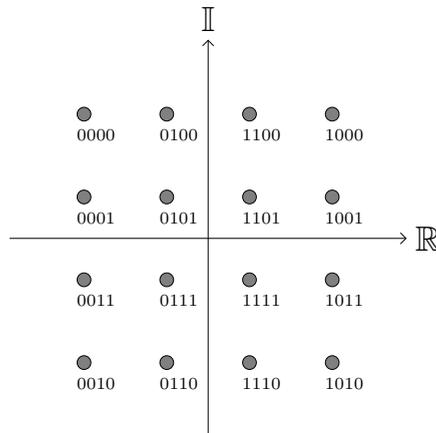


Figure 1.6: An illustration of one-to-one mapping between a group of four gray-coded bits to 16-quadrature amplitude modulation (16-QAM) symbols.

called “uniform M-QAM” signals. It can be useful to manipulate the probability of choosing these M-QAM symbols. This technique is called probabilistic constellation shaping (PCS).

PROBABILISTIC CONSTELLATION SHAPING

Probabilistic constellation shaping (PCS) is a technique to control the distribution of the complex-valued symbols used for communication. This enables the maximization of data rates and makes the system reach-and rate-adaptive [42], [43].

It is known that for an AWGN channel with a given noise variance and signal energy constraints, the optimal source distribution that maximizes the information rate is Gaussian [44], [45]. This result assumes that the source signal’s amplitude is continuous, which requires infinite resolution in DACs and ADCs. For a finite-size constellation set, a Maxwell-Boltzmann (MB) distribution is known to be optimal [45]. In PCS, the complex-valued symbols $a_k \in \mathcal{A}$ are drawn according to the MB distribution, independently with probability $P_A(a_k)$ given by

$$P_A(a_k) = \frac{e^{-\lambda_s |a_k|^2}}{\sum_{a_m \in \mathcal{A}} e^{-\lambda_s |a_m|^2}}. \quad (1.4)$$

Here, $\lambda_s \geq 0$ is a shaping parameter. An optimal λ_s has to be chosen for a given signal-to-noise ratio (SNR) of the channel. The number of bits represented by each symbol is measured by the entropy of the modulation format $\mathbf{H} = -\sum_{a_m \in \mathcal{A}} P_A(a_m) \log_2 P_A(a_m)$.

In PCS implementations, a distribution matcher is required in the transmitter DSP. Distribution matchers transform independently Bernoulli distributed input bits into output symbols with a desired distribution [46]. This transformation is reversed by a reverse DM at the receiver DSP [43], [47].

PULSE SHAPING

After the bits to symbol mapping, a pulse shaping filter with impulse response $g_{\text{tx}}(n)$ is applied to the sequence of symbols a_k . Here, $k, n \in \mathbb{Z}$ denote symbol index and discrete time sample, respectively. The pulse-shaped signal is given by

$$u(n) = \sum_{k=-\infty}^{\infty} a_k g_{\text{tx}}(n - kT). \quad (1.5)$$

Pulse shaping keeps the signal within the allowed transmission bandwidth. Raised cosine (RC) and root-raised cosine (RRC) filters are examples of commonly used pulse-shaping filters. An advantage of using a RRC filter for pulse shaping is that another RRC filter $g_{r,x}(n)$ can be used at the receiver to perform matched filtering. A matched filter is the optimum linear filter that maximizes the SNR in the presence of AWGN [48]. The impulse response of the matched filter is the same as that of a transmit pulse-shaping filter. Finally, an additional step of digital pre-distortion may be applied to the pulse-shaped signal.

TRANSMITTER DIGITAL PRE-DISTORTION (DPD)

The transmitter hardware components like DACs, DAs, and modulators are not ideal and have undesired responses. For example, the frequency responses of hardware components may not be flat and the gain response of the driver amplifier may be nonlinear. These responses degrade the transmit signal integrity. Therefore, applying a pre-compensation or digital pre-distortion (DPD) filter is often necessary. The DPD filter modifies the pulse-shaped signal such that the transmitter hardware distortions are minimized or even mitigated. A typical example of such a DPD technique is a linear DPD that overcomes the bandwidth limitation of transmitter hardware.

1.3.2. RECEIVER DSP (RX-DSP)

A simple block diagram of key steps in the coherent receiver DSP is shown in Figure 1.7. It takes four sequences of samples from the ADCs as input and produces a sequence of bits at the output. It compensates for several impairments that the signals have acquired in transit. These impairments include transceiver hardware distortion, chromatic, and polarization mode dispersion, clock errors, frequency offset and phase noise. The sequence of processing steps shown in Figure 1.7 is not fixed and often varies depending

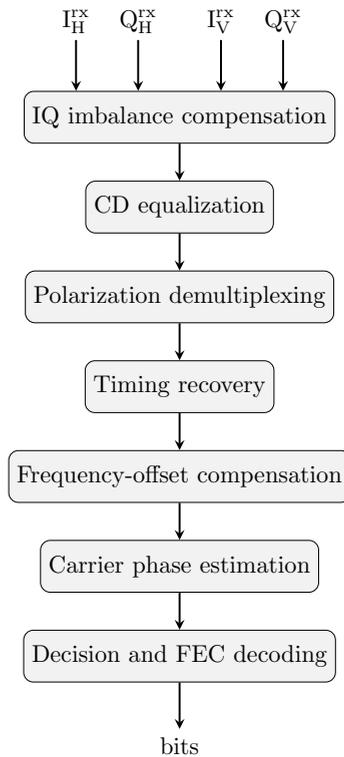


Figure 1.7: A block diagram of key stages in receiver DSP. [IQ crosstalk: in-phase and quadrature-phase crosstalk, CD: chromatic dispersion, FEC: forward error correction]

on the system design. Various DSP algorithms were proposed for the compensation of these impairments. They can be broadly classified into data-aided and non-data-aided (or blind) algorithms. The former requires the use of known training data for learning impairment compensation. Pilot symbols (training data) are transmitted along with information symbols when data-aided receiver DSP algorithms are used. Therefore, the use of data-aided algorithms reduces the effective data rate of the system. However, blind algorithms may have a higher computation and memory complexity.

IQ CORRECTION

A non-ideal optoelectronic front-end of the receiver can introduce gain imbalance and timing misalignment (also known as timing skew) between the I and Q tributaries of the signal. It is also possible that these effects have already been present in the optical signal due to an imperfect modulation at the transmitter. Ideally, the I and Q tributaries should be orthogonal to each other. Under this assumption, the gain imbalance can be compensated by using an orthogonalization process as done in two well-known techniques: Gram-Schmidt and Löwdin algorithms [49]–[51]. IQ time delay or skew correction requires time-shifting and interpolation using a finite impulse response (FIR) filter [52], [53]. It is possible to combine IQ correction with polarization demultiplexing [54], [55].

CD COMPENSATION

Usually, FOCs are operated in a linear regime with loss compensated by EDFAs. In that case, the effect of chromatic dispersion on the pulse propagating through fiber can be described by dropping the nonlinear term in the NLSE (1.1)

$$\frac{\partial Q(\ell, t)}{\partial \ell} = -j \frac{\beta_2}{2} \frac{\partial^2 Q(\ell, t)}{\partial t^2}. \quad (1.6)$$

By taking the Fourier transform with respect to the time coordinate, we get

$$\frac{\partial Q(\ell, \omega)}{\partial \ell} = j \frac{\beta_2}{2} \omega^2 Q(\ell, \omega). \quad (1.7)$$

Here, ω is the angular frequency of the optical carrier. With further simplification, the evolution of the Fourier transform of the input pulse is given as

$$Q(\ell, \omega) = Q(0, \omega) e^{j\beta_2 \omega^2 \ell / 2}, \quad (1.8)$$

which represents a delay in different frequency components of the pulse. The chromatic dispersion (CD) can be compensated by undoing these delays, by using an all-pass filter with a transfer function given by

$$G(L, \omega) = e^{-j\beta_2 \omega^2 L / 2}, \quad (1.9)$$

where L is the length of the fiber.

CD effect is largely time-invariant and has a long memory. The CD compensation can be implemented in the time or frequency domain [56].

POLARIZATION DEMULTIPLEXING

Optical fibers often have manufacturing defects and suffer from mechanical stress, which can be caused by the cabling process. Due to this, there is a slight difference in the propagation characteristics of lightwaves with different polarization states. For example, the pulse may experience different dispersion in two orthogonal polarizations, known as polarization mode dispersion (PMD). This leads to the convolutive mixing of signals that were modulated in two orthogonal polarizations at the transmitter. After removing the CD, an adaptive multi-input multi-output (MIMO) filter compensates for the polarization-dependent impairments and separates the polarizations. The adaptive filter is usually implemented as 2×2 MIMO FIR filter as shown in Figure 1.8. The MIMO structure contains four FIR sub-filters denoted by $\mathbf{h}_{XX}(n)$, $\mathbf{h}_{YY}(n)$, $\mathbf{h}_{XY}(n)$, and $\mathbf{h}_{YX}(n)$, where the former two compensate for intra-polarization and the latter two for inter-polarization cross-talk. The filter taps are obtained by using an adaptive algorithm, which is often required to work with only the magnitude of symbols as the phase is still impaired at this stage.

As an example of polarization demultiplexing, the constant modulus algorithm (CMA) [57], [58] is explained here. It is well-known that when two uniformly distributed signals are mixed, the distribution of the mixed signal tends to be Gaussian [59], [60]. This means the kurtosises of mixed signals are higher than the kurtosises of the signals before mixing. The CMA attempts to find filter taps that minimize the kurtosises of the signals at its output. The CMA adapts the filter coefficients $\mathbf{h}_{XX}(n)$, $\mathbf{h}_{YY}(n)$, $\mathbf{h}_{XY}(n)$ and $\mathbf{h}_{YX}(n)$ to minimize the error ϵ_X and ϵ_Y given by

$$\epsilon_X = R_2 - |x_o|^2, \quad \epsilon_Y = R_2 - |y_o|^2, \tag{1.10}$$

in a mean square sense. Here, R_2 is a real-valued parameter that depends on the modulation format of the transmit signal. It can be calculated using $R_2 = \mathbb{E}\{|a_k|^4\} / \mathbb{E}\{|a_k|^2\}^2$, where a_k represent the transmit symbols. The outputs of the CMA at time-index n are $x_o(n)$ and $y_o(n)$. If N is the memory of the MIMO filter, then for length N sequences of

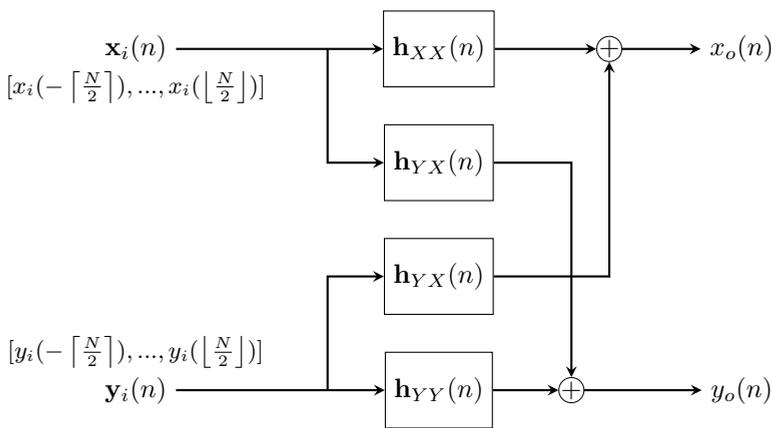


Figure 1.8: A schematic of an adaptive MIMO filter used for polarization demultiplexing.

input symbols $\mathbf{x}_i(n)$ and $\mathbf{y}_i(n)$ the output is given by [61]

$$x_o(n) = \mathbf{h}_X(n)\mathbf{u}_i(n), \quad y_o(n) = \mathbf{h}_Y(n)\mathbf{u}_i(n). \quad (1.11)$$

Here,

$$\mathbf{u}_i(n) = \begin{bmatrix} \mathbf{x}_i(n) \\ \mathbf{y}_i(n) \end{bmatrix}, \quad \mathbf{h}_X(n) = \begin{bmatrix} \mathbf{h}_{XX}(n) \\ \mathbf{h}_{XY}(n) \end{bmatrix}, \quad \text{and} \quad \mathbf{h}_Y(n) = \begin{bmatrix} \mathbf{h}_{YX}(n) \\ \mathbf{h}_{YY}(n) \end{bmatrix}.$$

Then, the filter coefficients are updated using the following rule

$$\mathbf{h}_X^\dagger(n+1) = \mathbf{h}_X^\dagger(n) + \mu\epsilon_X x_o(n)\mathbf{u}_i^\dagger[n], \quad \mathbf{h}_Y^\dagger(n+1) = \mathbf{h}_Y^\dagger(n) + \mu\epsilon_X x_o(n)\mathbf{u}_i^\dagger(n). \quad (1.12)$$

Here, μ is the step size and $(.)^\dagger$ denotes complex-conjugate transpose operation. Often, the filter coefficients are initialized with zero except for the center taps $\mathbf{h}_{XX}(0)$ and $\mathbf{h}_{YY}(0)$, which are set to one. CMA can work in a fully blind mode. Although CMA was originally proposed for constant modulus formats such as QPSK (4-QAM), it can still work for the M-QAM format. Furthermore, there are modified versions of CMA such as the multi-modulus algorithm (MMA) that can work for M-QAM formats. Some alternatives to CMA are independent component analysis (ICA) [59], and Stokes-space based [62] algorithms. The underlying principle of these algorithms is also the same as for CMA, which is to apply a MIMO filter on the mixed signals such that the distribution of output signals is non-Gaussian. However, these algorithms do not work properly for PCS-QAM signals. As PCS-QAM signals tend to be Gaussian distributed, these algorithms cannot identify whether the output signals are separated or still mixed. To circumvent this problem, a modification is made to the existing algorithms. In [63], [64], samples of the received signal are separated into two parts based on their magnitude. Only the part corresponding to a lower magnitude is processed, e.g. with a Stokes-space-based algorithm [64] or CMA [63]. This solution with magnitude-based partitioning of the signal into two parts works because the smallest magnitude part in the constellation of a PCS QAM signal are QPSK symbols. They are uniformly distributed. A conventional algorithm like CMA can be applied to them to estimate the MIMO filter coefficients. However, these modified algorithms require more iterations to converge because many samples belonging to the higher magnitude part are discarded. Furthermore, the mixing matrix can assign some signal samples to an incorrect category.

CLOCK RECOVERY

The sampling frequency of analog to digital converters (ADCs) often varies over time, for example, due to temperature change. Due to sampling clock drift, the detected waveform is then not sampled at uniform intervals. The clock recovery algorithms digitally align the sampling instants of the ADCs with the time best for signal detection. A matched filter $g_{r_x}(n)$ is applied before the clock recovery. Some clock recovery algorithms are the Gardner [65], Mueller-Müller [66], and Oerder and Meyr's algorithms [67]. We explain the working of Oerder and Meyr's algorithm here. In this algorithm, a magnitude square operation is applied to the signal sequence. Then, the resultant sequence is processed in a block-wise manner. By taking the Fourier transform of each block, the phase at the frequency equal to the symbol rate is extracted. In the absence of a timing error, this phase should not drift between the blocks. Oerder and Meyr's algorithm estimates the timing

errors from the computed phase drifts. Let us assume that the signal propagated through a linear fiber channel and, an ideal CD, polarization demultiplexing and matched filter are already applied. The timing errors caused by imperfect sampling of ADCs can be equivalently described as if the sampling was ideal but the continuous time signal already had timing errors. This continuous-time signal is given by

$$x(t) = \sum_{k=-\infty}^{\infty} a_k g_{\text{tx}}(t - kT - \epsilon(t)T) + \tilde{n}(t). \quad (1.13)$$

Here, a_k denotes transmit symbols with $k \in \mathbb{Z}$. The transmit pulse shape and symbol time period are $g_{\text{tx}}(t)$ and T , respectively. The timing error is $\epsilon(t)T$, which varies slowly with time t and is assumed to be constant over L symbols. Channel noise is denoted by $\tilde{n}(t)$.

The continuous-time signal after matched filtering is denoted by $\tilde{x}(t) = x(t) \star g_{\text{rx}}(t)$. It is given by (1.13) but with $g(t) = g_{\text{tx}}(t) \star g_{\text{rx}}(t)$ and $\tilde{n}(t) = g_{\text{rx}}(t) \star \tilde{n}(t)$ replacing $g_{\text{tx}}(t)$ and $\tilde{n}(t)$, respectively. Here, \star denotes the convolution operation. The sampled signal, with sampling rate $\frac{N}{T} > \frac{2}{T}$ is

$$\begin{aligned} \tilde{x}_p &= \tilde{x}\left(p \frac{T}{N}\right) \\ &= \sum_{k=-\infty}^{\infty} a_k g\left(\frac{pT}{N} - kT - \epsilon T\right) + \tilde{n}\left(\frac{pT}{N}\right), \end{aligned} \quad (1.14)$$

with each symbol represented by N samples. The phase at the frequency $\frac{1}{T}$ in the spectrum of $|\tilde{x}_p|^2$ is used to compute the timing errors. With each block having L symbols, the timing error of m^{th} block is computed using

$$\begin{aligned} X_m &= \sum_{k=mLN}^{(m+1)LN-1} |\tilde{x}_k|^2 \exp\left(-j2\pi \frac{k}{N}\right) \\ \hat{\epsilon}_m &= -\frac{1}{2\pi} \arg(X_m). \end{aligned} \quad (1.15)$$

The timing error can be corrected by finding a correct sampling instant using (1.14) and (1.15) and then by computing the sample value at that instant using interpolation of neighboring samples. After timing error correction, the received signal sequence is downsampled to 1 sample per symbol.

FREQUENCY OFFSET COMPENSATION

The frequency of the transmit laser and local oscillator often drifts over time. A frequency detuning (offset) Δf between these lasers adds a constant phase increment of $\phi_{\text{fo}} = 2\pi\Delta f T$ in the symbol sequence, where T is the symbol duration. A transmit symbol a_k corrupted with constant phase increment of ϕ_{fo} due to the frequency offset and phase noise $\theta(n)$ at time-index n is given by

$$x_i(n) = a_k e^{j(\theta(n) + \phi_{\text{fo}}n)}. \quad (1.16)$$

There exist many algorithms for frequency offset compensation. A simple time-domain algorithm suitable to QPSK signals works by raising the signal to the fourth power [68]

$$x_i^4(n) = \left(a_k e^{j(\theta(n) + \phi_{f_0} n)} \right)^4. \quad (1.17)$$

Here, $a_k = e^{jk\frac{\pi}{4}}$, $k \in \{1, 3, 5, 7\}$ denotes QPSK symbols. Raising it to the fourth power, $a_k(n)^4 = e^{jk\pi} = e^{j\pi}$, removes any phase changes introduced by the data modulation. Thus, the only phase variation remaining in $x_i^4(n)$ is due to the frequency offset and phase noise:

$$x_i^4(n) = \left(e^{j\pi} e^{j4(\theta(n) + \phi_{f_0} n)} \right). \quad (1.18)$$

The phase noise $\theta(n)$ is a slow process, i.e. $\theta(n+1) \approx \theta(n)$ as symbol time periods are in picoseconds. An average phase increment over L successive samples of $x_i^4(n)$ can be used to extract frequency offset

$$\hat{\Delta f} = \frac{1}{4 \times 2\pi TL} \sum_1^L \arg(x_i^4(n+1) (x_i^4(n))^*). \quad (1.19)$$

The frequency offset can be removed by applying phase shift $-2\pi\hat{\Delta f}Tn$ in the sequence $x_i(n)$. It is also possible to extend the above methods for higher-order M-QAM formats for e.g. by using only the outermost symbols for estimation. Another way of FO estimation is to detect the frequency at which the Fourier spectrum of $x_i^4(n)$ peaks. This frequency is $4 \times \Delta f$ as can be understood from (1.18). The compensation of FO in PCS-QAM signals is challenging for conventional blind algorithms. This is due to the fact that PCS QAM signals contain a lower number of outer symbols. Equivalently, in the spectrum of $x_i^4(n)$, the peak is not identified accurately. In [69], two modified algorithms are proposed for PCS-M-QAM signals.

CARRIER PHASE ESTIMATION

The output of an ideal oscillator has a constant phase change over a given interval of time, i.e., the output is perfectly periodic. However, the lasers used at the transmitter and the receiver are not ideal. Their phase transition over a given interval of time is random due to reasons like spontaneous emission and temperature drifts. This effect is observed as a non-zero linewidth (spectral width) in the power spectral density of their output. The random phase transitions lead to phase noise $\theta(n)$ in the modulated signals. Symbols corrupted with phase noise can be expressed by (1.16) with $\phi_{f_0} = 0$. Phase noise is a Wiener process [70] and can be compensated using algorithms such as Viterbi and Viterbi [71], blind phase search [72], or pilot-added [73] algorithms.

The Viterbi and Viterbi algorithm works for QPSK-format signals by raising the signal to the fourth power to remove the data modulation as explained for frequency compensation (see (1.18)). In the absence of FO, the only phase variation left in the sequence $x_i^4(n)$ is due to phase noise $\theta(n)$. A moving average window can be applied in order to reduce the impact of noise in the estimation of the phase noise. After estimation, the phase noise can be removed from $x_i(n)$ to get the estimated symbols \hat{a}_k . The Viterbi and Viterbi algorithm can be adapted to other M-QAM formats [74].

DECISION AND FEC DECODING

In this stage, the bits \hat{b}_k are decided from the symbols \hat{a}_k using either a soft-decision or a hard-decision approach. In hard-decision, a series of hard-decision thresholds can be applied on in-phase and quadrature-phase components separately. This kind of implementation is easy to design with logic circuits and outputs a bit sequence that can be fed to hard-decision FEC decoding. In soft-decision implementations, instead of a bit-sequence, the log-likelihood ratios (LLRs) are fed to the FEC decoding stage. A LLR is the logarithm of the ratio of likelihoods that a 1-bit was transmitted rather than a 0-bit. The output of the FEC decoding stage is the received bit sequence. In an ideal scenario, i.e., in the absence of noise and impairments, the transmit and received bit-sequences match.

1.4. ADVANCED TOPICS IN DSP OF COHERENT FIBER-OPTIC COMMUNICATION SYSTEMS

The use of DSP together with coherent architecture has enabled many-fold increases in the information rates of FOCSs in the past. It allowed digital compensation of many physical impairments and, thus, simplified complexity of FOCSs. However, the growth in global IP traffic continues exponentially as new bandwidth-hungry services like video-on-demand and cloud computing become common. Furthermore, remaining impairments that were not significant previously and that were therefore left uncompensated have become non-negligible as higher data rates are to be extracted from communication systems. One of them is the Kerr-nonlinear effect, shown as a bottleneck in increasing the data rates [75] around 10 years ago. In the absence of Kerr-nonlinearity, a higher information rate can be achieved simply by increasing the transmit signal power. It should follow a well-known relationship proposed by Claude E. Shannon, $C = B \log_2(1 + \text{SNR})$ [76]. Here, B is the bandwidth of the signal and C is the information rate. The interplay of the Kerr-nonlinear effect and the dispersion acting simultaneously on the signal, adds noise to both the magnitude and phase of the signal. These nonlinear distortions grow with transmit power therefore reduce data rates. A bell-curve-shaped relation exists between the data rate and the signal power. The best performance is obtained at optimum power.

In the last decade, various techniques have been proposed to mitigate these nonlinear impairments using DSP, such as digital back-propagation[77], Volterra series-based [78], and perturbation theory-based nonlinearity equalizers [79], [80]. However, the complexity of nonlinearity compensation algorithms makes it difficult to accommodate in a real-time DSP. Recently, nonlinear Fourier transform (NFT)-based transmission techniques were proposed as a novel approach to mitigate Kerr-nonlinearity [81]–[83]. However, the impact of noise on NFT-based systems has been shown to be significant [84]. Deep learning-based solutions are another new approach that is also being explored for Kerr nonlinearity compensation. They can offer lower computational complexity than a standard digital back-propagation algorithm [85], [86].

Another impairment that became non-negligible is transceiver hardware distortions. They dominate especially in short links like data center interconnects, where links with a very high data rate and short transmission distance of around 100 km are needed. A non-

linear digital pre-distortion (DPD) technique is required in the transmitter DSP to improve signal quality by mitigating transceiver hardware distortions [87]. Deep learning-based solutions are also being explored for DPD applications [88]–[90].

Furthermore, new advancements in DSP are also being directed to the next generation of coherent FOCs. These systems are moving toward symbol rates beyond 100 Giga-symbol per sec to reduce the cost per bit. A single transceiver pair operating at a high symbol rate can utilize a large bandwidth effectively by reducing the number of guard bands. Furthermore, the number of transceiver pairs needed to utilize the available bandwidth (e.g. ~ 4 THz for the C-band) also decreases, directly reducing the cost. In these systems, DSP faces different challenges. Equalization enhanced phase noise (EPPN) is an impairment that has been investigated a lot recently [91]. It is an amplitude noise generated by the interplay of the receiver laser (LO) phase noise and electronic CD equalization. As the CD effect grows quadratically with signal bandwidth (see (1.8)), a large-memory electronic CD equalization is needed in the receiver's DSP for high symbol rate systems. Thus, electronic CD equalization converts the LO phase noise to amplitude noise and distributes it over a much larger block of symbols, an effect difficult to compensate for. A technique proposed to minimize EPPN is digital subcarrier multiplexing (DSCM) [92], [93]. DSCM transceivers generate a large bandwidth signal composed of many subcarriers of low symbol rates multiplexed digitally. It is also shown to be beneficial in point-to-multi-point transmission scenarios [94]. However, DSP of DSCM systems suffer from other challenges. For example, IQ distortions in DSCM system are more detrimental compared to single carrier systems [95]. Finally, the DSP chain of space division multiplexing (SDM) systems requires an adaptation to account for interactions in the higher dimensional signal space because the impairments can spread across signals propagating through various modes or cores depending on the type of fiber and coupling [39].

Apart from increasing rates, DSP is also enabling FOCs to dynamically adapt to traffic demands and channel conditions. To achieve the maximum data transmission rate for a given channel condition, probabilistic shaping was proposed [42]. New commercial transceivers already use probabilistic shaping [93]. Yet the impact of PCS on crucial DSP stages such as clock recovery and polarization demultiplexing is still being explored. New advanced algorithms are being designed for PCS-QAM signals [64], [69], [96]–[98]. A crucial challenge that needs to be addressed with PCS is that FEC and PCS, both DSP stages operate on bit sequences and therefore can interfere with each other. Different kinds of PCS implementation methods are also under research [47], [99].

In this thesis, we focus on the DSP of coherent systems albeit that DSP is actively used in intensity modulation (IM) direct detection (DD) [100], [101] systems as well. Deep learning-based solutions for IM-DD systems are also being examined [102], [103]. DSP is enabling new architectures that are neither coherent nor IMDD. These architectures can be referred to as semi-coherent. Semi-coherent systems can recover phase or difference between phases of two polarizations, from intensity-only detection. They do not require a LO laser at the receiver, which reduces their cost. Like coherent systems, these systems can support the modulation of phase and amplitude. Many semi-coherent schemes were investigated recently which include Kramers-Kronig receiver [104], Stokes-vector receiver, [105], [106] and phase retrieval-based systems [107]. Kramers-Kronig receivers

are based on the concept of minimum phase systems in which the signal phase is related to its intensity using a standard relation. These receivers detect only the intensity of the signal and utilize DSP to recover the phase information. DSP in Stokes vector receivers allow to take the advantage of polarization diversity in DD systems. The phase retrieval-based receivers use DSP to estimate phase profile of the signal from detected intensity profiles [107]. DSP takes a heavy load in these systems. Various DSP techniques are being investigated to simplify hardware complexity and manage various impairments faced by semi-coherent systems.

Besides signal equalization, advanced algorithms and implementation strategies for DSP have been explored to reduce energy consumption [108]–[110]. DSP has a role to play in achieving this objective as it is a rather power-hungry part of coherent receivers. Furthermore, DSP can help in monitoring various link parameters, which can be used to make optical networks flexible and dynamic and, to optimize performance and resource allocation [111]–[113]. These parameters include the longitudinal power profile and strength of nonlinearity. Power profile monitoring helps in locating anomalies in the link without needing dedicated devices [114], while nonlinearity strength monitoring is useful to optimize the transmit power [115]. Live monitoring information is useful to obtain an accurate digital twin of an optical network. The tracking of the state of polarization using coherent receiver DSP can be helpful in the early detection of fiber damage [116]. In this regard, a new role of DSP is also to transform a coherent receiver into a sensor. Many advanced signal processing and machine learning techniques are being applied to achieve this objective [117].

1.5. NONLINEAR FOURIER TRANSFORM FOR FIBER-OPTIC COMMUNICATION SYSTEM

The nonlinear Fourier transform (NFT) is a mathematical tool that linearizes a lossless nonlinear fiber optic channel. The propagation of a pulse through a nonlinear fiber channel is complicated in the time domain due to simultaneously acting dispersive and nonlinear effects. The NFT translates this complicated propagation into the linear evolution of nonlinear spectral components [118]. The NFT is also known as the inverse scattering transform. It was originally developed for solving certain classes of nonlinear partial differential equations (PDEs). One such PDE is the Korteweg-de Vries (KdV) equation that arises in the study of water waves [119]. An attractive feature of such PDEs is particle-like solutions, also known as solitons, that preserve their shape during propagation. Fig. 1.9 shows the propagation of a fundamental soliton of NLSE. Furthermore, when solitons are launched toward each other, they recover their shape afterward as shown in Fig. 1.10. The surprising properties of solitons have attracted many researchers. In [120], a method is presented to solve the initial value problem of the KdV equation.

Later, the theoretical framework of Lax pairs was proposed in [121]. This framework allowed the use of the method in [120] for any nonlinear PDE for which Lax pairs are found. In 1972, a Lax pair for the NLSE was discovered by Zakharov and Shabat [118]. Later, a detailed method to generate integrable equations associated with Lax pairs of a certain structure was developed [122]. The class of nonlinear PDEs solvable by the NFT

is known as integrable equations [123] and includes the KdV, NLSE, modified KdV [124] and sine-Gordon [125] equations.

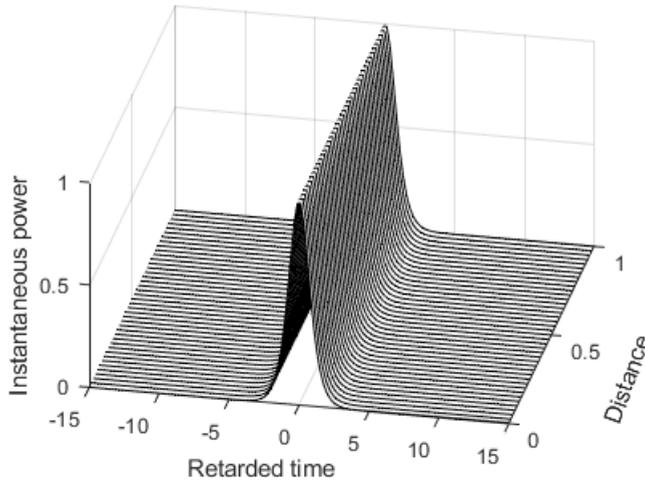


Figure 1.9: A fundamental soliton propagating through an ideal lossless optical fiber preserves its shape.

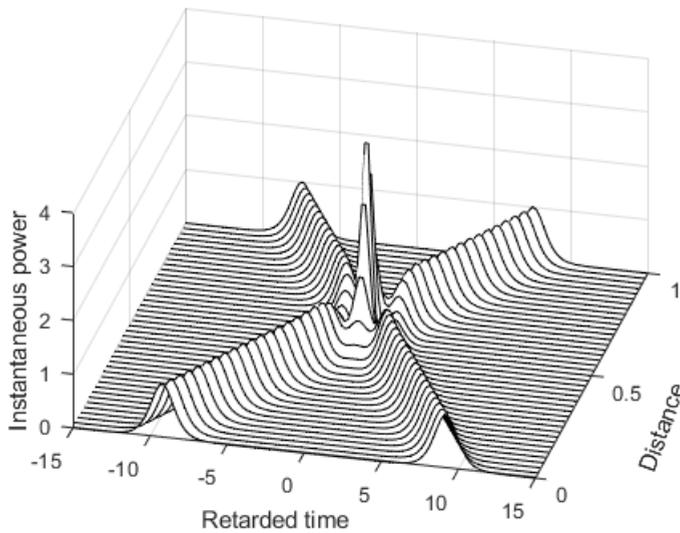


Figure 1.10: Two fundamental solitons propagating towards each other in an ideal lossless optical fiber collide. They recover their shapes after the collision.

1.5.1. SOLITON COMMUNICATION

NLSE solitons preserve their shape as the Kerr nonlinear effect counterbalances the pulse-broadening due to chromatic dispersion. They are immune to pulse-broadening effects, which makes them interesting as carriers of information. Therefore, soliton-based FOCSS were investigated as early as 1973 by Hasegawa and Tappert [126]. The soliton-based transmission was extensively investigated in the 1990s to increase transmission range and minimize the need for re-generators. However, the concept of solitons requires an ideal lossless fiber, while the real fibers are lossy. Although the power loss in a span can be compensated using EDFAs, a decrease in the power of the signal along the span also reduces the strength of the Kerr nonlinear effect. This eventually disturbs the counterbalancing between dispersion and the nonlinear effect. One of the solutions proposed for this problem was to decrease the dispersion [127]–[129]. Another solution was to use the path-average approximation together with low-dispersion fibers [130]. The amplification using an EDFA adds amplified stimulated emission (ASE) noise, which causes random shifts in the carrier frequency of solitons. These shifts mean changes in velocity that translate into jitter in the arrival times of solitons. This effect is known as the Gordon-Haus effect [131]. Periodic narrow-band optical filtering of solitons was shown to minimize this jitter [132]. However, the jitter increases with the propagation distance [132] as the ASE noise gets accumulated in the pass-band of filters over the spans. To overcome this, a novel link design was proposed where the center frequency of the periodic filters was shifted gradually over the spans. This gradual filtering does not impact solitons as they tend to recover their shape. They pass through the link while ASE noise is blocked. This kind of link is surprisingly interesting as it is opaque to noise, yet at the same time, it is transparent to solitons [133]. In a WDM configuration, solitons of different wavelengths travel with different velocities and collide with each other. In an ideal fiber, these collisions cause small time-displacements. In a real link with lossy fiber and lumped amplification, some collisions lead to unacceptably large time-displacements. A way to tackle such challenges requires the use of dispersion management in the link, i.e., the use of dispersion compensating fibers. However, as mentioned earlier, this solution was not found favorable. Therefore, the research in solitons-based transmission systems slowed down.

In recent years, fiber nonlinearity has become an obstacle in increasing the information rates of FOCSS. To address it, the idea of NFT-based transmission was revived by Yousefi et al. [81], [134] which goes beyond the simplest case of soliton-based transmission. This idea leveraged digital coherent technology, high-speed DACs, and fast NFT computation algorithms.

Consider the propagation of an optical pulse through an ideal optical fiber, which is given by the lossless NLSE, i.e. (1.1) with $\alpha = 0$,

$$\frac{\partial Q(\ell, t)}{\partial \ell} + j \frac{\beta_2}{2} \frac{\partial^2 Q(\ell, t)}{\partial t^2} - j\gamma |Q(\ell, t)|^2 Q(\ell, t) = 0. \quad (1.20)$$

Here, we consider $\beta_2 < 0$, i.e., the anomalous dispersion regime. Zakharov and Shabat [118] found that the above equation can be solved by the so-called nonlinear Fourier transform (NFT). To understand NFT for this case, a normalized form of equation (1.20)

is often used

$$\frac{\partial u(z, \tau)}{\partial z} - j \frac{1}{2} \frac{\partial^2 u(z, \tau)}{\partial \tau^2} - j |u(z, \tau)|^2 u(z, \tau) = 0. \quad (1.21)$$

Here, $u(z, \tau)$ is a scaled $Q(\ell, t)$, z is the normalized distance, and τ is normalized time. These quantities are given by

$$u(z, \tau) = \frac{Q(\ell, t)}{\sqrt{P_0}}, \quad z = \frac{\ell}{L_D}, \quad \tau = \frac{t}{T_0}. \quad (1.22)$$

Here, $L_D = T_0^2 / |\beta_2|$ is the dispersion length and $P_0 = 1 / (\gamma L_D)$ is the peak power of the pulse. The parameter T_0 governs the pulse width. Note that, L_D, P_0 and T_0 are intertwined, i.e., changing any one of them changes the remaining two.

The NFT of a vanishing signal $u(z, \tau)$ with respect to equation (1.21) is given by solving the so-called Zakharov-Shabat scattering problem [118]

$$\frac{\partial}{\partial \tau} \begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} = \begin{pmatrix} -j\lambda & u(z, \tau) \\ -u^*(z, \tau) & j\lambda \end{pmatrix} \begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} \quad (1.23)$$

with the boundary condition

$$\begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \exp(-j\lambda\tau) \text{ for } \tau \rightarrow -\infty. \quad (1.24)$$

The Jost scattering coefficients are defined as

$$\begin{aligned} a(\lambda, z) &= \lim_{\tau \rightarrow +\infty} \vartheta_1(z, \tau) \exp(j\lambda\tau), \\ b(\lambda, z) &= \lim_{\tau \rightarrow +\infty} \vartheta_2(z, \tau) \exp(-j\lambda\tau). \end{aligned} \quad (1.25)$$

The NFT decomposes a signal $u(z, \tau)$, for a fixed z , into a nonlinear spectrum consisting of two parts: a continuous and a discrete spectrum.

CONTINUOUS SPECTRUM

The continuous nonlinear spectrum consists of continuous spectral functions representing the radiative or dispersive components of the signal $\rho_c(\xi) = b(\xi) / a(\xi)$ for $\xi \in \mathbb{R}$. Here, ξ denotes λ for the continuous spectrum.

DISCRETE SPECTRUM

The discrete nonlinear spectrum consists of solitonic or non-dispersive components of the signal. It contains a set of isolated points in the upper half of the complex plane called eigenvalues λ_j and corresponding discrete spectrum residues $\rho_{d,j} := b(\lambda_j) / \frac{da}{d\lambda}(\lambda_j)$. The eigenvalues are the zeros of $a(\lambda_j, z)$ with respect to λ_j .

The evolution of the signal in the nonlinear Fourier domain with respect to the standard NLSE (1.21) is given by [118]

$$\begin{aligned} \rho_c(\xi, z) &= \rho_c(\xi, 0) e^{2i\xi^2 z}, \\ \rho_{d,j}(\lambda_j, z) &= \rho_{d,j}(\lambda_j, 0) e^{2i\lambda_j^2 z}, \\ \lambda_j(z) &= \lambda_j(0). \end{aligned} \quad (1.26)$$

Just as dispersion effects are translated into simple phase rotations in the ordinary Fourier domain (1.8), (1.26) shows that simultaneously acting dispersive and nonlinear effects are translated into trivial phase rotations in the nonlinear Fourier domain. That means fiber-propagation effects can be easily reversed with a simple closed-form formula by back-propagating the received signal in the nonlinear Fourier domain.

NONLINEAR FREQUENCY DIVISION MULTIPLEXING SYSTEMS

In the absence of noise, the nonlinear spectral components do not develop any cross-talk among them (1.26). This means if the nonlinear spectral components are used as data carriers, cross-talk compensating techniques will not be required. In addition, data rates will no longer be limited due to fiber channel effects. The technique of using nonlinear spectral components as data carriers is known as nonlinear frequency division multiplexing (NFDM) [134]. The data transmission using NFDM and channel equalization in the nonlinear Fourier domain is illustrated in Fig. 1.11. There are several possibilities to modulate the discrete spectrum, including OOK of eigenvalues [25], [135], modulation of the eigenvalues λ_k [136]–[138], modulation of the discrete spectral residues $\rho_d(\lambda_k)$ [139]–[142], or modulation of the $b(\lambda_k)$ coefficients [143]. For the continuous spectrum, information can be modulated on spectral functions $\rho_c(\xi)$ [144], [145] or the $b(\xi)$ -coefficients [146], [147]. A modulation of the full-spectrum, i.e., both the discrete and continuous spectrum, is also possible [148]–[150].

In discrete spectrum modulation, multi-soliton pulses were optimized for criteria such as time-bandwidth product and perturbation due to lossy propagation [141]. In later works, the phases of the discrete spectrum residues of multiple eigenvalues were modulated [142], [151], [152]. In [143], it was demonstrated that the modulation of $b(\lambda_k)$ coefficients is more robust to noise compared to the modulation of the discrete spectrum residues. To achieve higher spectral efficiency in discrete spectrum modulated NFDM systems, evolution of the time-bandwidth product of solitons was studied in [153] and joint modulation of eigenvalue and, amplitude and phase of the discrete spectrum residues were investigated [154]. The authors of [155] studied the evolution of time and bandwidth of the 2nd and 3rd order soliton pulses during propagation. They numerically searched for optimal pulses that produce a small time-bandwidth product by evaluating various eigenvalues and phases of discrete spectral residues. They showed that the optimal pulses look similar to a train of first-order solitons. It was found that modulating the real part of eigenvalues instead of the imaginary parts is more bandwidth-efficient [154]. However, changing the real part changes the speed of solitons and may cause overlapping. Insertion of guard times is one way to avoid such overlapping. In [156], an alternative method was presented where the real parts of eigenvalues were carefully modulated such that soliton pulses collide during propagation, however, at the receiver they are well-separated without overlapping. This design improved the transmission bandwidth and distance.

For continuous spectrum modulation, an orthogonal frequency division multiplexing (OFDM)-like scheme was considered in [145], [157], [158] where the ρ coefficients were modulated. The continuous spectrum NFDM systems discussed in this thesis use the NFT with vanishing boundary conditions. In order to satisfy these boundary conditions, each signal pulse needs to have a large time interval around it such that its energy becomes sufficiently small. Furthermore, the pulses broaden during propagation and

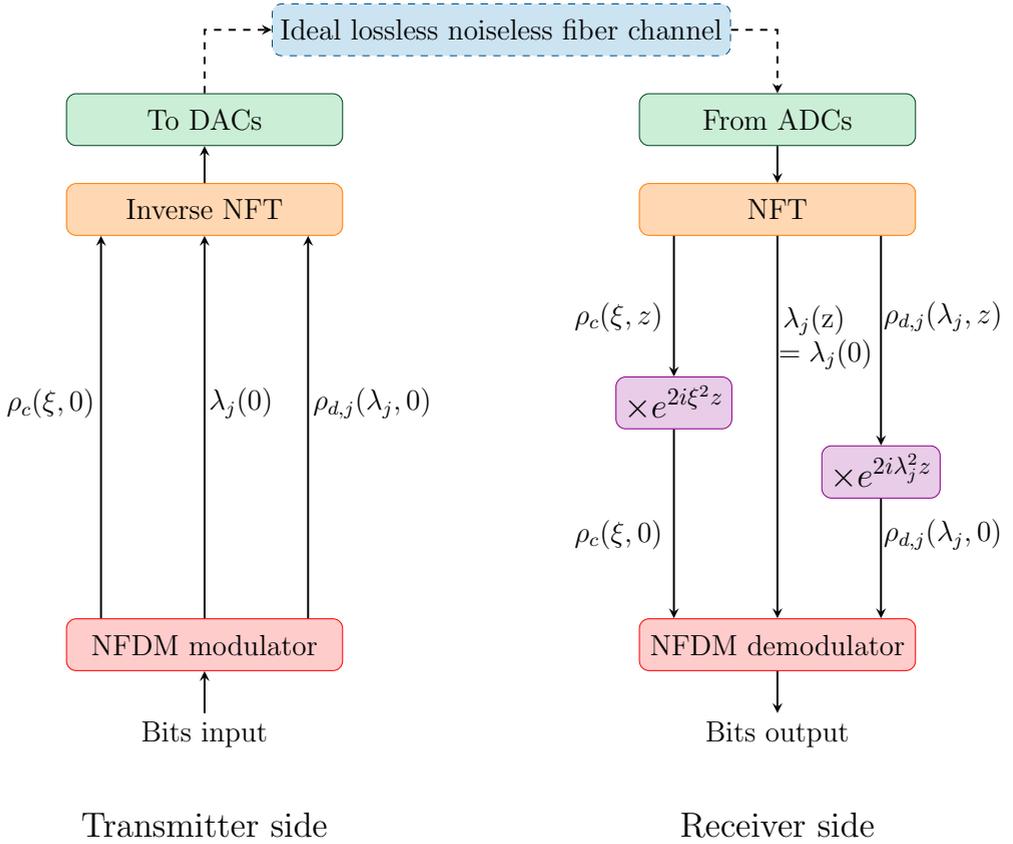


Figure 1.11: A simplified schematic of data transmission using nonlinear frequency division multiplexing (NFDN) technique. [ADCs: analog-to-digital converters, DACs: digital-to-analog converters, NFT: nonlinear Fourier transform]

can overlap, which may lead to interference. Therefore, a guard time interval between each pulse is necessary. This kind of transmission is spectrally inefficient if guard intervals are inserted frequently. To achieve a larger SE, the long decaying tails of each pulse should be reduced. However, a premature truncation of these tails causes performance degradation. To overcome this problem, the modulation of the b -coefficients was suggested for the continuous spectrum in [159]. A b -modulation method was designed by using flat-top carriers and constellation shaping in [160], providing better control over signal power and avoiding the generation of weak carriers. In [150], [161], modulation of both the continuous and discrete spectra was presented.

NFDN systems for dual-polarized transmission i.e. Manakov equations were presented in [162]–[165]. To increase SE, many avenues were explored. For example, in [144], channel pre-compensation was applied to reduce the guard intervals. In [166], b -modulation of the discrete spectrum over a record of thirteen eigenvalues was presented. It was shown recently that multi-soliton pulses can be designed to decay fast by

adding a controlled continuous spectrum to it. This method was shown to reduce the time-support of multi-soliton pulses, improving their SE by 40% in a 2 soliton-OOK system [167]. Recently, a method to jointly modulate the discrete and continuous parts of the nonlinear spectrum under finite time duration constraint was presented [168]. The authors of [147] used Hermite-Gaussian functions-based carriers that could produce a lower time-bandwidth products, for continuous spectrum modulation. They demonstrated the highest SE of 12 bits/sec/Hz so far for a dual-polarization NFDM system in simulations.

Several works studied the impact of practical impairments on NFDM systems. The authors of [169], [170] studied the impact of laser phase noise on a NFDM system that modulates $b(\lambda_k)$ coefficients of the discrete spectrum and found that phase noise is not a dominant factor that can limit the performance when lasers with 100 kHz linewidth are used. NFDM systems are based on the NFT for the lossless fiber model, but fibers have loss in practice. An approximation, known as the path-average model [130], was proposed for designing NFDM systems with lossy fibers in [171], [172]. The impact of fiber-loss on NFDM systems was studied in [141]. In [173], authors studied the impact of Raman effects, third-order dispersion and fiber-loss that are usually not accounted for in the analysis of NFDM systems. The authors concluded that fiber-loss and amplification noise are much more dominant impairments.

1.6. DEEP LEARNING FOR DSP IN FOCS

Machine learning (ML) is a field of study that involves training algorithms and building models to make predictions or decisions based on data. They enable machines to make intelligent decisions on specific tasks without being explicitly programmed for those tasks, instead by relying on experience and training examples. Deep learning is a subset of ML that uses many layers of linear and nonlinear operations on raw input data to progressively learn higher-level representations. DL methods are applied in many areas like pattern recognition, computer vision, prediction classification, and natural language processing. Although ML methods have been around since the 1950s [174], [175], in recent decades they became very popular again due to the availability of training data in large volumes and high computing power [176].

The backbone of DL is neural networks (NNs), which are a type of ML model. Many types of NNs are used in DL such as feed-forward (FF) NN, recurrent NN, and convolutional NN. Feed-forward neural networks are also known as multi-layer perceptrons (MLPs). An exemplary architecture of a FFNN is shown in Fig. 1.12. For simplicity, we consider a real-valued NN. It is composed of a layered structure with each layer containing many computational units referred to as neurons [177]. A neuron of a layer takes inputs from the neurons of the immediately preceding layer and performs a linear and nonlinear operation on it. These computations inside a single neuron are graphically shown in Fig. 1.13. Afterwards, the neuron forwards the computed output to the neurons of the immediately following layer. Usually, the first (left-most) and the last (right-most) layers are called the input and output layers, respectively. The layers in between the input and output layers are referred to as hidden layers. Commonly, NNs with more than one hidden layer are classified as DL. The architecture of the FFNN shown in Fig. 1.12 has $L + 1$ layers. The layers are indexed with $l = 0, 1, \dots, L$, i.e., $l = 0$ is the input layer,

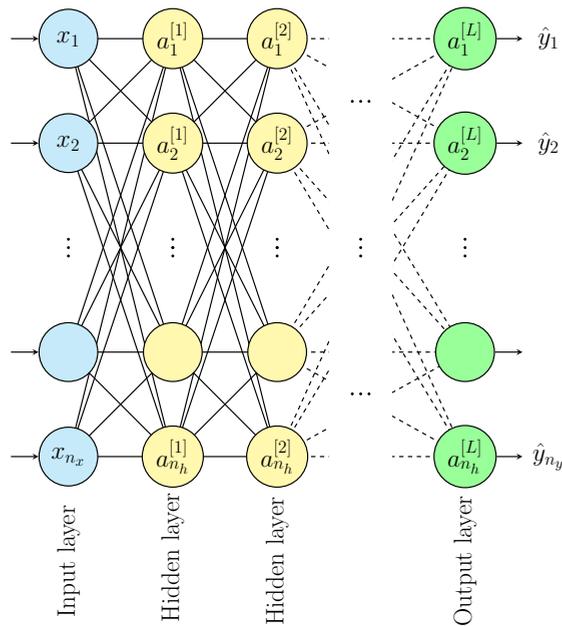


Figure 1.12: An example schematic of a L layer feedforward (FF) neural network (NN). The j^{th} neuron of the l^{th} hidden layer is denoted by $a_j^{[l]}$. The input and output at the j^{th} index are x_j and $\hat{y}_j = a_j^{[L]}$, respectively.

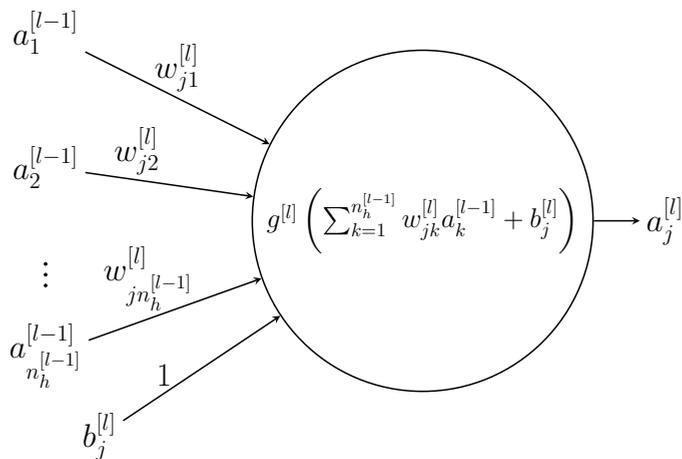


Figure 1.13: An illustration of computations in the j^{th} neuron of the l^{th} hidden layer which is denoted by $a_j^{[l]}$ in Fig. 1.12.

$l = 1, 2, \dots, L - 1$ are the hidden layers and the last layer, $l = L$, is the output layer. A neuron is identified by an index $k = 1, 2, \dots, n_h^{[l-1]}$ with $n_h^{[l-1]}$ being the number of neurons in the $(l - 1)$ th layer. The output of the j th neuron of the l th hidden layer is referred to as an activation and denoted by $a_j^{[l]}$.

A point-wise nonlinear operation $g^{[l]}$, also known as activation function, on

$$z_j^{[l]} = \sum_{k=1}^{n_h^{[l-1]}} w_{jk}^{[l]} a_k^{[l-1]} + b_j^{[l]} \quad (1.27)$$

is applied to compute the activation

$$a_j^{[l]} = g^{[l]}(z_j^{[l]}). \quad (1.28)$$

A typical example of a nonlinear activation function is the rectified linear unit (ReLU) $\text{ReLU}(z) = \max(0, z)$.

The trainable parameters $w^{[l]}$ and $b^{[l]}$ in the above equation are known as the weights and biases of layer l , respectively. At the first hidden layer (i.e., $l = 1$), $z_j^{[1]}$ is computed by using (1.27) with $n_h^{[0]} = n_x$ and $a_k^{[0]} = x^{(i)}$, where $x^{(i)} \in \mathbb{R}^{n_x}$ is the i th training input example to the NN. The output of a L -layer NN $\hat{y}^{(i)} \in \mathbb{R}^{n_y}$ is given by $a_j^{[L]}$. The computation of $\hat{y}^{(i)}$ from $x^{(i)}$ is often called a forward pass through the NN.

The parameters of a NN can be trained using supervised or unsupervised learning. In supervised learning, a labeled data set is provided to the NN. This data set is composed of inputs $\mathbf{X} \in \mathbb{R}^{n_x \times m}$ and corresponding target outputs $\mathbf{Y} \in \mathbb{R}^{n_y \times m}$, where m is the batch size or total number of training examples. Each column in \mathbf{X} and \mathbf{Y} are respectively the input and target output of the NN. The goal of supervised learning is to learn an unknown mapping $\mathbf{Y} = f(\mathbf{X})$. In unsupervised learning the data set is not labeled, i.e., the corresponding target for any input example of \mathbf{X} is unknown. Unsupervised learning can be used to learn latent representations of data by using autoencoder NNs. To train such an autoencoder NN, the input and output data are the same (i.e., $\mathbf{Y} = \mathbf{X}$). The latent representation of input data is an output of some intermediate (hidden) layer that has a lower dimensionality than the input layer. Learning latent representations using autoencoders has several applications such as reducing the number of features input to a NN and learning a mapping from bits to complex-valued symbols over a given communication channel.

The weights and biases are determined using an iterative training process. In each training iteration, a single batch of input examples \mathbf{X} is passed forward to compute $\hat{\mathbf{Y}}$. A cost function is then computed. There are many kinds of cost functions used in DL, usually depending on the intended task. A cross-entropy loss is often used for classification tasks. For regression tasks, the mean square error (MSE)

$$J(\mathbf{W}, \mathbf{B}) = \frac{1}{m} \sum_{i=1}^m \|y^{(i)} - \hat{y}^{(i)}\|^2 \quad (1.29)$$

is often computed. Then, the NN weights \mathbf{W} and biases \mathbf{B} are adapted to minimize this cost function. An optimization algorithm like stochastic gradient descent (SGD) can be

used for this. In SGD, first the gradients of the cost function with respect to each of the weights and biases, i.e. $\frac{\partial J(\mathbf{W}, \mathbf{B})}{\partial \mathbf{W}}$ and $\frac{\partial J(\mathbf{W}, \mathbf{B})}{\partial \mathbf{B}}$ are computed. The chain rule of differentiation is used to compute the gradients of $J(\mathbf{W}, \mathbf{B})$ with respect to each parameter. The process of computing the gradients with respect to each parameter is often called backward pass or back-propagation, because the computation starts from the last layer $l = L$ until it reaches layer $l = 1$. Finally, the parameters are updated using

$$\begin{aligned} w^{[l],\{n+1\}} &= w^{[l],\{n\}} - \eta \frac{\delta J(\mathbf{W}, \mathbf{B})}{\delta w^{[l],\{n\}}}, \\ b^{[l],\{n+1\}} &= b^{[l],\{n\}} - \eta \frac{\delta J(\mathbf{W}, \mathbf{B})}{\delta b^{[l],\{n\}}}. \end{aligned} \quad (1.30)$$

Here, η is the learning rate (or step size) and the superscript $\{n\}$ denotes the n^{th} iteration or the batch index, respectively.

Before the first training iteration, the NN parameters are initialized randomly. Once sufficient accuracy is achieved on the training data, the NN can be used to predict outputs for new unseen inputs. It was shown that NNs are universal approximators, i.e., they can learn any measurable function with any desired degree of accuracy [178].

Neural network models were explored with deeper architectures as well [179]. A deep NN contains many layers. In such a NN, computing the derivatives of the cost function with respect to the parameters of the initial layers requires the chain rule to be applied many times. As a consequence, it is possible that the resulting gradients are too weak to update those parameters. This problem is known as the vanishing gradient problem. To overcome this problem, the residual learning method was introduced [180], [181]. In this method, a shortcut connection that bypasses several layers is added. This shortcut provides an alternate path to compute the gradients for the beginning layers. It reduces the number of times that the chain rule has to be applied and preserves a sufficient magnitude in gradients such that initial layers can be updated. Residual learning is faster when a significant part of the desired output can be produced with a linear mapping of input. Training a large NN with millions of parameters is difficult because a vast amount of memory is needed to store activations that are later used during the back-propagation step. To simplify this, new NN layers known as reversible layers that compute activations on-the-go instead of storing them were proposed in [182], [183]. Deep NN models were shown successful in image classification [176], [179] and other parallel application domains such as object detection and tracking [184], [185], and video classification [186]. The availability of a variety of deep learning (DL) toolboxes [187] provides a head-start for exploring DL for various research problems.

DL methods have also found their place in communication systems [188]–[191]. They were applied in wireless communication for channel estimation and equalization [192] as well as for spectrum allocation [193]. In optical networks, DL techniques were investigated for network optimization, such as resource allocation, traffic routing [189], modulation format identification, multi-impairment monitoring, and performance monitoring [117], [194].

As DSP in coherent FOCS offers a plethora of flexible parameters for optimization, such as baud rate, modulation format, pulse shaping, constellation shaping, and FEC, the search for an optimum configuration for a given link scenario is a complex task and,

it is believed that DL techniques can be useful for such applications. Several applications of DL for the physical layer are presented in [195]. In end-to-end (E2E) learning, the complete communication system is modeled using NNs such that the transmitter and receiver can be jointly optimized for a common metric like SNR. This optimization is different from the conventional design of DSP where each block is responsible and tuned for only a particular task. A step towards E2E learning on FOCs was to model nonlinear fiber channels using NNs [196], [197] and generative models [198], [199]. In later works, E2E learning of autoencoder-based transceivers was presented [199], [200]. Similarly, signal constellations were optimized over the nonlinear channel in an E2E manner [197], [201], [202] [203], [204]. In some works [205]–[208], DSP stages, which are also a part of the E2E channel, were included in the E2E optimization. Besides E2E learning, DL methods for specific DSP tasks were investigated. These include the compensation of impairments caused by, for example, Kerr-nonlinearity and transceiver distortions, the improvement of symbol decision boundaries for nonlinear fiber channels and the computation of the optimum nonlinear demapping of symbols to log-likelihood ratios (LLRs) [189], [191], [209], [210]. FFNNs with input features such as time-domain waveforms or nonlinear Fourier spectral components were also demonstrated for equalization tasks in NFDM systems [211]–[218].

1.7. OUTLINE OF THE DISSERTATION

In this thesis, we address problems in increasing the achievable data rates of the current state-of-the-art fiber-optic communication systems, which are often limited by fiber nonlinearity and transceiver hardware distortions. Furthermore, the benefits of using probabilistically shaped signals to maximize data rates over given channel conditions are challenged because the non-data-aided algorithms of receiver digital signal processing (DSP) do not work well for shaped signals. Therefore, new algorithms are required. We present novel solutions for each of these problems.

CHAPTER 2

We investigated the degrading effects of fiber-loss on NFDM systems. Normally, the loss in real fibers cannot be accounted for completely in NFDM systems. We proposed a novel approach to incorporate fiber-loss into NFDM systems that uses a modified NFT together with specialized dispersion-decreasing fibers (DDFs). In an ideal mathematical scenario, our proposed NFDM system design eliminates degrading effects due to fiber-loss completely. Two of the popular NFDM systems [142], [160] are redesigned using the new method and compared against the approximation-based systems for conventional fibers in numerical simulations.

This chapter is based on the following publication:

V. Bajaj, S. Chimmalgi, V. Aref and S. Wahls, "Exact NFDM Transmission in the Presence of Fiber-Loss". In *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 3051-3058, 1 June, 2020, doi: 10.1109/JLT.2020.2984041.

CHAPTER 3

The mitigation of hardware distortions is critical for short-reach systems such as data-center interconnects. These systems operate transceiver hardware at much faster symbol rates and modulate QAM formats of higher order. Under these conditions, the nonlinear effects of the hardware are more pronounced and complex digital pre-distortion (DPD) techniques are required. We investigate such a 128 G symbol/second PCS-256QAM experimental setup and propose a neural network (NN)-based DPD technique. The NN-based DPD was compared against various popular DPD methods and shown to provide significant gains. A record transmission of 1.61 Terabits/sec/wavelength over 80 km of fiber was achieved experimentally. Furthermore, the computation complexity of the NN-DPD was compared with a Volterra series-based DPD. We significantly reduced the computation complexity of NN-DPD with a minor loss in the performance.

This chapter is based on the following publication:

V. Bajaj, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "Deep Neural Network-Based Digital Pre-distortion for High Baudrate Optical Coherent Transmission". In *Journal of Light-wave Technology*, vol. 40, no. 3, pp. 597-606, 1 Feb. 2022, doi: 10.1109/JLT.2021.3122161.

CHAPTER 4

We modified the NN-based DPD from Chapter 3 to handle IQ crosstalk and evaluated its performance over a 48 km field-deployed fiber in a wavelength division multiplexing (WDM) scenario. A net C-band transmission rate of 54.5 Tb/s was demonstrated.

This chapter is based on the following publication:

V. Bajaj, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion". In *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, San Francisco, CA, USA, 2021, pp. 1-3.

CHAPTER 5

Recently, recurrent NNs have been shown to perform quite well in fiber nonlinearity equalization [219] and E2E transceiver optimization [220]. In Chapter 5, we investigate RNN-based DPD in simulations and compare it against our previously proposed NN-based DPD from Chapter 2. The investigations show that RNN-based DPD achieves benchmark performance but with much higher computational complexity.

This chapter is based on the following publication:

V. Bajaj, V. Aref and S. Wahls, "Performance Analysis of Recurrent Neural Network-based Digital Pre-Distortion for Optical Coherent Transmission". In *2022 European Conference on Optical Communication (ECOC)*, Basel, Switzerland, 2022, pp. 1-4.

CHAPTER 6

Conventional blind polarization demultiplexing algorithms are not suitable for PCS M-QAM signals because the distribution of these signals tends to be Gaussian. It is known that independently and identically distributed (iid) Gaussian signals, when mixed, cannot be recovered/separated from their mixture. In this chapter, we discuss a method wherein we add temporal correlations at the transmitter, which are subsequently exploited at the receiver in order to separate the polarizations. Two algorithms are proposed, one for weak and one for moderate PMD channels, respectively. A numerical proof-of-concept investigation was carried out against a benchmark algorithm.

This chapter is based on the following publication:

V. Bajaj, R. Van de Plas and S. Wahls, "Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations". In *Journal of Lightwave Technology*, vol. 42, no. 2, pp. 560-571, Jan. 2024, doi: 10.1109/JLT.2023.3315370.

2

EXACT NFDM TRANSMISSION IN THE PRESENCE OF FIBER-LOSS

Nonlinear frequency division multiplexing (NFDM) techniques encode information in the so called nonlinear spectrum which is obtained from the nonlinear Fourier transform (NFT) of a signal. NFDM techniques so far have been applied to the nonlinear Schrödinger equation (NLSE) that models signal propagation in a lossless fiber. Conventionally, the true lossy NLSE is approximated by a lossless NLSE using the path-average approach which makes the propagation model suitable for NFDM. The error of the path-average approximation depends strongly on signal power, bandwidth and the span length. It can degrade the performance of NFDM systems and imposes challenges on designing high data rate NFDM systems. Previously, we proposed the idea of using dispersion decreasing fiber (DDF) for NFDM systems. These DDFs can be modeled by a NLSE with varying-parameters that can be solved with a specialized NFT without approximation errors. We have shown in simulations that complete nonlinearity mitigation can be achieved in lossy fibers by designing an NFDM system with DDF if a properly adapted NFT is used. We reported performance gains by avoiding the aforementioned path-average error in an NFDM system by modulating the discrete part of the nonlinear spectrum. In this paper, we extend the proposed idea to the modulation of continuous spectrum. We compare the performance of NFDM systems designed with dispersion decreasing fiber to that of systems designed with a standard fiber with path-average model. Next to the conventional path-average model, we furthermore compare the proposed system with an optimized path-average model in which amplifier locations can be adapted. We quantify the improvement in the performance of NFDM systems that use DDF through numerical simulations.

The text in this chapter has previously appeared in modified form in open access article V. Bajaj, S. Chimmalgi, V. Aref and S. Wahls, "Exact NFDM Transmission in the Presence of Fiber-Loss". In *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 3051-3058, 1 June, 2020, doi: 10.1109/JLT.2020.2984041. [221]. Its reuse is licensed under CC BY 4.0 [222].

2.1. INTRODUCTION

THE ability of the nonlinear Fourier transform (NFT) to linearise the lossless nonlinear fiber-optic channel has attracted much research in recent years. An optical pulse propagates through an ideal optical fiber in a complicated manner as dispersive and nonlinear effects act simultaneously on it. This complicated propagation translates into simple rotations in the nonlinear spectrum [118]. As the nonlinear spectrum evolves linearly through nonlinear lossless fiber, the cross-talk among the spectral components is absent. This has led to the emergence of new types of optical fiber transmission technologies where fiber-nonlinearity is no longer seen as an undesired element. Because of the immunity to nonlinear cross-talk and simple equalization, NFT based transmission techniques are seen as interesting approach to mitigate fiber-nonlinearity [81], [134]. The propagation of a pulse in an ideal optical fiber is modeled by the lossless NLSE, which belongs to a certain class of evolution equations, known as integrable evolution equations. The property of integrability of the lossless NLSE makes it exactly solvable by NFT [118]. The NFT decomposes a signal into a nonlinear spectrum which consists of two parts: continuous and discrete spectrum. The continuous part consists of continuous spectral functions representing the radiative components of the signal. The discrete part consists of a set of isolated points called eigenvalues and their corresponding spectral values. The discrete part represents the solitonic components of the signal. The absence of interference between these nonlinear spectral components (in an ideal fiber) encourages the idea of encoding information on them and new transmission techniques were proposed known as nonlinear frequency division multiplexing (NFDN) [134]. NFDN systems face several challenges such as complicated noise statistics [84], [139], numerical complexity, and the lack of general efficient optical methods to (de-)multiplex several users in the nonlinear frequency domain [223], [224]. Another, fundamental issue is the loss in optical fibers. The fiber-loss breaks the integrability property of NLSE and hence, the NFT is not exactly applicable. Traditionally in NFDN systems, this challenge is addressed with the path-average model [9], [130], where the lossy propagation of the signal is approximated by lossless propagation in another fiber (with a path-averaged nonlinear parameter). Many NFDN transmission results have been demonstrated utilizing this model [82], [141], [142], [145], [225]. An improvement in the accuracy of the path-average model was shown by a shift in the amplifier locations in the link [171], [226]. However, the signal propagation described by the path-average model always deviates from the actual propagation due to the use of approximation. Perfect nonlinearity compensation cannot be achieved even in absence of noise [145]. Further, the approximation errors become stronger when signal power, bandwidth and span length increase [130], [145], [171].

In order to avoid the approximation errors associated with the path-average model, we investigated an idea from classical single soliton systems [227]. In these systems dispersion decreasing fiber (DDF) was introduced to prevent soliton broadening in lossy fiber [127]–[129], [228], [229]. These fibers are made such that the balance between the dispersive and nonlinear effects is preserved along the fiber. Furthermore, it has been shown in [230] that a DDF can be approximated efficiently using a fiber with a stepwise dispersion profile. It was found in the study that an n -fold stepwise dispersion profiling provides an equivalent reduction in the errors due to the path-average approximation as

an n -fold decrease in the span length. Nevertheless, if a fiber with desired dispersion decreasing profile is designed then its propagation model is exactly integrable and hence can be solved exactly with NFTs [231]. Thus in an NFDM system designed with DDF, the nonlinear and dispersive impairments in the signal introduced during the noiseless propagation can be perfectly mitigated even though there is fiber-loss. An NFDM transmission system designed with DDF requires a suitably adapted NFT. We previously demonstrated such an exact NFDM transmission over DDF using the modulation of the discrete part of the nonlinear spectrum [227].

In this paper, we extend our previous results when the data is modulated on the continuous spectrum. Further, we compliment the previous results of discrete spectrum modulation by comparing the performance of an NFDM system designed using DDF to an NFDM system designed using constant dispersion fiber (CDF), where the parameters of the CDF are chosen in the range of standard non-zero dispersion shifted fiber. We furthermore include the performance of an NFDM system designed using CDF when the location of amplifiers are optimized in the transmission-link. The paper is organized as follows. In Sec. 2.2, we review the basics of NFDM and the path-average model. In Sec. 2.3, we discuss the advantages of using DDF in designing NFDM system in the presence of fiber-loss along with the corresponding NFT. Then, in Sec. 2.4, we describe the simulation setups and present results for discrete spectrum and continuous spectrum modulation respectively. The paper is concluded in Sec. 2.5.

2.2. BASICS OF NFDM

The propagation of an optical pulse $Q(\ell, t)$ in an ideal lossless single-mode fiber can be modeled by the NLSE [9, Ch. 2.6.2]

$$\frac{\partial Q}{\partial \ell} + i \frac{\beta_2}{2} \frac{\partial^2 Q}{\partial t^2} - i \gamma |Q|^2 Q = 0, \quad (2.1)$$

where ℓ represents the propagation distance and t is retarded time. The parameters β_2 and γ are the dispersion and nonlinear parameters respectively. Here, we consider the anomalous dispersion case $\beta_2 < 0$. The above equation is integrable and can be solved exactly by NFTs. By a change of variables,

$$u = \frac{Q}{\sqrt{P}}, \quad z = \frac{\ell}{L_D}, \quad \tau = \frac{t}{T_0},$$

where $L_D = \frac{T_0^2}{|\beta_2|}$, $P = \frac{1}{\gamma L_D}$ and T_0 is a free parameter, (1) can be transformed into the normalized form

$$\frac{\partial u}{\partial z} - i \frac{1}{2} \frac{\partial^2 u}{\partial \tau^2} - i |u|^2 u = 0, \quad u = u(z, \tau). \quad (2.2)$$

The NFT of a vanishing signal $u(z, \tau)$ with respect to (2.2) is obtained by the solution of the so-called Zakharov-Shabat scattering problem [118]

$$\frac{\partial}{\partial \tau} \begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} = \begin{pmatrix} -j\lambda & u(z, \tau) \\ -u^*(z, \tau) & j\lambda \end{pmatrix} \begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} \quad (2.3)$$

with the boundary condition

$$\begin{pmatrix} \vartheta_1(z, \tau) \\ \vartheta_2(z, \tau) \end{pmatrix} \rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} \exp(-j\lambda\tau) \text{ for } \tau \rightarrow -\infty. \quad (2.4)$$

The Jost scattering coefficients are defined as

$$\begin{aligned} a(\lambda, z) &= \lim_{\tau \rightarrow +\infty} \vartheta_1(z, \tau) \exp(j\lambda\tau), \\ b(\lambda, z) &= \lim_{\tau \rightarrow +\infty} \vartheta_2(z, \tau) \exp(-j\lambda\tau). \end{aligned} \quad (2.5)$$

The NFT of $u(z, \tau)$, for fixed z , consists of two parts:

CONTINUOUS SPECTRUM

For $\lambda \in \mathbb{R}$, now onward denoted by ξ , consisting of spectral functions $\rho_c(\xi) = b(\xi)/a(\xi)$.

DISCRETE SPECTRUM

For $\lambda \in \mathbb{C}^+$, consisting of eigenvalues λ_j and their corresponding discrete spectral values $\rho_{d,j}$,

$$\left(\lambda_j, \rho_{d,j} := b(\lambda_j) / \frac{da}{d\lambda}(\lambda_j) \right),$$

where the eigenvalues are the zeros of $a(\lambda, z)$ with respect to λ in the complex upper half-plane.

The evolution of the signal in the nonlinear Fourier domain with respect to the standard lossless NLSE (2.2) is given by [118]

$$\begin{aligned} \rho_c(\xi, z) &= \rho_c(\xi, 0) e^{2i\xi^2 z}, \\ \rho_{d,j}(\lambda_j, z) &= \rho_{d,j}(\lambda_j, 0) e^{2i\lambda_j^2 z}, \\ \lambda_j(z) &= \lambda_j(0). \end{aligned} \quad (2.6)$$

Eq. (2.6) implies that the impairments due to the dispersive and nonlinear effects acting simultaneously on a signal can be equalized easily in the nonlinear Fourier domain. Further, the cross-talk due to fiber-propagation is absent among the nonlinear spectral components (2.6). In NFDN techniques, these spectral components are utilized as data-carriers.

So far, we have considered a lossless fiber but real fibers have non-negligible loss, which can be accounted in (2.1) by a loss parameter α [9],

$$\frac{\partial Q}{\partial \ell} + i \frac{\beta_2}{2} \frac{\partial^2 Q}{\partial t^2} - i\gamma |Q|^2 Q = -\frac{\alpha}{2} Q. \quad (2.7)$$

The above equation is not integrable, hence it is not suitable for NFDN. One way to address this issue is by using the path-average model [130]. In the path-average model, the variation in the signal power due to loss is transformed into variations in the nonlinear parameter. Then, by approximating the varying nonlinear parameter with its average value over a span, a lossless fiber model is obtained.

NFDM techniques can be applied to the fiber model obtained using the path-average approach. However, as the model has been derived with an approximation, errors are introduced and complete nonlinearity compensation cannot be achieved even in absence of noise [145]. In the next section, we discuss how by using a suitably designed fiber exact nonlinearity compensation can be achieved even in the presence of loss, assuming that higher order effects in the fiber such as third order dispersion or scattering are negligible.

2.3. DISPERSION DECREASING FIBER AND MODIFIED NFT

In order to preserve the integrability property in presence of loss, a fiber can be designed with varying dispersion and/or nonlinear parameter profile [127]. As these two parameters are linked with the effective core radius of fiber, a fiber with a desired profile can be manufactured practically by tapering the optical fiber during the draw process [9, Ch. 9.3.1],[129], [228].

In this paper, we assume a simplified approximate relation between the effective core radius r (in μm) and the dispersion parameter β_2 that was given in [127],¹

$$r(\beta_2) = (-\beta_2/\kappa + 20)/8, \quad (2.8)$$

where $\kappa = \lambda_0^2/2\pi c \times 10^{-6}$ and λ_0 and c are the wavelength and speed of light in free space respectively. The nonlinear parameter depends on the effective core radius as follows [9, Ch. 2.6.2]¹

$$\gamma = 2n_2/(\lambda_0 r^2), \quad (2.9)$$

where n_2 is the nonlinear-index coefficient. Hence, by controlling the radius r , a fiber with variable dispersion parameter $\beta_2(\ell) = \beta_2(0)D(\ell)$ and a variable nonlinear parameter $\gamma(\ell) = \gamma(0)R(\ell)$ can be designed. The propagation of an optical pulse $Q(\ell, t)$ in such a fiber is then given by [231]

$$\frac{\partial Q}{\partial \ell} + i \frac{\beta_2(0)D(\ell)}{2} \frac{\partial^2 Q}{\partial t^2} - i\gamma(0)R(\ell)|Q|^2 Q = -\frac{\alpha}{2} Q. \quad (2.10)$$

The above equation can be transformed as before, but with $L_D = \frac{T_0^2}{|\beta_2(0)|}$ and $P = \frac{1}{\gamma(0)L_D}$, into the normalized form

$$\frac{\partial q}{\partial z} - i \frac{D(z)}{2} \frac{\partial^2 q}{\partial \tau^2} - iR(z)|q|^2 q = -\frac{\alpha L_D}{2} q. \quad (2.11)$$

It was shown in [231] that the above equation can be solved exactly via NFT if

$$\alpha L_D = -\frac{R(z)D'(z) - R'(z)D(z)}{R(z)D(z)}, \quad (2.12)$$

where the prime denotes differentiation.

¹Equation (2.15) above differs slightly from the results reported in [227] because there were typos in the equations (2), (3) and (8) of [227].

The NFT of $q(z, \tau)$ with respect to (2.11) is now defined as the conventional NFT of the signal $u(z, \tau) = \sqrt{\frac{R(z)}{D(z)}} q(z, \tau)$. The reader is referred to Appendix 2.5 for a detailed description. The evolution of the signal in the nonlinear Fourier domain with respect to (2.11) is given by

$$\begin{aligned}\rho_c(\xi, z) &= \rho_c(\xi, 0) e^{2i\xi^2 \int_0^z D(\zeta) d\zeta}, \\ \rho_{d,j}(\lambda_j, z) &= \rho_{d,j}(\lambda_j, 0) e^{2i\lambda_j^2 \int_0^z D(\zeta) d\zeta}, \\ \lambda_j(z) &= \lambda_j(0).\end{aligned}\quad (2.13)$$

The above relation enables us to recover the signal impaired simultaneously with loss, dispersion and nonlinearity. In order to satisfy (2.12), the required dispersion profile has to satisfy

$$\frac{\beta_2(\ell)}{\gamma(\ell)} = a e^{-\alpha \ell}, \quad (2.14)$$

where $a = \frac{\beta_2(0)}{\gamma(0)}$. By combining (2.8), (2.9) and (2.14), we arrive at¹

$$8r^3(\ell)\kappa - 20\kappa r^2(\ell) + \frac{2an_2}{\lambda_0} e^{-\alpha \ell} = 0. \quad (2.15)$$

If equation (2.15) is written as

$$c_3 r^3 + c_2 r^2 + c_1 r + c_0 = 0. \quad (2.16)$$

Then for our case ($\kappa > 0$ and $a < 0$), the coefficients are real and $c_3 = 8\kappa > 0$, $c_2 = -20\kappa < 0$, $c_1 = 0$, $c_0 = 2an_2/\lambda_0 < 0$ and discriminant $\Delta = 18c_3c_2c_1c_0 - 4c_2^3c_0 + c_2^2c_1^2 - 4c_3c_1^3 - 27c_3^2c_0^2 < 0$. Thus, the equation will always have one real root and two complex conjugate roots [232]. The real root is positive and corresponds to the radius that satisfies equation (2.15). The real-valued solution at different ℓ provides us the distance dependent effective core radius. Once the radius $r(\ell)$ is known we can find $\beta_2(\ell)$ and $\gamma(\ell)$ from (2.8) and (2.9) respectively.

We chose realistic values of fiber parameters and controlled the effective core radius to achieve the desired profile for fiber-loss of 0.2 dB/km as shown in Fig. 2.1. The value

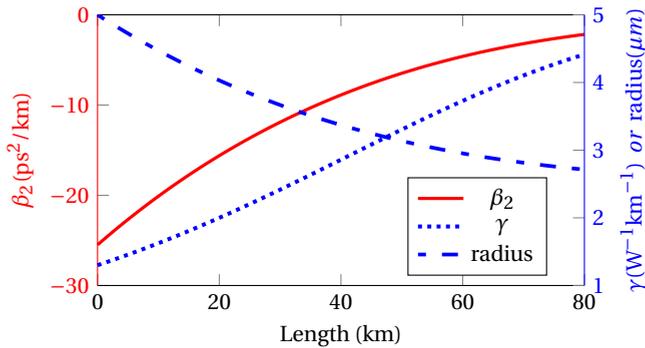


Figure 2.1: DDF parameters for a single span as used in the simulations.

for the nonlinear-index coefficient n_2 and λ_0 was taken as $2.52 \times 10^{-20} \text{ m}^2/\text{W}$ and $1.55 \text{ }\mu\text{m}$ respectively. The dispersion parameter β_2 is varied from $-25 \text{ ps}^2/\text{km}$ to $-2.17 \text{ ps}^2/\text{km}$, while the nonlinear parameter is varied from $1.3 \text{ W}^{-1}\text{km}^{-1}$ to $4.4 \text{ W}^{-1}\text{km}^{-1}$. Here, the effective core radius of fiber is decreased slowly over 80 km length so that the radiation losses are expected to be negligible [233]. Third order dispersion was not implemented since realistic values around $\beta_3 = 0.05 \text{ ps}^3/\text{km}$ are expected to be insignificant in our case due to the high average dispersion parameter and nanosecond pulse durations [233].

It must be noted that the integrability of the channel is still preserved in a link designed using DDF where each span consists of DDF followed by an noiseless amplifier with its gain equal to span-loss. In such a link, at an amplifier not only the gain is added, but also the fiber parameters (β_2, γ) change as a new span starts. In the normalized domain, these two effects cancel each other and the integrability of the channel is preserved.

2.4. SIMULATION SETUP AND RESULTS

In this section, we numerically compare the performance of NFD systems in DDF to that in CDF. The parameters of the DDF are shown in Fig. 2.1 as described earlier. Due to the different fiber characteristics, it is not obvious how fair the two setups can be compared. The parameters of the CDF were chosen in such a way that both systems have the same power, bandwidth and time duration at the transmitter². This requires that the time-scale parameter T_0 and power-scale parameter P , which scales the normalized signal before transmission should be the same for both systems. The obtained CDF parameters are summarized in Table 2.1 and are very close to realistic parameters of standard non-zero dispersion shifted fibers. It is worth mentioning that for the DDF case, the normalization length L_D is smaller and the overall dispersion and nonlinearity is higher. The simulations were carried out with the open-source software environment NFDmlab

Table 2.1: Fiber parameters used in simulations.

Fiber type	CDF	DDF
α (dB/km)	0.2	0.2
β_2 (ps ² /km)	-6.75	-25 to -2.17
γ (1/W/km)	1.3	1.3 to 4.4

[234], which uses FNFT[235], a software library to compute NFTs and INFTs. We compare the performance between the NFD systems designed with the two fibers, CDF and DDF. First we evaluate the NFD systems using discrete spectrum modulation and later using continuous spectrum modulation.

NFDM WITH DISCRETE SPECTRUM MODULATION

We considered the multi-soliton transceiver presented in [142], in which the spectral values $\rho_{d,j}$ of seven eigenvalues λ_j were modulated independently with QPSK. The system parameter were kept the same as presented in [142] unless otherwise stated. The simu-

²It may be interesting to see a comparison of these systems designed in other possible ways. In our prior work [227], a different comparison was carried out.

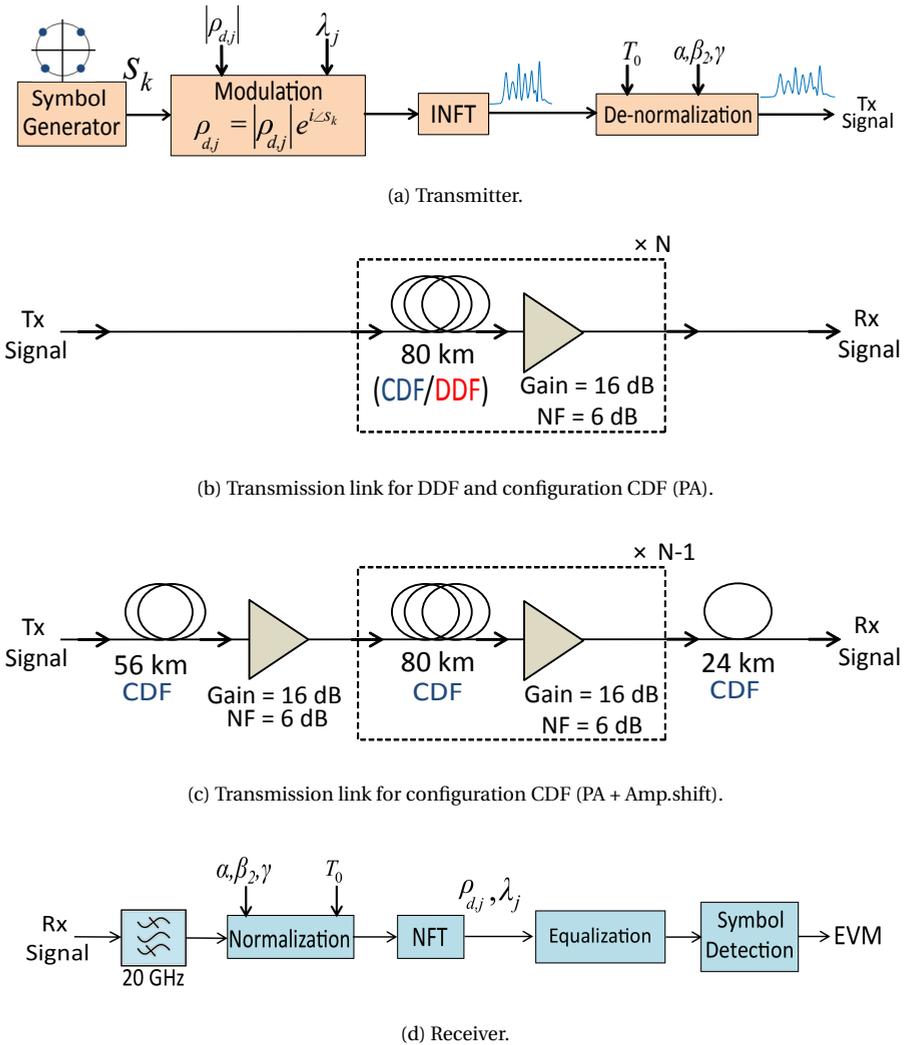


Figure 2.2: Simulation setup of NFDm system with discrete spectrum modulation.

lation setup is shown in Fig. 2.2. At the transmitter, randomly generated QPSK symbols were modulated on the discrete spectral values of the corresponding seven eigenvalues given in [142] and shown in Fig. 2.3. Then the time-domain multi-soliton pulse was obtained with the inverse NFT (INFT) operation. The duration of the normalized multi-soliton pulse was set to 18π in order to avoid truncation effects and pulse-overlap during propagation. The normalized multi-soliton pulse is then scaled using the fiber parameters and time-scale parameter T_0 . T_0 controls the duration of de-normalized pulse and hence the transmit power. A train of 1023 pulses was transmitted through the link for

each evaluation. For the case of NFDN system designed using CDF, we considered two types of link configurations. Fig. 2.2(b) shows the transmission link for the first configuration (referred as CDF-PA). In this transmission link, each span in the link consists of 80 km fiber followed by an Erbium-doped fiber amplifier (EDFA) to compensate the span-loss of 16 dB. The same link configuration was used for the NFDN system designed with DDF. The second configuration of link, shown in Fig. 2.2(c), is same as the first one except for the first and the last spans which have different lengths (referred as CDF-PA+Amp-shift). The lengths of the first and the last span were optimized according to the analysis in [171]. It was shown in [171] that the approximation error of path-average model is minimized at those optimal lengths. The transmit power in the second configuration refers to the power at the amplifier outputs. The noise figure of the EDFAs was set to 6 dB in both link configurations. The fiber propagation was simulated using a split-step Fourier method. At the receiver (shown in Fig. 2.2(d)), the signal was filtered to remove out of band noise. After normalization, the signal was then equalized in the nonlinear Fourier domain using (2.6) for CDF and (2.13) for DDF. The QPSK symbols were demodulated from the spectral values of the eigenvalues. Finally, performance is measured in terms of the error vector magnitudes (EVMs) of received symbols.

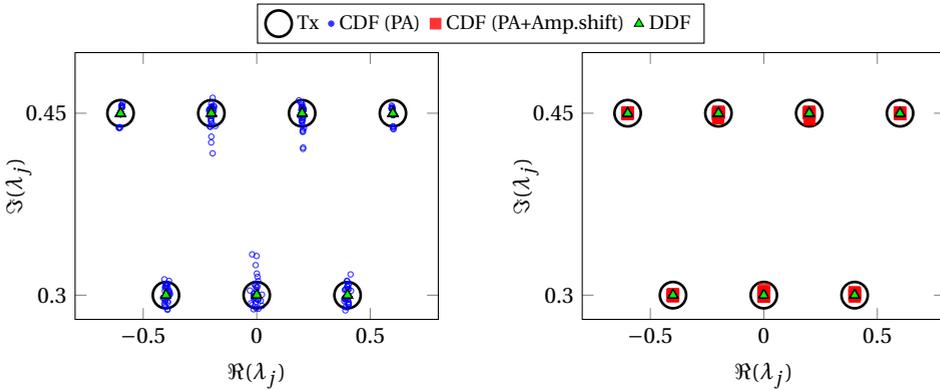


Figure 2.3: Transmitted and received eigenvalues for 8×80 km noiseless transmission at 2.82 dBm transmit power (left) and 4.7 dBm transmit power (right).

Fig. 2.3 shows the 7-eigenvalues transmitted over noiseless 640 km link and the corresponding received eigenvalues for the case of NFDN systems designed using DDF and CDF. We can see that in the absence of noise the NFDN system designed using DDF is exact and preserves the eigenvalues while there is fluctuation in the received eigenvalues of the NFDN system designed using CDF due to the involved path-average approximation. Fig. 2.4 shows EVM over transmit power for 16×80 km noiseless transmission. In the low transmit power region, the performance of the NFDN system designed using DDF is limited only by the accuracy of the numerical approximations. The performance of the NFDN systems designed using CDF (the other two cases) is mainly limited by errors due to the path-average approximation. We also observe that the NFDN system with CDF (PA + Amp. shift) configuration performs very close to the performance of the NFDN system designed using DDF. However, EVM for NFDN system with CDF (PA

+ Amp. shift) configuration is rising with increase in power. The EVM for the NFDM system designed using DDF is rising very slowly and then there is a sudden steep rise. This steep rise happens at higher powers (shorter pulse durations) where pulses start to overlap for DDF case due to the following reason.

2

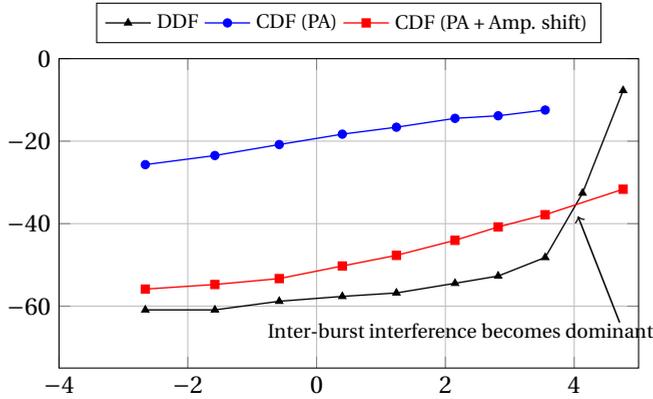


Figure 2.4: Discrete spectrum modulation: EVM over transmit power for 16×80 km noiseless transmission.

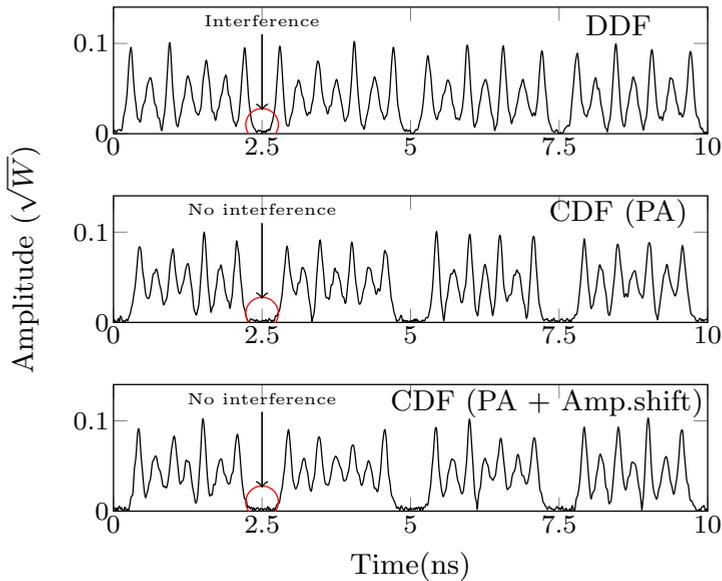


Figure 2.5: Signal after 16×80 km transmission over different link-configurations (at transmit power of 2.8 dBm). The pulses spread more for the case of DDF, resulting into interference among them.

Remark: The magnitude of the dispersion parameter of the DDF ($|\beta_2(0)|$) at the beginning of the span is larger than the corresponding value of the CDF. Thus, the normalization length ($L_D = T_0^2 / |\beta_2(0)|$) of the NFDM system designed using DDF is smaller

compared to the other two NFDM systems for a given transmit power. This means that the multi-soliton pulses transmitted over DDF evolves over a longer normalized distance (link length/ L_D). This results into the spreading of the individual pulses beyond its allocated duration at 16×80 km as shown in Fig. 2.5, causing interference with neighbouring pulses. This interference is the reason for the degraded performance in the NFDM system designed using DDF at higher power (smaller pulse-duration). However, the other two NFDM systems will also face the inter-pulse interference at longer transmission distances or higher power.

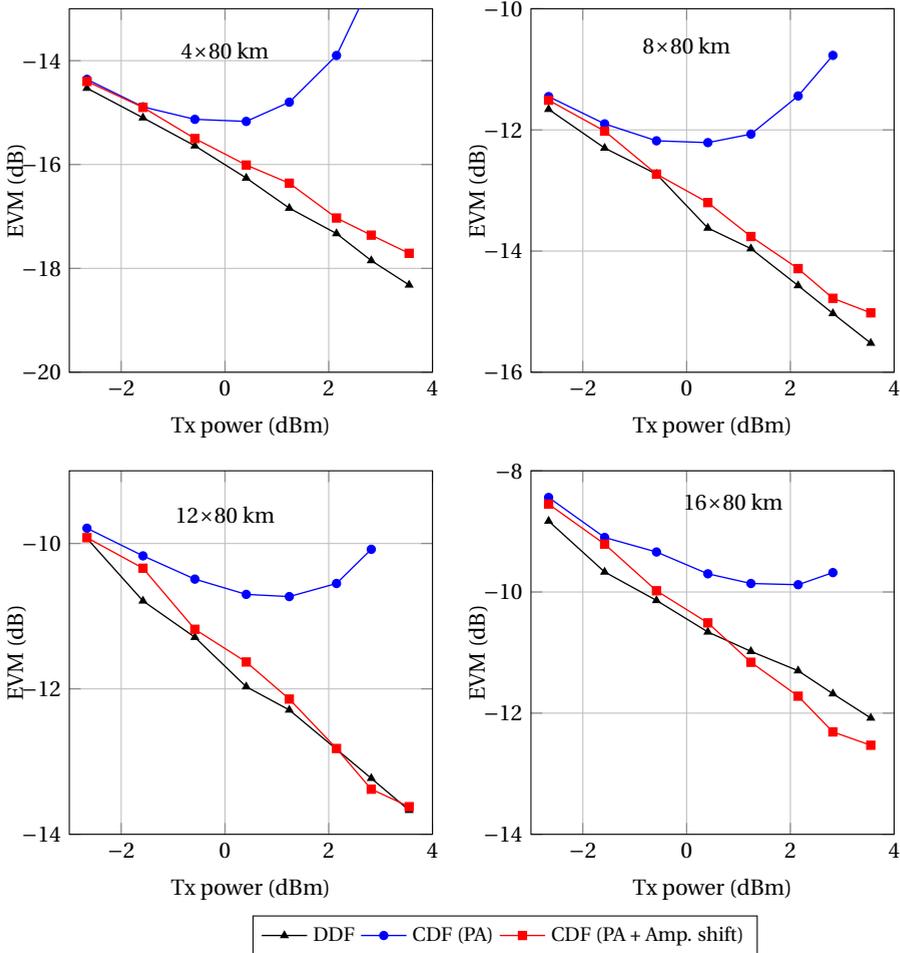


Figure 2.6: Discrete spectrum modulation : EVM over transmit power for different transmission distances.

In order to visualize the error due to the path-average approximation in presence of noise, the EVM is plotted in terms of the transmit power for different transmission dis-

tances in Fig. 2.6. For the NFDM system designed using CDF that uses the configuration CDF-PA, the EVM initially decreases with transmit power due to increase in effective signal to noise power ratio (SNR). But after a threshold transmit power, the approximation error due to the path-average model dominates and hence, the EVM starts rising with transmit power. For the NFDM system with CDF-PA+Amp-shift configuration, we did not observe any rise in the EVM in the simulated power range. The NFDM scheme that uses DDF performs slightly better than the NFDM scheme that uses CDF-PA+Amp-shift configuration except for the 16×80 km transmission. For 16×80 km transmission in DDF case, the pulses start overlapping at higher powers due to the reason explained earlier. We observe a large gain of up to ≈ 3 dB in the EVM for 640 km transmission over DDF in comparison to the NFDM system designed with CDF-PA configuration. We also observe that the NFDM system designed with CDF-PA+Amp-shift configuration performs as good as the system designed with DDF in this example, which however has been designed to keep the path-average approximation error small. In the next example we observe much higher gains in the NFDM system designed with DDF.

NFDM WITH CONTINUOUS SPECTRUM MODULATION

We considered the NFDM system with b -modulation presented in [160] to modulate continuous spectrum. In b -modulation, the information is modulated on the b -coefficient (2.5) instead of spectral function $\rho_c(\xi)$. One advantage of b -modulation is that it is easier to control the signal duration using b -modulation in comparison to spectral function $\rho_c(\xi)$ modulation [159], [160]. Another advantage is that the noise impact on the b -coefficient is less severe than the reflection coefficient [164]. The simulation setup of NFDM system is shown in Fig. 2.7. The nonlinear spectrum consists of nine flat-top shaped $b(\xi)$ carriers with carrier spacing of 15 (in the normalized NFT domain). The average energy of each carrier is controlled by an energy per carrier parameter E_d , which

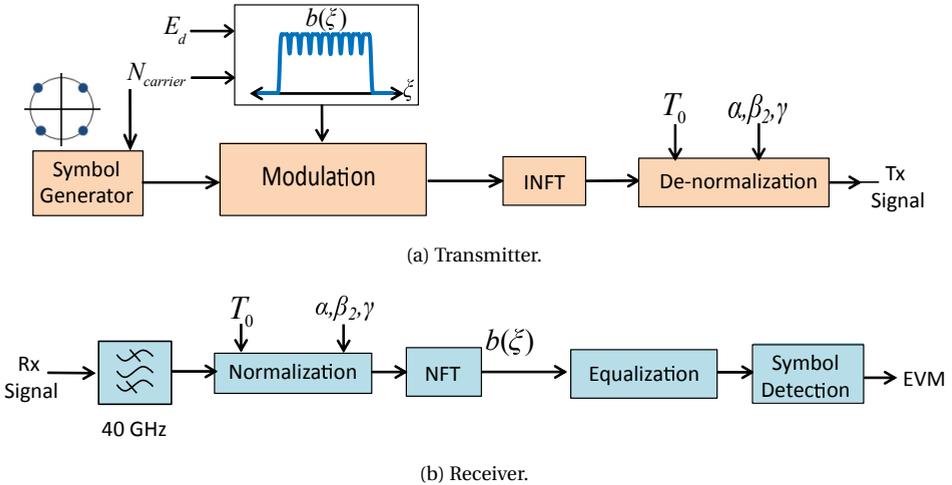


Figure 2.7: Simulation setup of NFDM system with continuous spectrum modulation. The transmission link is the same as described before in Fig. 2.2(b), (c).

in turn controls the transmit power. Each carrier is modulated with randomly generated QPSK symbols. Then, the INFT operation is performed to obtain time-domain signal with a normalized duration of 4.5. The time-domain pulse is then de-normalized using the fiber parameters and a time scale parameter T_0 of 1.25 ns. A train of 127 pulses is then transmitted through the link. The net data rate and signal bandwidth were 3.2 Gb/s and approximately 40 GHz respectively. The transmission link configurations are the same as described earlier in the case of discrete spectrum modulation and are shown in Fig. 2.2(b), (c). At the receiver, the signal is filtered and normalized. After the NFT operation the b -coefficients are obtained, which are then equalized using (2.6) for CDF and (2.13) for DDF. Finally, the symbols are detected from the $b(\xi)$ carriers and EVMs were computed for different transmission lengths and transmit powers controlled by E_d .

It must be noted that as the magnitude of b -coefficient cannot be greater than one, we have an upper limit on the carrier energy E_d [160], and thus on the transmit power [236]. It was shown theoretically in [236] that the system in [160] cannot exceed a finite power limit. Fig. 2.8 shows EVM over transmit power for 16×80 km noiseless transmission. We see that EVM for the NFDN system that uses DDF is significantly lower than the other two NFDN systems. In all three NFDN systems, the EVM increases with increase in transmit power. However, the rise in the EVMs for the NFDN systems that use CDF is steeper, which is mainly due to the errors from the path-average approximation. The rise in EVMs for the NFDN system designed using DDF is purely from the accuracy of numerical approximations. At higher power, we get closer to the aforementioned finite power limit and numerical errors occur which is visible for the case of DDF. The b -modulator is prone to breakdown in this region. Fig. 2.9 shows EVM over transmit power for the case of transmission in the presence of noise. The additive noise increases signal power slightly during transmission. Close to the power bound, this increase in signal power due to the noise is enough to push the received signal outside the range of the b -modulator (in this case the energy gets transferred into discrete spectrum that are not accounted for in the receiver). Thus, the impact of noise becomes severe in the higher power region when the signal approaches the aforementioned finite power limit. This results in the rise in the EVMs of all the considered NFDN systems in the high power region. In Fig. 2.9, the errors due to the path-average approximation are clearly

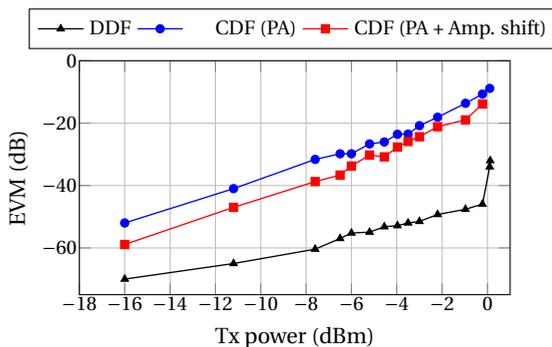


Figure 2.8: Continuous spectrum modulation: EVM over transmit power for 16×80 km noiseless transmission.

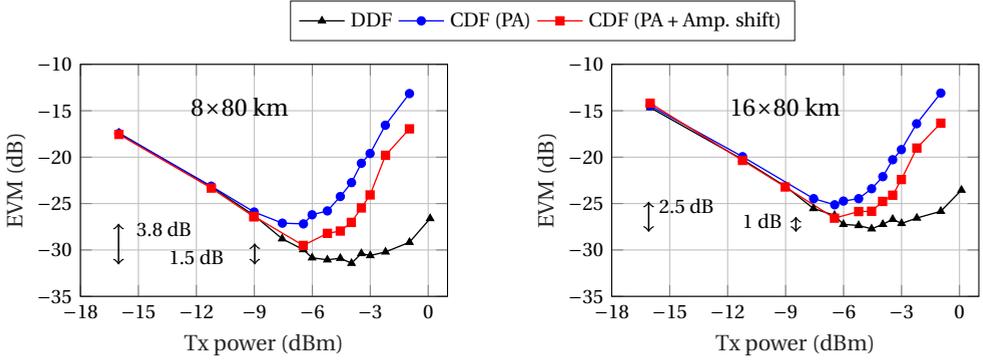


Figure 2.9: Continuous spectrum modulation : EVM over transmit power for different transmission distances.

visible. We can see that for 8×80 km transmission, EVM improves with increasing transmit power but after a threshold the EVM for CDF starts degrading, as the path-average error increased at higher power. Further, we see that approximation error in the path-average model is reduced by the amplifier-shift, hence the NFD M system that uses CDF-PA+Amp-shift configuration has better performance than CDF-PA. The performance of DDF improves with transmit power till -5 dBm, thereafter EVM degrades which is a result of the b -saturation effect. A similar trend is observed for the case of 16×80 km transmission. Fig. 2.10 shows the minimum EVMs obtained at the corresponding distances for all of the three NFD M systems. We obtained EVM gains of approx. 3.8 dB and 1.5 dB at 640 km with respect to path-average without amplifier-shift and path-average with amplifier-shift respectively. At 1280 km, these gains reduce to 2.5 dB and 1 dB respectively.

In this section, we compared NFD M systems in order to highlight potential advantages of using DDF in NFD M systems. The systems were designed to ensure the system parameters such that the transmitter requirements (bandwidth, power and burst duration)

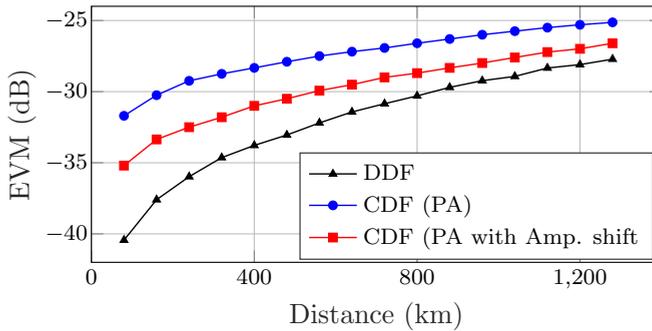


Figure 2.10: EVMs over transmission reach at optimal transmit powers.

are identical for all three systems. Our comparisons thus isolate and highlight the differences arising solely from choosing DDF instead of CDF. For the chosen system parameters, we also observe that time-spreading of bursts is smaller for the NFDM systems designed using CDF in comparison to the NFDM systems designed using DDF. This difference in time-spreading comes from the fact that the dispersion and nonlinear parameters of CDF and DDF are different. Thus, a smaller guard time interval could be allocated for the NFDM systems designed using CDF, resulting in higher spectral efficiency. A comparison between NFDM systems designed using DDF and CDF independently optimized to maximize performance (bit rate/spectral efficiency) would be an interesting next step.

2.5. CONCLUSION

We have presented numerical results for exact NFDM transmission by discrete and continuous spectrum modulation over lossy fiber. We have shown that by using a suitably designed fiber together with an adapted NFT, the approximation error from the path-average model can be avoided. We applied this approach for the case of fiber-links with EDFA based amplification. For discrete spectrum modulation, the NFDM system designed with DDF has EVM gains of up to 3 dB and 2 dB at 640 km and 1280 km respectively in comparison to the NFDM system with path-average model. The EVM performance of the NFDM system with path-average model and amplifier shifts was close to performance of NFDM system with DDF. For the continuous spectrum modulation case, by using DDF, we obtained EVM improvements of approximately 3.8 dB and 2.5 dB for 640 km and 1280 km transmission respectively in comparison to CDF with path-average model. By shifting the amplifiers to optimum locations for the path-average approximation, the gain obtained is approximately 1.5 dB and 1 dB respectively. We observed through simulations that in the presence of fiber-loss, the NFDM systems designed using DDF have clear performance advantage over the NFDM systems designed using CDF for the case of discrete as well as continuous spectrum modulation. It must be noted that the performance-gain obtained in an NFDM system designed with DDF depends on system design parameters. An NFDM system which uses a path-average model will have degraded performance at higher transmit power, bandwidth and longer span, and hence, the gains of employing DDF increase further.

APPENDIX

NONLINEAR FOURIER TRANSFORM FOR DDF

It was shown in [231, Sec. 3] that if (2.12) is satisfied, the propagation model given by (2.11) can be solved with suitable NFTs. In order to compute and evolve the NFT with respect to (2.11), the following set of linear equations have to be solved [231]

$$\vartheta_{\tau} = \begin{pmatrix} -j\lambda & \sqrt{\frac{R(z)}{D(z)}}q \\ -\sqrt{\frac{R(z)}{D(z)}}q^* & j\lambda \end{pmatrix} \vartheta, \quad (2.17)$$

$$\vartheta_z = \begin{pmatrix} -jD(z)\lambda^2 + j\frac{R(z)}{2}|q|^2 & \sqrt{R(z)D(z)}(\lambda q + \frac{j}{2}q_\tau) \\ \sqrt{R(z)D(z)}(\lambda q^* + \frac{j}{2}q_\tau^*) & jD(z)\lambda^2 - jR(z)|q|^2 \end{pmatrix} \vartheta. \quad (2.18)$$

where $q = q(z, \tau)$, λ is an eigenvalue and ϑ is an eigenvector of the eigenvalue problem (2.17). Here, the subscript denotes differentiation with the corresponding variable.

From (2.3), (2.17) and [81, equation (16)], it is clear that computing the NFT with respect to (2.11) is same as computing the conventional NFT (with respect to (2.2)) with the potential $q(z, \tau)$ scaled by a factor of $\sqrt{\frac{R(z)}{D(z)}}$. The evolution of the nonlinear spectrum is described with respect to (2.18). Following [81, equation (24)], one finds that the evolution of nonlinear spectrum is given by the rotation described in (2.13).

3

DEEP NEURAL NETWORK-BASED DIGITAL PRE-DISTORTION FOR HIGH BAUDRATE OPTICAL COHERENT TRANSMISSION

High-symbol-rate coherent optical transceivers suffer more from the critical responses of transceiver components at high frequency, especially when applying a higher order modulation format. We recently proposed a neural network (NN)-based digital pre-distortion (DPD) technique trained to mitigate the transceiver response of a 128 GBaud optical coherent transmission system. In this paper, we further detail this work and assess the NN-based DPD by training it using either a direct learning architecture (DLA) or an indirect learning architecture (ILA), and compare performance against a Volterra series-based ILA DPD and a linear DPD. Furthermore, we deliberately increase the transmitter nonlinearity and compare the performance of the three DPDs schemes. The proposed NN-based DPD trained using DLA performs the best among the three contenders. In comparison to a linear DPD, it provides more than 1 dB signal-to-noise ratio (SNR) gains at the output of a conventional coherent receiver DSP for uniform 64-quadrature amplitude modulation (QAM) and PCS-256-QAM signals. Finally, the NN-based DPD enables achieving a record 1.61 Tb/s net rate transmission on a single channel after 80 km of standard single mode fiber (SSMF).

The text in this chapter has previously appeared in modified form in open access article V. Bajaj, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "Deep Neural Network-Based Digital Pre-distortion for High Baudrate Optical Coherent Transmission". In *Journal of Lightwave Technology*, vol. 40, no. 3, pp. 597-606, 1 Feb. 2022, doi: 10.1109/JLT.2021.3122161. [89]. Its reuse is licensed under CC BY 4.0 [222].

3.1. INTRODUCTION

THE exponential increase in the internet traffic due to the emergence of bandwidth-hungry services such as cloud-based applications and video on demand is pushing the existing optical transport network to its limit. To increase the aggregate bit rate carried by a single fiber strand, one must find ways to best utilize the available optical spectrum while minimizing the number of components required to do so. The three main avenues to attain this objective are to increase the symbol rate and the average number of bits conveyed per symbol on a carrier, and decrease the spectral guard band between multiplexed carriers. Thus, it is desirable to operate such systems at high symbol rates on a tight spectral grid using high-order modulation formats to maximize the information rate [93], [237]–[239]. Therefore, the signal-to-noise ratio (SNR) should be as high as possible, a necessary condition for operating on a high order format.

The generation of signals with high integrity is challenging due to the impairments stemming from different sources along the information transmission system. At the transmitter side, these impairments include the limited bandwidth and the nonlinear characteristics of components. A common practice to mitigate these distortions is by digitally pre-compensating them in the digital signal processing (DSP) stack, a technique usually termed "digital pre-distortion" (DPD).

A linear DPD is generally employed to compensate for linear inter-symbol interference stemming from the limited bandwidth and/or the imperfect spectral response of the transmitter components [240], [241]. It is a common practice to limit the amplitude of the signals applied to transmitter components exhibiting a nonlinear response (e.g. driver amplifier (DA), electro-optic modulator, etc.) when using a linear DPD, which in turn limits the SNR because of the small signal power. A larger signal swing can improve the SNR, but may require to be accompanied by a nonlinear DPD to pre-compensate the increased nonlinear distortions. To increase the information rate, transmitters will require DPDs that can compensate for both the linear and nonlinear responses. The most common nonlinear DPDs are based on Volterra series which have been investigated for both radio frequency (RF) amplifiers [242]–[246] and coherent optical transmitters [87], [247]–[251].

Another type of DPD is based on neural networks (NNs) whose application dates back to 1980s [252], [253]. Recently, NN-based DPDs have received more attention [88], [254]–[260]. A simple feed-forward NN (FFNN) was used in [256] to mitigate the response of RF amplifier, without considering any memory effects. The memory effects were included in the DPDs based on time-delay NNs (TDNNs) [254], [258], [261] and on convolutional NNs (CNNs) [259]. Recently, some of the above schemes have been compared in [255] and shown experimentally that adding residual neural network (ResNet) structure improves the nonlinearity mitigation of RF amplifiers. However, NN-based DPDs for optical coherent transmitters are so far not well explored. Memoryless FFNNs were proposed to pre-compensate for Mach-Zehnder modulator (MZM) responses [262] and a simulated low-resolution digital to analog converter (DAC) response [257]. In [260], a DPD based on recurrent NNs (RNNs) is applied to the simulated aggregate response of an optical coherent transmitter. In another paper [263], a NN-based DPD was designed by considering the collective response of a coherent transmitter as a Wiener-Hammerstein (WH) model.

Very recently, we proposed an NN-based DPD designed using simple FFNNs and CNNs for a high-baud rate (128 GBaud) coherent optical transmitter [264]. In this paper, we demonstrate that the considered NN-based DPD leads to a record 1.61 Tb/s data rate over a 80 km fiber link, detail the proposed NN-based DPD and further investigate its performance by training it using the two well-known learning architectures, namely the indirect learning architecture (ILA) [242], [252] and the direct learning architecture (DLA) [243], [244]. In addition, we also consider a Volterra series-based ILA DPD and a linear DPD in our investigation. All considered DPDs are disjoint-IQ implemented using real values i.e. two separate DPDs each for I and Q channel. The DPDs are applied to two modulation formats, namely 64-quadrature amplitude modulation (QAM) and probabilistic constellation shaped (PCS)-256-QAM, and trained and evaluated for varying transmitter nonlinearity. Our results show that NN-based DPD trained using DLA performs the best among the considered candidates and obtains gains of 1.6 dB and 1.2 dB in received SNR with respect to the linear DPD for uniform 64-QAM and PCS-256-QAM, respectively. In addition, we compare the complexity of DPDs considered in the study. We reduce the complexity of the proposed NN-based DPD by applying a pruning method.

The outline of the paper is as follows: in Sec. 3.2 the experimental setup is described. The DPD techniques considered in this work and their implementation are discussed in Sec. 3.3. The proposed NN-based DPD is detailed in Sec. 3.4, and performance assessments are presented in Sec. 3.5. A complexity comparison is added to Sec. 3.6. The paper is concluded in Sec. 3.7.

3.2. HIGH BAUD RATE COHERENT OPTICAL TRANSMITTER

A schematic of our experimental setup is shown in Fig. 3.1. The transmitter (Tx) comprises three major components: digital to analog converters, driver amplifiers and an optical modulator (external IQ modulator). Each of the components have linear and nonlinear characteristics. In addition, the signal reflections originating by the radio frequency (RF) cables/connections add to the linear effects and further spread the impulse response. Overall, analog signals leaving each DAC flow through a chain of linear and nonlinear responses.

We used uniform 64-QAM and probabilistic constellation shaped 256-QAM (PCS-256-QAM) signals. The PCS 256-QAM format was shaped using the Maxwell-Boltzmann distribution with an entropy of 7.5 bits/symbol which was previously determined as a good choice for 19 dB SNR, the SNR limit in the setup. In the transmitter, the data consisting of a 2^{15} symbol sequence passes through the DPD block. The pre-distorted sequence is then loaded into the DACs after clipping and quantization. The DACs sample the signal at 128 GSa/s and operate at 1 sample per symbol (sps). The DACs have an effective number of bits (ENOB) of 4 at 64 GHz and 24 GHz 3-dB bandwidth. The two DACs produce the two electrical tributaries feeding a single-polarization optical IQ modulator. The outputs of the DACs are first amplified using DAs with 60 GHz 3-dB bandwidth. The DAs' outputs are then fed to the lithium-niobate (LiNbO_3) IQ modulator which has around 41 GHz 3-dB bandwidth. The optical carrier is generated by an external cavity laser (ECL) at 193.5 THz with a line-width of <100 kHz. The optical carrier is fed into the IQ modulator, where it gets modulated by the signals from the DAs' outputs. A po-

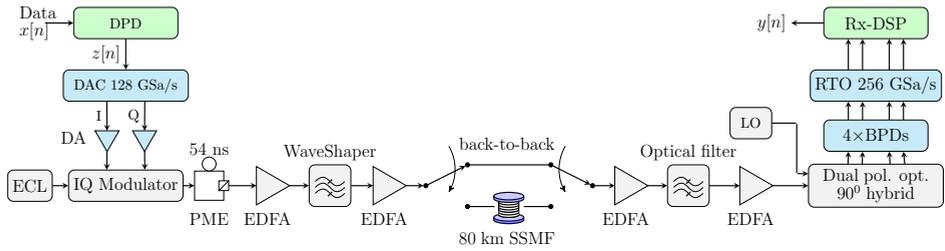


Figure 3.1: A schematic of the 128 GBaud experimental setup configured either in back-to-back or in an 80 km fiber transmission arrangement.

larization multiplexing emulator (PME) with a decorrelation delay of 54 ns is used to generate a polarization multiplexed signal. The polarization multiplexed signal is amplified using an Erbium doped fiber amplifier (EDFA). A Finisar WaveShaper is used to compensate the low pass response of the IQ modulator by increasingly attenuating frequencies closer to the carrier in order to flatten the optical spectrum at its output. The WaveShaper is configured once when a linear DPD was employed at the transmitter and kept fixed. The signal is then amplified using another EDFA and either sent directly to the coherent receiver or transmitted through 80 km of standard single mode fiber (SSMF) before coherent reception. The coherent receiver is preceded by an EDFA, an optical filter of 128 GHz 3-dB bandwidth to remove the amplified spontaneous emission noise, and a second EDFA. The resulting optical signal beats with a local oscillator through a dual-polarization 90° hybrid. Four balanced photo-diodes (BPDs) detects the signal. A Keysight high bandwidth real time oscilloscope (RTO) is used to sample and record the four detected waveform at 256 GSa/s. The RTO has a nominal resolution of 10 bits. An offline DSP for symbol recovery after optical coherent detection is applied. Note that, we intend to compensate the transmitter impairments using a DPD. While it is difficult to isolate the transmitter and the receiver impairments in an experimental setup, methods like homodyne detection to mitigate certain receiver impairments can be used. Also, some of the receiver impairments such as low pass response of photo-detectors can be separately determined and compensated by employing static filters. However, such scenarios are challenging for integrated transceivers. To train the DPD parameters in the transmitter DSP, we chose to always apply the same DSP stages at the receiver. Applying coherent receiver DSP allows to convert the inherently dynamic channel response (e.g.: rotation of the state of polarization followed by a polarization beam splitter at the receiver, beating of the incoming optical signal with a free running, non-phase locked laser source) into a stationary channel response as also applied in previous DPD research works [87], [247], [248].

The receiver DSP first re-samples the signal at 2 samples per symbol. Then chromatic dispersion is removed and timing errors are corrected. The polarizations of the signal are de-multiplexed by using a 2×2 complex-valued multiple-input multiple-output (MIMO) equalizer updated by a multi-modulus algorithm (MMA). Intermediate frequency offset and phase noise are then compensated. The residual signal distortions are compensated by another MIMO equalizer, operated as 4×4 on real values. Note that some of the im-

pairments that originate at the transmitter may get corrected by this adaptive 4×4 real-valued MIMO equalizer used in the receiver DSP. Consequently, the transmitter may not compensate for some of the impairments that actually occurred at the transmitter because they are always handled by the receiver DSP. As transmitter pre-distortion is the main focus of this work, we try to compensate most of the impairments at the transmitter. There could be different ways to do this, such as training the DPDs by excluding the 4×4 real-valued MIMO equalizer from the training loop [265]. In this work, the DPDs are trained in a step by step procedure by changing the length of the receiver 4×4 real-valued MIMO equalizer.

The transmission quality is measured in terms of SNR, mutual information (MI) and generalized MI (GMI). For decoding, we used a family of 130 optimized spatially coupled LDPC codes [237], [266] with variable overheads ranging from 3% to 100%. For each channel, the code with the smallest overhead capable of decoding the bits error-free is chosen.

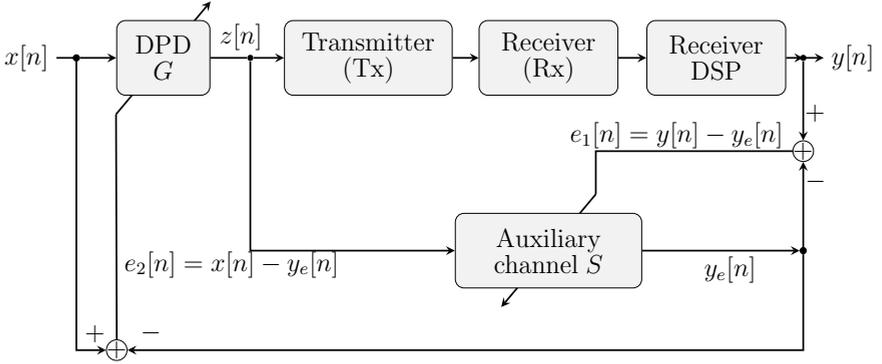
3.3. REVIEW OF VARIOUS DPD TECHNIQUES

In our work, we evaluated the performance of a linear DPD as well as nonlinear DPDs based on either Volterra series or the proposed neural network architecture. In the following, we first describe the two general training methods well-known in literature namely direct learning architecture (DLA) and indirect learning architecture (ILA). Then, we briefly review the Volterra series-based DPD. The proposed NN-based DPD is described in the next section.

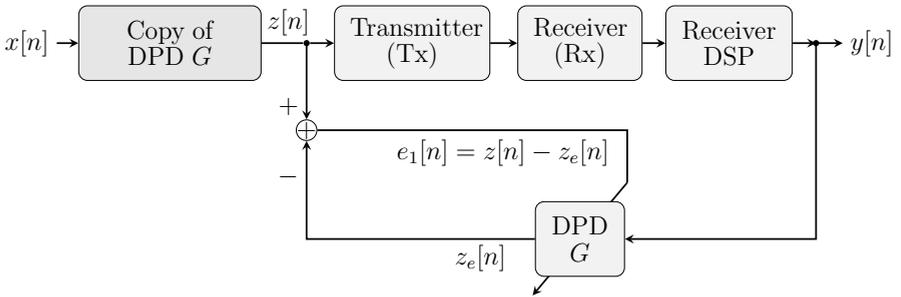
3.3.1. DIRECT VS. INDIRECT LEARNING ARCHITECTURE

An example schematic of the DLA is shown in Fig. 3.2(a). In DLA, the "communication channel" is modelled by a differentiable auxiliary channel model S with the help of which the DPD is determined in the following two steps. In the first step, the auxiliary channel model is trained by minimizing the objective function $J_1 = \sum_n \frac{1}{2} (e_1[n])^2 = \sum_n \frac{1}{2} (y[n] - y_e[n])^2$. Here, $y[n]$ is the soft symbols output of the receiver DSP and $y_e[n]$ is the output of the auxiliary channel model. Both sequences $y[n]$ and $y_e[n]$ are corresponding to a sequence $z[n]$ injected into both the communication channel and the auxiliary channel model. In the second step, once the auxiliary channel model S cannot further minimize the objective J_1 , S is fixed to its current state and only the DPD G is updated to minimize $J_2 = \sum_n \frac{1}{2} (e_2[n])^2 = \sum_n \frac{1}{2} (x[n] - y_e[n])^2$. The gradients of the loss function are back-propagated through S in order to train G ; the parameterized digital pre-distortion function. The input to the DPD is $x[n]$ which consists of ideal QAM symbols. The DPD obtained using the second step changes the statistics of the input signal $z[n]$; consequently, changing the response of the communication channel. So, the auxiliary channel model needs to be retrained. Thus, using the DLA architecture, the parametrized functions S and G are iteratively trained until no more gains are obtained.

In contrast to DLA, the ILA architecture does not require an auxiliary channel model, as is shown in Fig. 3.2(b). With ILA, the DPD G is trained at the output of the communication channel as a post-equalizer, while a copy of the pre-distorter G (obtained from the previous iteration) is used at the input of the transmitter. The new DPD G is trained



(a)



(b)

Figure 3.2: An example representation of training a DPD G using (a) direct learning architecture (DLA) and (b) indirect learning architecture (ILA). S is the auxiliary channel model.

by minimizing the objective function $J_1[n] = (e_1[n])^2 = \sum_n \frac{1}{2} (z[n] - z_e[n])^2$. The input to the DPD (training block) is the soft symbols output of the Rx-DSP $y[n]$. The “communication channel” response changes with signal statistics, hence, several iterations are needed to achieve good convergence. As ILA does not require an auxiliary model, its computational complexity in the training phase is almost halved compared to DLA. However, it suffers from a bias caused by nonlinear operations of DPD on the transmitter output which is often noisy, as explained in [267]. Further, the nonlinear DPD trained using ILA may not be the optimum as nonlinear blocks may not be commutative.

3.3.2. LINEAR DPD

The output of a M_1 memory linear DPD $z_e[n]$ for inputs $y[n]$ is given by

$$z_e[n] = \sum_{\tau_1=-M_1}^{M_1} g[\tau_1] y[n - \tau_1], \quad (3.1)$$

where $g[\tau_1]$ are filter coefficients. The above relation can be written as

$$\vec{z}_e = \mathbf{Y} \vec{g}, \quad (3.2)$$

where $(\vec{\cdot})$ denotes vector quantities and \vec{g} is the vector of the linear filter coefficients $g[\tau_1]$. The matrix \mathbf{Y} is made of columns of shifted vectors $\vec{y}_k = [y[k], y[k+1], \dots, y[k+n]]^T$. Here, $(\cdot)^T$ denotes the transpose operation. The matrix \mathbf{Y} is represented mathematically as $\mathbf{Y} = [\vec{y}_{-M_1} \dots \vec{y}_{+M_1}]$. The coefficients vector \vec{g} can be obtained using the Moore-Penrose inverse by the relation

$$\vec{g} = (\mathbf{Y}^H \mathbf{Y})^{-1} \mathbf{Y}^H \vec{z}_e \quad (3.3)$$

where $(\cdot)^H$ represents complex-conjugate transpose and $\vec{z}_e = [z_e[0], z_e[1], \dots, z_e[n]]^T$. Two separate disjoint linear DPD models were determined for I and Q tributaries.

3.3.3. VOLTERRA SERIES-BASED DPD

Volterra series are well-known to model nonlinear systems with memory [268]. The output $z_e[n]$ of a nonlinear system up to a nonlinearity order of three with memory can be written in terms of its input $y[n]$ in the form of

$$\begin{aligned} z_e[n] = & g_0 + \sum_{\tau_1=-M_1}^{M_1} g_1[\tau_1] y[n-\tau_1] + \sum_{\tau_2=-M_2}^{M_2} \sum_{d_2=0}^{D_2} g_2[\tau_2, d_2] y[n-\tau_2] y[n-\tau_2-d_2] \\ & + \sum_{\tau_3=-M_3}^{M_3} \sum_{d_2=0}^{D_2} \sum_{d_3=d_2}^{d_2+D_3} g_3[\tau_3, d_2, d_3] y[n-\tau_3] y[n-\tau_3-d_2] y[n-\tau_3-d_3] \end{aligned} \quad (3.4)$$

where M_p is the memory length and g_p are Volterra kernel coefficients in the p^{th} order. We considered a Volterra series-based DPD trained using ILA. The Volterra kernel (pre-distorter) coefficients g_p that map y to z_e can be represented in the matrix form like (3.2). Here, \vec{g} is the vector of the Volterra kernel coefficients g_p and the matrix \mathbf{Y} is made of columns of shifted vectors $\vec{y}_k = [y[k], y[k+1], \dots, y[k+n]]^T$ and columns generated by element-wise multiplications of shifted versions of \vec{y} . The input $y[n]$ is normalized to have a unit variance prior to generating the matrix \mathbf{Y} is represented mathematically as $\mathbf{Y} = [\vec{y}_{-M_1} \dots \vec{y}_{+M_1} \dots \vec{y} \odot \vec{y} \dots \vec{y}_{M_2} \odot \vec{y}_{M_2-D_1} \dots]$. In the above relation, the element-wise multiplication operation is denoted by \odot . The coefficients g_p can be obtained by using the Moore-Penrose inverse with relation (3.3). In this paper, we do not consider IQ cross-talk compensation at the transmitter, hence, two separate disjoint Volterra-series based DPD models were determined each for I and Q tributaries which were implemented using real values.

3.4. NEURAL NETWORK-BASED DPD

First, we modelled the optical coherent transmitter using experimentally acquired linear responses of the DACs and the DAs and simulated nonlinear responses of DACs, DAs and MZMs. Then, we tested different architectures in numerical simulations. The linear memory was accounted using convolutional neural networks (CNNs) which are easy to

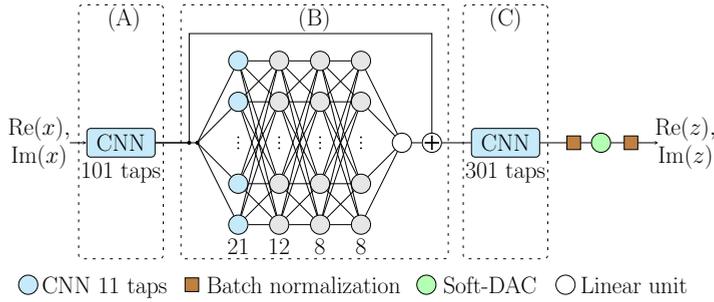


Figure 3.3: Architecture of the proposed NN-based DPD.

interpret. A total memory of around 400 taps was needed for the CNNs due to the signal reflections with a large time delay present in the experimental setup. These reflections are shown in Fig. 3.5 and explained in the next section.

To account for nonlinearity, we used fully connected layers with leaky ReLU activation functions which can be evaluated with a few simple operations. Nonlinearity mixed with memory was introduced in the NN by adding a convolution layer before the fully connected layers. A sufficient number of layers and neurons were then determined over the simulation setup by different trials and by observing the performance (in terms of SNR) of a given architecture. It was known that DACs have strong low pass response (8 dB attenuation at 64 GHz [238]) and a non-negligible degradation of the signal quality at the transmitter stems from the power amplifiers nonlinearity. Hence, we design an NN-based DPD architecture by inspiring from the so-called Wiener-Hammerstein (WH) structure [268] i.e. by keeping the nonlinear fully connected FFNN layers in between the two linear memory (CNNs). This architecture gave better performance than others. Note that, although the structure of the proposed DPD is similar to a WH system, it does not belong to this class. The reason is that the nonlinear part is not static, but has memory. We found that as the linear effects are more dominant, an additional shortcut bypass (ResNet connection [180]) to the nonlinear FFNN improves the performance and speeds up the training process. Later, the number of layers and neurons was further reduced over the experimental setup to reach a final architecture described as follows.

The architecture of the DPD is shown in Fig. 3.3. The sections (A) and (C) are uni-dimensional (1-D) linear CNNs, which are equivalent to finite impulse response (FIR) filters that compensate the linear responses while section (B) in the middle mainly corrects for the nonlinear responses. In section (B), the first layer (in cyan color) consists of short 1-D linear convolutions of 11 taps feeding to a layer with 21 neurons. The following three layers are fully connected FFNN layers with leaky rectified linear unit (Leaky ReLU) activation functions. The last layer in the FFNN has a single linear unit. The size of each layer is detailed in Fig. 3.3 and in Tab. 3.1.

The NNs were implemented using real values. The complex signal was processed separately using two disjoint NNs. We used both DLA and ILA based training for the NN-based DPD and refer them as “NNDLA” and “NNILA” in the rest of the paper, respectively. For NNDLA based training, we used another NN serving as an auxiliary channel model S . Its architecture was designed as a mirrored version “(C)→(B)→(A)” of the DPD

architecture shown in Fig. 3.3.

In order to model the DAC in our NN-based DPD, an approximation of the DAC was used to avoid the vanishing gradient problem. We call this customized unit as Soft-DAC and describe it in the following subsection.

3.4.1. SOFT-DAC ACTIVATION UNIT

Soft-DAC unit models the DAC with resolution of m bits and should quantize its input uniformly to 2^m discrete levels. As activation functions in NNs should have a non-zero derivative to pass gradients back, using a staircase activation function is not possible as its derivative is zero everywhere.

The output of the Soft-DAC unit, u_{out} , is defined by

$$f(u_{in}; s) = \begin{cases} \lfloor u_{in} \rfloor + sr & r \leq th \\ \lfloor u_{in} \rfloor + 0.5 + (r - 0.5)/s & th < r < 1 - th \\ \lfloor u_{in} \rfloor + 1 + s(r - 1) & r \geq 1 - th, \end{cases}$$

$$u_{out} = \max\{\min\{f(u_{in}; s), 0\}, 2^m - 1\},$$

where u_{in} is the input, $\lfloor \cdot \rfloor$ is the floor function, $r = u_{in} - \lfloor u_{in} \rfloor$, s is the softening factor and $th = 0.5/(1 + s)$. The Soft-DAC unit is implemented as a piece-wise linear function with $2^{m+1} + 1$ linear pieces. The slope of pieces is alternatively s and $1/s$. The behavior of the Soft-DAC can be changed from clipping-only operation to clipping and m -bits quantization operation by varying the slope s from 1 to 0, as shown in Fig. 3.4. The input to the Soft-DAC should be scaled and shifted properly such that it fits around the range of 0 to $2^m - 1$. A batch normalization (BN) layer is used prior to the Soft-DAC. The scaling parameter in the BN layer optimizes the clipping and is optimized by manually decreasing s from 1 towards 0 during the training of the pre-distorter. Finally, when $s = 0$ the outputs of Soft-DAC are discrete levels and any preceding NN layers to the Soft-DAC cannot be trained.

The Soft-DAC unit is one way to apply quantization in the NN framework. There are alternative ways available in the literature to implement quantization in the NN framework. We refer to [269] and references there in the paper. The problem at hand is not

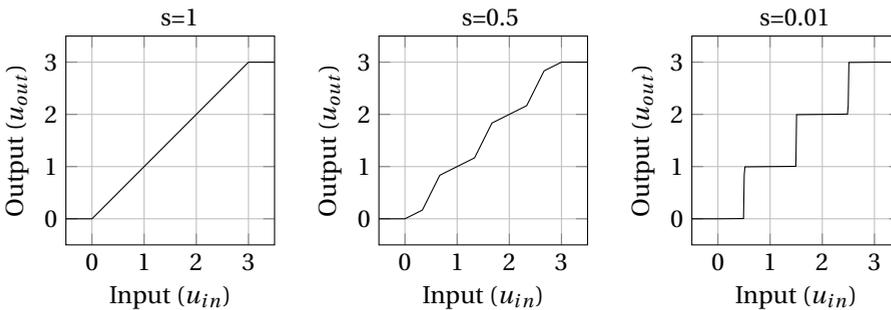


Figure 3.4: A 2-bit Soft-DAC activation unit for different softening factor (s).

the quantization of the weights (NN model parameters) but the activations. Some of the popular methods of activation quantization are by using approximation methods. In these methods, the forward pass through NN usually has an ideal quantization and the corresponding backward pass is implemented by using some approximated function such as a smoothed version of the ideal quantization function or a straight through estimator. One challenge with these approximation methods is that they have a gradient mismatch problem as described in [270]. The soft-DAC avoids gradient mismatch as it is a differentiable function.

3

3.4.2. INITIALIZATION AND TRAINING OF THE NN-BASED DPDs

The linear CNN layers in Fig. 3.3 are equivalent to FIR filters. The tap values of the CNN layer is initialized using “impulse initialization” i.e. all weights are set to zeros except for the center weight whose value is set to one. The weights and biases of the remaining FFNN part were initialized using Kaiming uniform initialization i.e. by randomly sampling a uniform distribution $U(-1, 1)\sqrt{k}$, where $k = 1/(\text{number of weights or biases})$. The network size hyperparameters and initialization are summarized in Tab. 3.1. It is also possible to initialize the first and the last 1-D CNNs using the known linear DPD response. We observed that by doing so the overall performance did not change, however, convergence is achieved quicker.

For both NN-based DPDs i.e. NNILA and NNDLA, we used gradient descent back-propagation with the mean square error loss function and the Adam optimizer [271]. We used sequences of 2^{18} symbols to train the NNs. The batch size should be large enough to capture the transmitter memory. We used a larger batch size to minimize the fluctuation in the mean and the variance of individual batches because of the BN layer operation. The learning rates were determined by doing a grid search on a logarithmic scale. The training hyperparameters are summarized in Tab. 3.2.

For NNILA, the training data consists of the received signal $\mathbf{y}[n] = [\text{Re}(y[n]), \text{Im}(y[n])]$ and pre-distorted signal $\mathbf{z}[n] = [\text{Re}(z[n]), \text{Im}(z[n])]$ as inputs and targets, respectively. In the first step of NNDLA, the auxiliary channel model is trained by using pre-distorted signal $\mathbf{z}[n] = [\text{Re}(z[n]), \text{Im}(z[n])]$ and the soft symbols output of the receiver DSP $\mathbf{y}[n] = [\text{Re}(y[n]), \text{Im}(y[n])]$ as inputs and targets. While in the second step, the cascaded NN i.e. pre-distorter followed by auxiliary channel model, uses $\mathbf{x}[n] = [\text{Re}(x[n]), \text{Im}(x[n])]$ as its input and target. Note that, in the second step, only the pre-distorter part of the cascaded NN is updated and the auxiliary channel model part is kept fixed.

3.5. RESULTS

In this section, we explain the procedure and the results of our experimental study. The training of the considered DPDs was done by using either the DLA or the ILA described in section 3.3. In the first iteration of the training, a pre-distorted signal z obtained by a linear static pre-distortion filter was used. This linear pre-distortion filter was already known from a characterization of the DACs and the DAs in the electrical domain. In the following subsection, we describe how we determined the memory needed in the considered DPDs.

Table 3.1: Table of network size hyperparameters of the NN-based DPD.

DPD-NN layer	Weight	Bias	Weight initialization	Bias initialization	Activation function
First CNN layer	101	1	Impulse	Zero	-
FFNN layer 1 (CNN)	$11 \times 21 = 231$	21	Impulse	Zeros	-
FFNN layer 2	$21 \times 12 = 252$	12	Kaiming uniform	Kaiming uniform	Leaky ReLU (0.1)
FFNN layer 3	$12 \times 8 = 96$	8	Kaiming uniform	Kaiming uniform	Leaky ReLU (0.1)
FFNN layer 4	$8 \times 8 = 64$	8	Kaiming uniform	Kaiming uniform	Leaky ReLU (0.1)
FFNN layer 5	$8 \times 1 = 8$	1	Kaiming uniform	Kaiming uniform	-
Last CNN layer	301	1	Impulse	Zero	-
BN layer 1	1	1	One	Zero	-
BN layer 2	1	1	One	Zero	-
Auxiliary-NN layer	Weight	Bias	Weight initialization	Bias initialization	Activation function
First CNN layer	301	1	Impulse	Zero	-
Last CNN layer	101	1	Impulse	Zero	-

Table 3.2: Table of training hyperparameters of the NN-based DPD.

Hyperparameter	Value
Optimization	Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)
Learning rate (DPD-NN)	1×10^{-4}
Learning rate (Auxiliary-NN)	5×10^{-4}
Batch size (DPD-NN)	4096
Batch size (Auxiliary-NN)	2048
Data length	2^{18}
Training epochs	30

3.5.1. REQUIRED FILTER LENGTH

A linear DPD with very long memory was trained in order to determine the memory required in the pre-distorter. The impulse response of the linear DPD filter after conver-

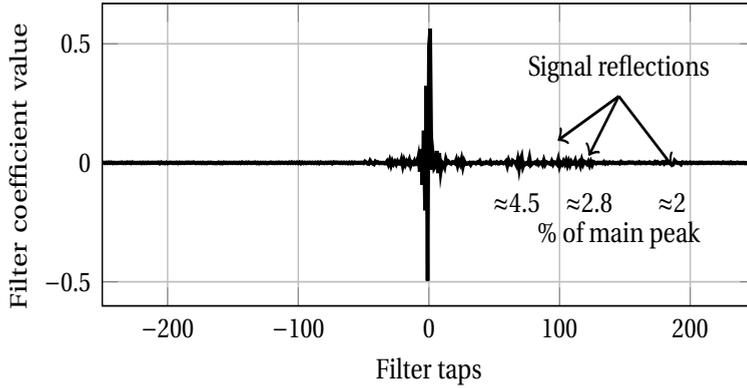


Figure 3.5: The experimentally obtained coefficients of the “I” tributary Linear DPD filter. A filter with long memory is required to capture the reflections of the signal delayed by ≈ 200 symbol periods.

gence of the adaptive training algorithm is shown in the Fig. 3.5. We observe that signal reflections are present even around 200 symbol duration delay. These reflections were possibly generated by the RF cables/connections. Thus, we set the length of linear DPD as well as the first order coefficients of Volterra to 441 taps. In NN-based DPDs (NNDLA, NNILA), memory (the sum of tap-lengths in CNNs) was set to around 440 taps.

The required memory for the second and the third order Volterra kernels for the DPD are determined in a similar way. The memory order and depth for the second order terms are $M_2 = 10$ and $d_2 = 4$, respectively. For the third order terms, we used $M_3 = 5$, $d_2 = 2$ and $d_3 = 3$, respectively. In total, the Volterra series-based DPD uses 105 second order and 99 third order coefficients along with 441 linear coefficient and one bias coefficient.

3.5.2. TRAINING PROCEDURE

As explained previously, we use the following training procedure so that most of the transmitter impairments are compensated at the transmitter side via the DPD and not by the adaptive 4×4 real-valued MIMO equalizer at the receiver DSP. We first train the DPD when the MIMO equalizer has a memory length of 101 taps. Then, after convergence, the memory length of the 4×4 real-valued MIMO equalizer is increased, and the DPD is trained again. More specifically, the MIMO filter lengths is increased from 101 to 241, 361 and 521. In Fig. 3.6, we show the training of the considered DPDs over the iterations for the case of uniform 64-QAM signal. We observe that all DPDs converged within 9 iterations. In these results, the DACs output voltages were set to optimum values for the corresponding DPD as described in the following subsection.

3.5.3. DAC VOLTAGE VARIATION

We vary the output voltage of the DACs to change the strength of the transmitter non-linearity. We re-train each of the DPDs for every value of the DACs’ output voltage. In Fig. 3.7(a), the SNR values of each DPD technique are plotted against the variation in the DACs’ output voltage. We observe that with a linear DPD employed at the transmitter,

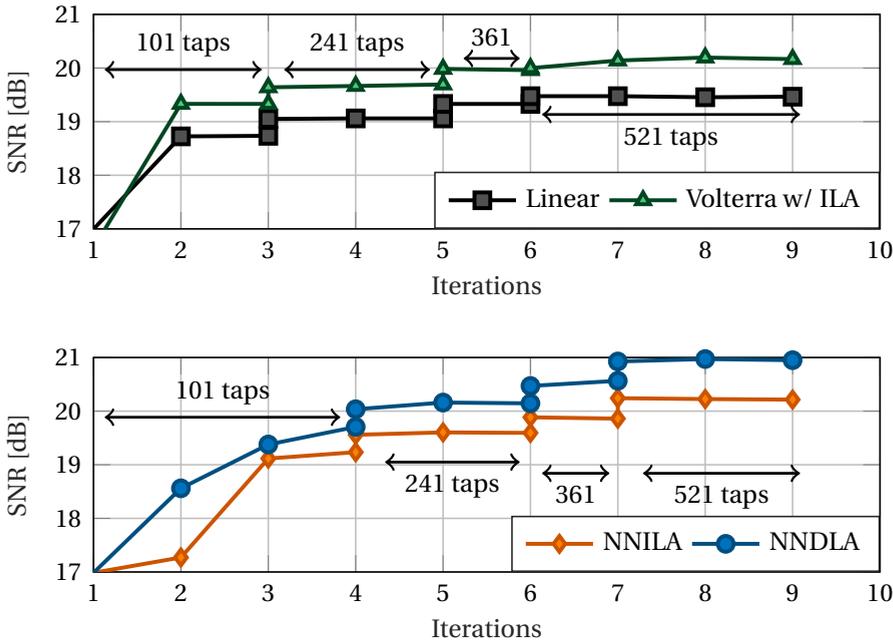


Figure 3.6: Training of different DPDs at 460 mV DACs' voltage by increasing the memory length of the 4×4 real-valued MIMO equalizer in the receiver DSP.

the SNR decreases with increasing the DACs' output voltage due to the increased nonlinear distortions from the transmitter. For linear DPD, the optimum operating point is around 300 mV.

Further, applying nonlinear DPDs at the transmitter gives improvements in the SNR even at a lower DAC output voltage of 300 mV. This shows the presence of significant nonlinear distortions at the transmitter even at low voltage. Furthermore, we see that unlike linear DPD, increasing the DACs' output voltage beyond 300 mV improves the SNR. The optimum DACs' voltage for nonlinear DPDs is approximately 450 mV which is 50% higher than that of the linear DPD. The NN-based and the Volterra series-based DPD, both trained using the ILA architecture, attain almost similar performance. Although the SNR gains obtained by using the former is slightly higher. Moreover, the NNDLA (i.e. the NN-based DPD trained using DLA) provides the highest gain in SNR which is 1.6 dB with respect to the linear DPD, after 9 iterations of Tx DPD training. NNDLA DPD obtains better SNR values than NNILA mainly because of the difference in the training architectures: DLA and ILA. A detailed explanation of this point can be found in [272].

At this point, we show how the proposed NN-DPD is different from WH architecture. For this purpose, the WH-DPD is obtained from reconfiguration of the proposed NN-DPD as follows: The section (A) and (C) and the short-cut connection between them is kept as it is. While, we remove the memory from the section (B) as the nonlinear part of a typical WH structure is memoryless. In detail, the section (B) was modified by removing

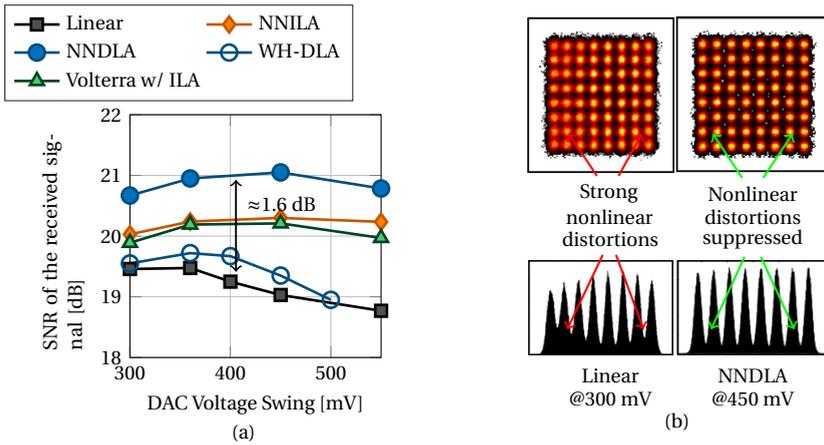


Figure 3.7: (a) Performance of the considered DPDs at different DACs' voltages with uniform 64-QAM. (b) Constellations and histograms of the soft symbol outputs of the receiver DSP y after the convergence, for the Linear and the NNDLA DPD.

the CNN layer (FFNN layer 1) i.e. the layer with 21 11-tap convolutions such that the output from the section (A) fans out directly to the second layer (FFNN layer 2, the layer with 12 neurons) of the section (B). Overall, the fully connected FFNN part has layers with 12, 8, 8, 1 neurons with leaky ReLU activation function. The WH-DPD was trained using DLA by using an auxiliary channel whose architecture is identical to the one used for NN-DLA. The SNR performance at different DAC voltages is shown in the Fig. 3.7. We see that WH-DPD only adds up to 0.2 dB gain in the SNR to the linear DPD. This also shows that the nonlinearity is mixed with memory in the system and the proposed NN-DPD architecture captures this nonlinearity mixed with memory.

Remark: We also tried a Volterra-based DPD trained using DLA. In order to have a fair comparison, we used the auxiliary channel-NN of NNDLA as a surrogate for training the Volterra-DLA-DPD. The Volterra-DLA was implemented within the NN-framework. This has advantage as the Volterra-DLA-DPD-NN can be trained easily by using the auxiliary channel-NN in the same manner as done for the NN-DLA. The Volterra-DLA-DPD NN takes all possible Volterra terms as its input features and learns the required weights and bias in order to produce the output. A batch normalization is applied at the output. Surprisingly, we did not observe significant gains with Volterra-DLA-DPD-NN. At 460 mV DACs' voltage, the Volterra-DLA-DPD-NN added only around 0.1 dB gain in the SNR in comparison to Volterra-ILA. The Volterra-DLA needs more investigation.

Next, we visualize the impact of applying the NNDLA at the transmitter by carrying out a spectral analysis and by plotting the signal constellation after the entire Rx-DSP. In Fig. 3.7(b), we plot the constellation diagram of one of the polarizations of the experimentally obtained signals along with corresponding histogram of one dimension. It can be clearly seen that when only a linear DPD is applied, the received signal is contaminated with power dependent distortions. In contrast, these distortions are suppressed and not visible in the received signal when the NNDLA DPD is applied at the transmitter.

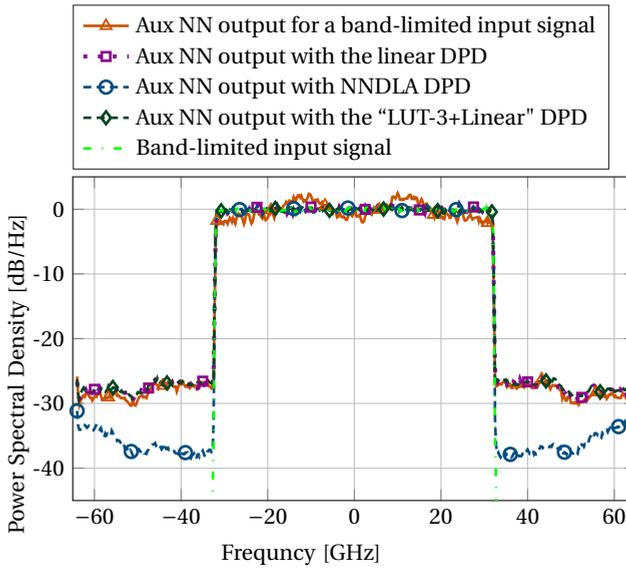


Figure 3.8: Noiseless spectral analysis of a brick-wall band-limited 64 Gbaud 64-QAM (at 2 sps) signal using experimentally trained NN models shows the transmitter distortions and its compensation. With the band-limited 64-QAM signal (dashed, green curve) as input, there is out-of-band noise at the output of auxiliary NN due to the nonlinearity (solid, orange curve). With NNDLA DPD (densely dash-dotted, blue curve) this nonlinear noise is suppressed.

Fig. 3.8 presents a spectral analysis of the output signal of the auxiliary NN when there is a test signal is applied as input. The test signal is a 64 Gbaud 64-QAM signal upsampled to 2 sample per symbol (sps) by zero-insertion and filtering using a brick-wall filter of 64 GHz bandwidth. We see that when the test signal is applied, without any DPD, to the trained auxiliary NN model, the auxiliary NN output has a distorted in-band spectrum along with out-of-band ($> |32|$ GHz) spectral components resulted from the nonlinearity. When the linear DPD is applied to the test signal before feeding it into the auxiliary NN, the output signal of the auxiliary NN has only its in-band spectrum corrected while the out-of-band spectrum generated by nonlinearity stays as is. On the other hand, when the test signal is passed through the NN-DPD and then fed to the auxiliary NN, the auxiliary NN output has a flat in-band spectrum together with a suppressed out-of-band spectrum.

We have also trained and tested a look-up table (LUT)-based DPD using this auxiliary NN model of the transceivers. We consider a the LUTs of memory 3 i.e. a correction to a symbol is based on that symbol and its adjacent symbols. The correction coefficients of the LUTs were learned by the method given in [273], [274] while the trained "Linear" DPD of 401 taps was applied in order to compensate for the large channel memory. This combined DPD is referred as "LUT-3+Linear" DPD. Two LUTs were trained each for the in-phase and the quadrature-phase tributary at 1 sps for a 128 Gbaud 64-QAM signal. We compared the performance of the "LUT-3+Linear" DPD with that of the Linear and the NNDLA DPD in terms of the normalized mean square error (NMSE) between the transmit and the auxiliary NN output symbols. We observed that "LUT-3+Linear" gave 1.6 dB smaller NMSE than the Linear DPD. The NNDLA gave about 10 dB smaller NMSE

than the “LUT-3+Linear” DPD. The NMSE difference between the “LUT-3+Linear” and the NNDLA shows that nonlinearity is mixed with the memory. One may expect more compensation gains by increasing the LUT memory. However, the current LUT has already $8^3 = 512$ entries and its size grows exponentially with memory if implementation is not optimized. The spectrum of the auxiliary output signal when the test signal with “LUT-3+Linear” DPD is fed at its input is also shown in Fig. 3.8.

In the next experiment, we test the DPDs using the PCS-256-QAM format and quantify the SNR performance. Fig. 3.9 shows the SNR values obtained for different DAC voltages. We observed a trend similar to the uniform 64-QAM format. The SNR gain that NNDLA provides with respect to the linear DPD is approximately 1.2 dB.

3.5.4. VERIFICATION OF PATTERN INDEPENDENCE

A common concern in the NN-based techniques is dependence on patterns. We applied our proposed NNDLA predistortion on 100 statistically independent symbol sequences which were not used in the training. The corresponding 100 pre-distorted waveform were transmitted through the experimental setup in back-to-back configuration with 450 mV DAC voltage and their performances were evaluated. In Fig. 3.9(b), a histogram of the observed 2-D MI values is plotted. A fitted curve with Gaussian approximation shows that the standard deviation is very small of around 0.014 bits/symbol/polarization indicating that the NNDLA is nearly pattern independent.

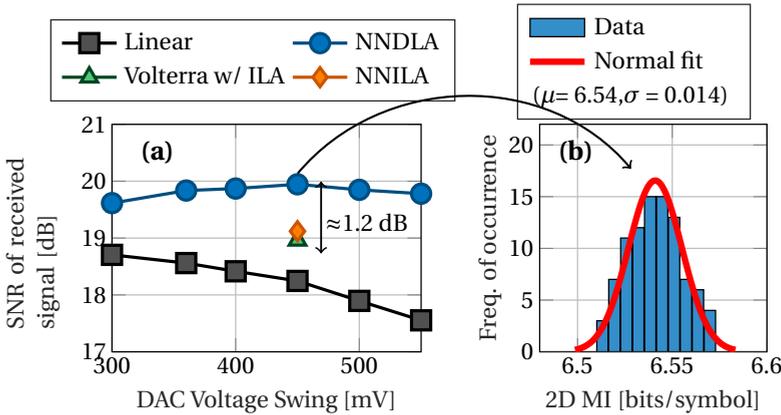


Figure 3.9: (a) Received signal SNR over a variation in DACs' voltage for PCS 256 QAM, (b) A histogram of the 2-D MI values obtained by transmitting 100 statistically independent symbol sequences.

3.5.5. EVALUATION IN THE FIBER TRANSMISSION SCENARIO

In further investigations, we apply our trained NNDLA and the linear DPD at the transmitter with PCS 256-QAM and test it over a link of 80 km SSMF. Fig. 3.10 shows the SNR, GMI and net rate of the received signal over different transmit powers. The figure indicates that the optimum launch power is around 6 dBm for both DPDs. Furthermore, applying NNDLA at the transmitter results into a significant SNR gain in comparison to the linear DPD. At the optimum launch power, the SNR gain is around 1.2 dB which

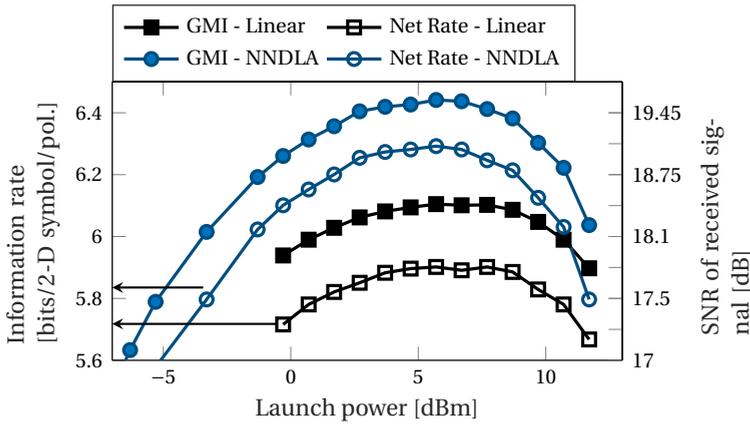


Figure 3.10: SNR, GMI and net information rate at different launch powers for 80 km SSMF transmission of PCS 256-QAM format signal pre-distorted using the Linear or the NNDLA DPD.

is the same as observed in experiments with the back-to-back configuration for PCS 256-QAM. The NNDLA transmission achieves GMI of 6.44 bits/symbol/polarization in comparison to 6.1 bits/symbol/polarization obtained by applying the linear DPD. Furthermore, the net rate increases from 5.9 bits/symbol/polarization for the linear DPD to 6.3 bits/symbol/polarization for the case of the NNDLA. Moreover, we observed that the FEC decoding loss is slightly less when the NNDLA is applied. This is attributed to the more Gaussian-like distribution of the soft symbols at the Rx-DSP output (y in Fig. 3.7(b)) when using the NNDLA instead of the linear DPD. The lower decoding loss is due to the assumption of conventional FEC decoding algorithms that received symbols follow a Gaussian likelihood. Overall, our proposed NNDLA increases the net bit rate to a record 1.61 Tb/s over a single-channel of 80 km SSMF.

3.6. COMPUTATIONAL COMPLEXITY AND PRUNING

In this section, we compare the computational complexity of each of the considered DPDs. As a figure of merit, we compute the required number of real-valued multiplications to implement each DPDs. The number of real-valued multiplications per stage of the NN-based DPDs is provided in Tab. 3.3. The batch normalization layers and soft-DAC are not accounted for in the computation as similar processing is required for the other DPDs in order to generate integers prior to loading the DACs.

The linear DPD has 441 coefficients, thus, require 441 real-valued multiplications for each channel (I/Q). The Volterra filter has three kernel orders requiring 441, 105 and 99 coefficients for the first, second and third order respectively. Additionally, 14 real-valued multiplications are needed to generate the second and third order terms. By following [275], we consider that already computed lower-order Volterra terms are used to generate other possible higher-order terms so that the real-valued multiplications for Volterra DPD are not over-counted. In total Volterra implementation requires 659 multiplications. Tab. 3.4 summarizes the complexity. We see that our proposed NN-DPD requires about 64% more real-valued multiplications than the Volterra DPD.

Table 3.3: Real-valued multiplications per channel needed for NN-based DPD.

NN layer	Real-valued multiplications
First CNN layer	101
Last CNN layer	301
CNN layer before FFNN	$11 \times 21 = 231$
FFNN layer 1	$21 \times 12 + 12 = 264$
FFNN layer 2	$12 \times 8 + 8 = 104$
FFNN layer 3	$8 \times 8 + 8 = 72$
FFNN layer 4	$8 \times 1 = 8$
Total	1081

Table 3.4: Real-valued multiplications needed per channel for each of the considered DPDs.

DPD type	Real-valued multiplications
Linear	441
Volterra	659
NNDLA/NNILA	1081

To understand performance-complexity trade-off, we reduce the complexity of NN-based DPD by pruning it after the NNDLA has converged in the experiments. We prune only the middle FFNN structure (ie. the section (B) from Fig. 3.3) as the other layers are linear and are common to other DPDs as well. The pruning method proposed in [276] was applied on each channel (I/Q) separately. We used L1-norm as pruning criteria such that the smallest weights and biases in the FFNN layer are forced to zero after pruning. A target or final pruning factor s_f is achieved in N training steps or epochs by pruning with a s_c factor every ΔN epochs. The pruning factor for a given epoch s_c is given by the following relation

$$s_c = s_f + s_f \left(1 - \frac{\lfloor n/\Delta N \rfloor}{N} \right)^3, \tag{3.5}$$

We used $N = 20$ epochs to achieve a final pruning factor where pruning by s_c was applied every ΔN epochs. Pruning reduces the complexity by avoiding several multiplications. It is even possible that all the weights feeding a neuron are zeroed by the pruning process, thereby further reducing the complexity as the computation of the activation of that neuron can then also be omitted. For our case, the decrease in the complexity due to deactivated neurons is very small, and thus, ignored. The decrease in the number of real-valued multiplications in the middle FFNN structure as a result of pruning is considered as $651 \times s_f + 28$.

We also pruned the nonlinear part of the Volterra w/ ILA DPD kernels i.e. the coefficients of the second and the third order. The pruning was again done by forcing the smallest magnitude coefficients to zero. In Fig. 3.11, we plot the performance of NNDLA

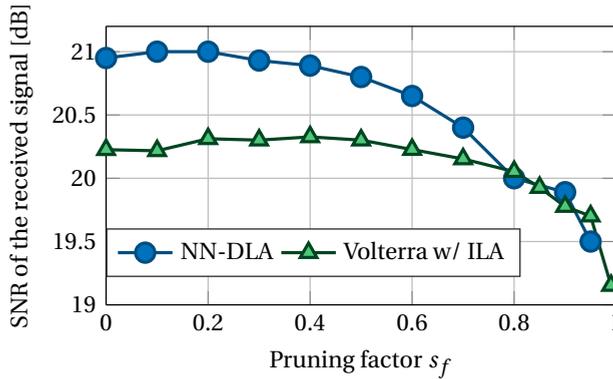


Figure 3.11: SNR of the received signal for different pruning factor s_f applied on nonlinear parts of NNDLA and Volterra w/ ILA DPDs. The number of real-valued multiplication reduces by $651 \times s_f + 28$ and $218 \times s_f$ for NNDLA and Volterra DPD, respectively.

and Volterra DPD over the experimental setup at various pruning factors. For Volterra DPD, a pruning up to a factor of 0.6 does not add any penalty, instead, we observe some improvement (0.1 dB) in the SNR when pruning factor is around 0.3. This is attributed to the fact that a reduced number of Volterra kernel increases the accuracy of the least squares based Volterra DPD. At pruning factor s_f of 0.6, Volterra DPD has around 530 kernel coefficients.

For NNDLA, we see that a pruning by a factor of 0.2 can be applied without causing performance degradation, while larger pruning factors add penalty to the received SNR. At pruning factor of 0.8 both DPDs have similar performance while it requires 560 and 485 real-valued multiplications for NNDLA and Volterra DPD, respectively. A pruning by a factor of 0.4 still gives good performance while reducing the overall per channel complexity of the NNDLA to around 820 real-valued multipliers.

3.7. CONCLUSION

In this paper, we reported on a new record transmission of 1.61 Tb/s data rate over a single channel of 80 km of standard single mode fiber that was achieved using a novel neural network-based digital pre-distorter. The proposed DPD has been compared with a Volterra series-based ILA-DPD and a linear DPD. In addition, we evaluated the performance of the proposed DPD by training it using direct learning and indirect learning architecture. The NN-based DPD trained using DLA adds SNR gain of around 1.6 dB and 1.2 dB with respect to a linear DPD for uniform 64-QAM and PCS 256-QAM formats, respectively. Further, we show that by applying pruning the computation complexity of the proposed DPD can be reduced significantly with no or only minor losses in the SNR.

4

54.5 TB/S WDM TRANSMISSION OVER FIELD DEPLOYED FIBER ENABLED BY NEURAL NETWORK-BASED DIGITAL PRE-DISTORTION

We demonstrate a record 54.5 Tb/s WDM transmission at 11.35 bit/s/Hz over 48 km of field-deployed SMF connecting business and academic parks enabled by a novel joint I-Q Neural Network-based transmitter digital pre-distortion technique.

The text in this chapter has previously appeared in modified form in V. Bajaj, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion". In *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, San Francisco, CA, USA, 2021, pp. 1-3. [265]

4.1. INTRODUCTION

THE rapid growth in internet traffic is driving the need for high data rate optical fiber transmission systems. To increase the aggregate bit rate carried by a single fiber strand while minimizing the component count, these systems need to be operated at high symbol rates on a tight spectral grid using high-order modulation formats (HOMFs) while maintaining the highest possible signal-to-noise ratio (SNR) [93], [237]–[239]. Transmitter digital pre-distortion (DPD) techniques enable the generation of signals with high integrity by mitigating undesired component responses. Usually, a linear DPD is used. In such cases, the uncompensated nonlinear response of components still adds distortions and limits the SNR. As HOMFs are sensitive to distortions, addressing these distortions at the transmitter in a way that is tailored to the system and its components is paramount to increase the information rate (IR).

DPD techniques based on neural networks (NN) [253] have recently received more attention [88], [255], [259], [260], [262], with notable simulation results in [88] by implementing NN-based DPD using direct learning architecture (DLA) for optical coherent systems. We recently demonstrated a NN-based DPD technique that independently mitigates the distortions of the “I” and “Q” tributaries on a single channel fiber-optic transmission system where an SNR improvement of up to 1.6 dB was achieved [264]. Another recent in-lab demonstration was of a total C-band capacity of 52.1 and 51.6 Tb/s over 80 km of TXF™ fiber and standard single mode fiber (SSMF), respectively [277].

In this paper, we extend our previous work [264] by developing a NN-based DPD technique to *jointly* pre-compensate for the distortions on the I and Q tributaries of the transmitter. The joint DPD is learned from “scratch” on a field-deployed 48 km link connecting industrial, business and academic parks in a metropolitan area. The DPD is applied to a fully populated C-band, consisting of 35 channels spanning a total bandwidth of 4.8 THz. It improves the SNR by about 1 dB compared to a linear DPD, and allows achieving a net capacity of 54.5 Tb/s.

4.2. PROPOSED NEURAL NETWORK-BASED JOINT DPD

We depict the architecture of the proposed NN-based DPD in Fig. 4.1. The upper and lower parts belong to the I and the Q tributaries respectively and have identical structure. The DPD comprises three sections. Section (A) contains a 2×2 multiple-input multiple-output (MIMO) composed of linear uni-dimensional convolution filters (1D-CNNs) of 101 taps intended to mitigate the linear response of the IQ modulator (through and cross). Section (C) contains two linear (1D-CNN) filters of 301 taps meant to compensate the linear effects of the digital to analog converters (DACs) and their driver amplifiers (DAs). The two feed-forward NNs (FFNNs) in Section (B) are each composed of five dense fully connected layers. The first layer is a linear 11 tap 1D-CNN layer that fans out the digital signal. The following 3 layers have leaky rectified linear unit activations (ReLUs). The last layer has a single linear output. The sizes of all layers are shown in Fig. 4.1. As the linear effects are dominant, adding a shortcut bypassing the FFNNs boost the performance and the training speed [264]. We also model the DAC using a customized activation unit, named Soft-DAC, that scales, clips and quantizes the signal in order to minimize the quantization distortion. We refer to [264] for more details.

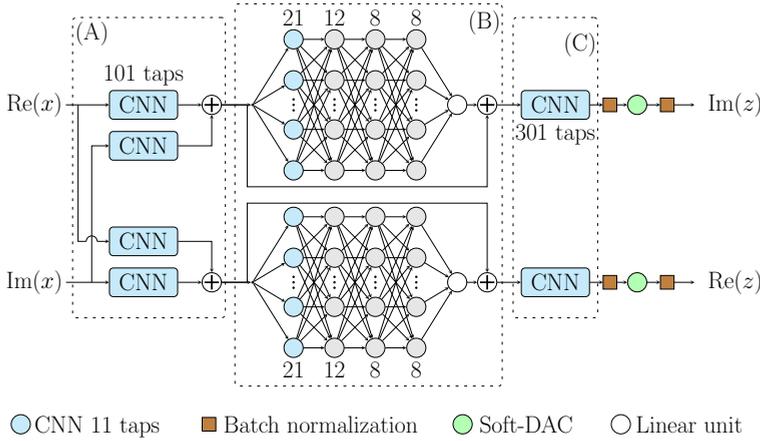


Figure 4.1: Architecture of the proposed joint I-Q NN-based DPD.

As depicted in in Fig. 4.2, we used a DLA to train the DPD. We model first the “communication channel” using a mirror version “(C) \rightarrow (B) \rightarrow (A)” of the differentiable DPD structure, serving as an auxiliary channel model (Aux). The DPD and the Aux are trained iteratively until the SNR converges; about 10 iterations in our experiments. The input to the Aux is the DPD output $z[n]$, which is also the input to the channel. A coherent receiver digital signal processing (Rx-DSP) is employed to stabilize the dynamic response of the channel. The Aux is trained to minimize the mean squared error (MSE) between its output and the soft symbol outputs of the Rx-DSP $y[n]$. The DPD is trained by using ideal QAM symbols $x[n]$ as its input and the gradients of the MSE between the Aux output and the desired output of the channel (i.e. ideal QAM symbols $x[n]$), back propagated through the Aux. The both NNs are optimized using Adam algorithm by using data sequences of 2^{18} length at 1 sample per symbol (sps).

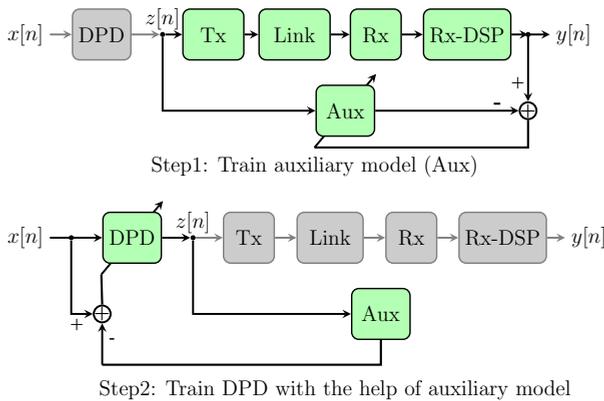


Figure 4.2: Direct learning architecture (DLA).

4.3. FIELD TRIAL EXPERIMENTAL SETUP

The experimental setup, shown in Fig. 4.3, is configured for the WDM transmission of 35 channels equally spaced by 137.5 GHz at 128 Gsymbol/s. The transmitter has two IQ modulators: one for the channel under test (CUT) and one for the bulk modulation of the other 34 WDM channels. All channels employ an external cavity laser (ECL) of 20 kHz linewidth. The inverted and non-inverted outputs of two DAC modules sampling at 128 GSa/s are amplified before driving the two IQ-modulators [238]. Dual-polarization (DP) is emulated using two separate delay and add interferometers. An SSMF of 5 km in the WDM branch decorrelates the channels. Both branches are amplified before combination via a Waveshaper, also serving as a per-channel power equalizer to compensate for the modulator's response. The 35 channel WDM optical signal is amplified and launched into a SSMF interconnecting an industrial and business area to an academic and research park around Stuttgart, Germany. The link spans 48 km but exhibits a total loss of 16 dB, comparable to DCI-type distances of 80 km. The Maxwell-Boltzmann probabilistic constellation shaping (PCS) of a 256-QAM format with an entropy of 7.5 bits/symbol is employed, as previously determined as a good choice for roughly 19 dB SNR [278].

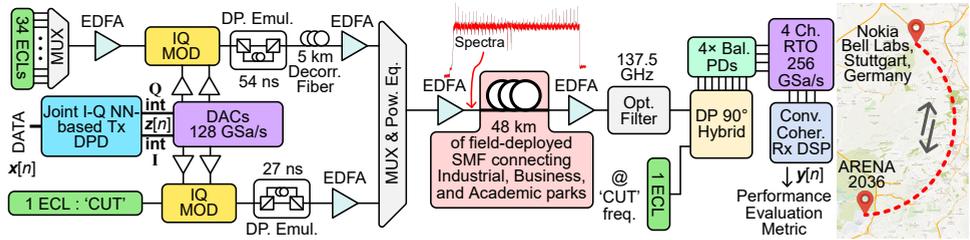


Figure 4.3: Experimental setup of the WDM transmission system. Right: map shows the field trial area.

At the receiver (Rx), the signal is first amplified before the CUT is isolated by a tunable optical filter. A local oscillator (LO), another ECL of 20 kHz linewidth, beats with the CUT through a DP 90° optical hybrid. A 4-channel real-time oscilloscope (RTO) digitizes the 4 balanced photo-detector outputs at 256 GSa/s. The Rx-DSP processes the 4 digitized signals in the following stages. First, the signal is re-sampled at 2 sps. Then chromatic dispersion is removed and timing errors are corrected. The signal polarizations are de-multiplexed by using a 2×2 complex valued MIMO equalizer updated by a multi-modulus algorithm (MMA). Frequency detuning correction and phase noise correction operations are subsequently applied. The residual signal distortions are then compensated by a real valued 4 × 4 MIMO equalizer[237]. For decoding, we used variable overhead SC-LDPC codes described in [237].

4.4. RESULTS

Datasets obtained by repeated transmission of a 2^{15} long PCS 256-QAM pre-distorted pattern were used for training the NN-based DPD and the linear DPD. The DACs voltages were set to 300 mV and 470 mV for the linear and the NN-based DPD, respectively,

which were found optimal in our previous study [264]. First, both DPDs were trained when the real valued MIMO equalizer was not included in the Rx-DSP. In this case, the I-Q impairments are not compensated by the Rx-DSP and the NN-based joint I-Q DPD learns to mitigate them at the Tx, providing a gain of 3 dB with respect to the disjoint I-Q linear DPD. Next, when the NN-based joint I-Q DPD is employed and the real-valued 4×4 MIMO stage is included in the Rx DSP stack, the SNR after the MIMO stage is 1 dB higher compared to when only a disjoint I-Q linear DPD is employed. This not only indicates the significant gains provided by the NN-based joint I-Q, but also that the joint I-Q DPD can learn to compensate for linear IQ cross-talk that would otherwise need to be compensated at the receiver using noisy receiver inputs.

Fig. 4.4(a) shows the achievable bit rate and the corresponding generalized mutual information (GMI) obtained when the CUT is swept through the 35 channels. We observed that the NN-DPD outperforms the linear DPD by giving a considerable gain in GMI. The average GMI (in bits/symbol/pol./carrier) is increased from 5.95 for the linear DPD to 6.31 for the NN-DPD. For the linear DPD, we achieved an average net bitrate of 1.45 Tb/s, which is slightly lower than what was obtained in [238]. We attribute this to the greater impairments from the 48 km of fibers criss-crossing the city compared to 80 km of spooled fiber in a lab. To verify that the NN did not learn the transmit pattern, we applied 100 different random sequences of PCS-256QAM symbols to the NN-DPD and measured their performance (Fig. 4.4(b)). The variance in achievable bitrate is very small, <1% (± 10 Gb/s) at most. The variations in GMI and net bitrate versus channel number are attributed to the EDFAs' gain profile. Moreover, the average decoding loss (in bits/symbol/pol./carrier) when employing the NN-DPD is reduced to 0.22 compared to 0.27 for the linear DPD, thanks to the more Gaussian-like distribution of the soft symbols at the Rx-DSP output ($y[n]$ in Fig. 4.2) when using the NN-DPD instead of the linear DPD (see Fig. 4.4(c)), since the FEC decoding algorithm conventionally assumes a Gaussian likelihood for the received symbols. The average net IR increases from 5.68 to 6.09 bits/symbol/pol./carrier, which corresponds to an average net bit rate of 1.55 Tb/s per channel. Overall, the 35 WDM channels result in a net transmission capacity record of 54.5 Tb/s, which is 7.2% higher than the capacity achieved by the linear DPD. Finally, we assessed the stability of the system performance when using a trained and fixed NN-based DPD. Fig. 4.4(d) shows the performance of 1 channel (out of 35) over the course of 3 days, with measurements taken every 20 minutes. The fluctuations are very small, even at a high GMI of 6.31 bits/symbol/pol., indicating that the learned Tx NN-based nonlinear DPD is temporally stable.

4.5. CONCLUSION

We demonstrated a field trial of 35 channel WDM transmission employing a novel joint I-Q neural networks-based DPD (NN-DPD). The NN-DPD adds a gain of 0.41 bits/symbol/pol. in net information rate and enabled us to achieve a record net rate of 54.5 Tb/s at 11.35 bit/s/Hz spectral efficiency. The performance of the NN-DPD evaluated over time shows its good stability.

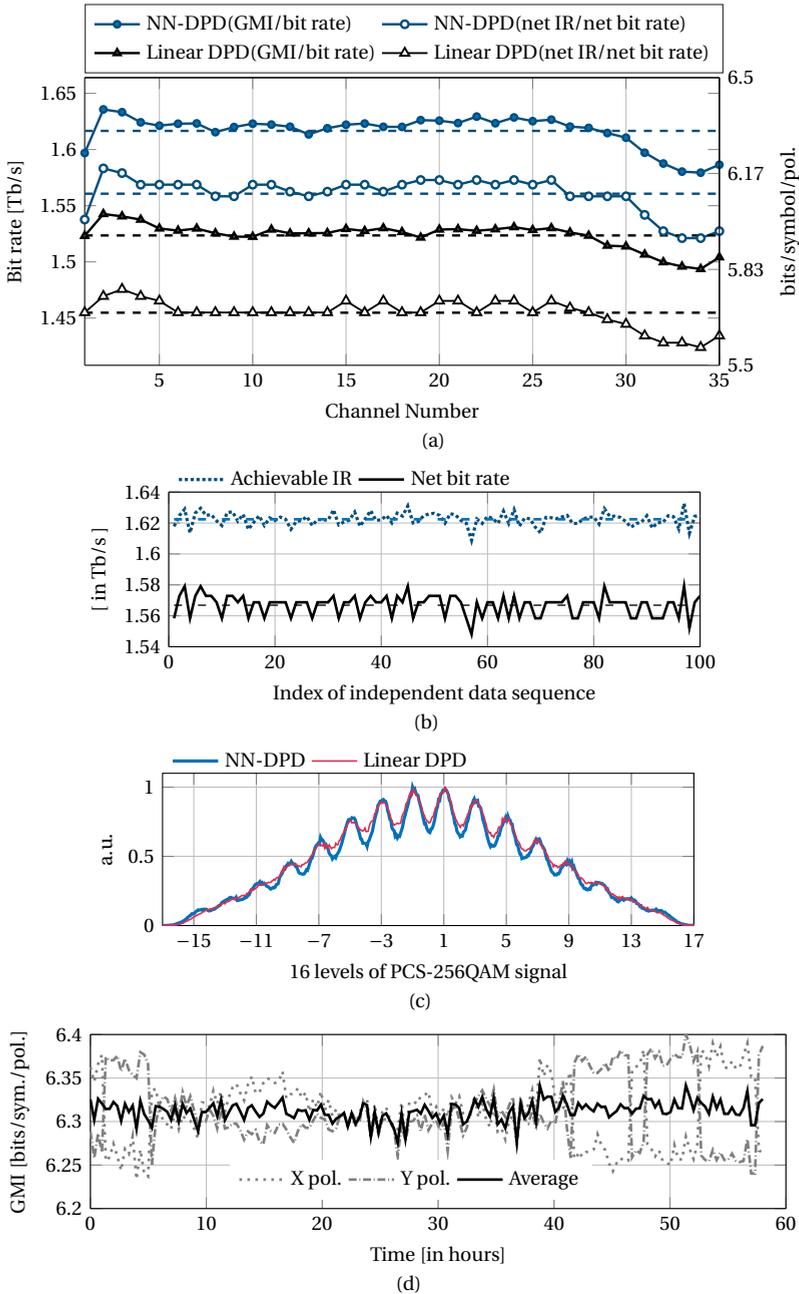


Figure 4.4: (a) Data rates and information rates at different WDM channels (dashed lines: average values over the C-band), (b) achievable information rate and net bit rate over a test of 100 statistically independent sequences, (c) histograms of averaged 4-dimensions of the soft symbols output from the Rx-DSP and (d) GMI of the soft symbols output from the Rx-DSP plotted over time, when NN-DPD is employed at Tx.

5

PERFORMANCE ANALYSIS OF RECURRENT NEURAL NETWORK-BASED DIGITAL PRE-DISTORTION FOR OPTICAL COHERENT TRANSMISSION

A recently developed neural network (NN)-based digital pre-distortion method for a high baud rate (128 GBaud) optical coherent transmitter utilized a feed-forward architecture. In this paper, we investigate the performance of recurrent NN architectures for this task.

5.1. INTRODUCTION

ONE current approach to address growing capacity demand is to increase the symbol rate and to use higher order modulation formats in optical fiber communication systems[237]. A main challenge is though the transmitter impairments[87]. For higher symbol rates and modulation formats, the undesired nonlinear responses of transmitter components gets more severe and limits the quality of the transmit signal. Transmitter digital pre-distortion (DPD) techniques compensate these undesired component responses and increase signal integrity, maximizing the net data rate. Various DPD techniques have been investigated in the past. DPDs based on Volterra-series [247], its simpler versions [249] as well as look-up tables [274] were investigated.

In recent years, deep neural networks (NNs) are being used for DPD as well in optical communication systems[88], [257]. NNs are seen as capable to model complex non-

The text in this chapter has previously appeared in modified form in V. Bajaj, V. Aref and S. Wahls, "Performance Analysis of Recurrent Neural Network-based Digital Pre-Distortion for Optical Coherent Transmission". In *2022 European Conference on Optical Communication (ECOC)*, Basel, Switzerland, 2022, pp. 1-4. [279].

linear systems and also make it easier to implement direct learning architectures [88], [280]. For instance, a recurrent NN-based DPD was investigated over a simulated transmitter [88]. Furthermore, a LUT-based DPD was simplified by trading LUT storage with a FFNN-based DPD [90]. Reinforcement learning was also applied to learn DPD without deriving a model of the transmitter [281]. In our previous work, a NN-DPD architecture with a convolutional NN (CNN) and a feed-forward NN (FFNN) part was investigated in experiments over a single channel [264] and wavelength division multiplexing (WDM) field trial experiments [265]. It was observed that the nonlinear effects have memory and require a DPD structure that can capture such effects [89]. Recurrent NNs have the capability to model nonlinearities with very long memory in time-series data. In optical fiber communications, RNNs have shown improvement over feed-forward architectures for nonlinear equalizers [219] and end-to-end transceiver optimization [220]. Thus, RNNs can also be advantageous for DPD applications.

In this paper, we investigate bidirectional RNNs for DPD using experimentally trained NN transmitter models based on our previous experiments [264]. We consider the three popular RNN structures and compare them against the previous non-recurrent architecture [264].

5

5.2. RNN-BASED DPDs

In our previously proposed DPD architecture [264] (shown in Fig. 5.1) two large CNNs ($\ell_1 = 101, \ell_2 = 301$) that capture large memory effects surround a FFNN that deals with nonlinearity, while a structure with CNN followed by multi-layer perceptrons (MLPs), i.e. CNN+MLP (referred to as FFNN structure in [264]) in the middle captures the nonlinearity mixed with memory.

In this investigation, we replace the central CNN+MLP structure with a RNN+MLP structure for a fair comparison and evaluate performance and complexity. We consider three candidate RNNs for this purpose, namely vanilla RNN (vRNN), LSTM and GRU which were implemented in PyTorch deep learning toolbox [282]. The considered configurations of DPDs for this comparative study are shown in Fig. 5.2. Similar to the previous architecture, these RNN-based DPDs were implemented using real-values, i.e., separate DPDs are used for I and Q tributaries.

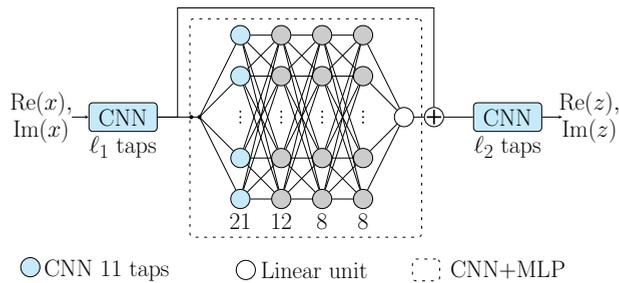


Figure 5.1: Architecture of feed-forward NN-based DPD ($\ell_1 = 101, \ell_2 = 301$) and auxiliary model ($\ell_1 = 301, \ell_2 = 101$).

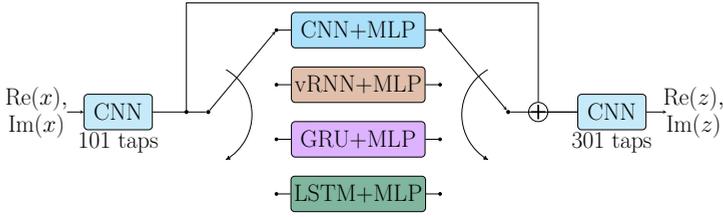


Figure 5.2: Architecture of NN DPDs derived with either CNN+MLP or considered RNNs+MLP.

For a real-valued input sequence sample x_t at time t , the output hidden state is

$$\vec{h}_t = \tanh(W_{ih}x_t + b_{ih} + W_{hh}\vec{h}_{t-1} + b_{hh}), \quad (5.1)$$

where \vec{h}_{t-1} is the previous hidden state vector and n_h is the size of hidden states. The trainable parameters of a vRNN cell are the weight matrices $W_{ih} \in \mathbb{R}^{n_h}$, $W_{hh} \in \mathbb{R}^{n_h \times n_h}$ and the bias matrices $b_{hh} \in \mathbb{R}^{n_h}$, $b_{ih} \in \mathbb{R}^{n_h}$. For the reverse direction, the hidden states, denoted by a vector arrow pointing from right to left or an r in the superscript, are given by

$$\vec{h}_t = \tanh(W_{ih}^r x_t^r + b_{ih}^r + W_{hh}^r \vec{h}_{t-1} + b_{hh}^r), \quad (5.2)$$

where the input x_t^r is $x_{L-(t-1)}$ and L is the length of the sequence fed into vRNN cell. The hidden state from both directions at time t are combined by averaging to generate an averaged hidden state at time t , i.e. $\bar{h}_t = 0.5(\vec{h}_t + \vec{h}_t^r)$. In a vRNN, the averaged hidden state is then passed to a following MLP structure. LSTM[283] and GRU[284] cells are also bidirectional. Their hidden states from both directions at time t are averaged (as explained for the vRNN cell above) and passed on to the following MLP structure. In our implementation of LSTM, the input, forget and output gates use Sigmoid activation function and the cell gate uses hyperbolic tangent function. Finally, the RNN+MLP structure is integrated with the rest of the architecture as shown in Fig. 5.2.

A major portion of the computational complexity of NN is contributed by multiplications. Thus, we omit other operations and count the number of real-valued multiplications (RVMs) required to produce one sample (real or imaginary part of the symbol) at the DPD output for comparisons. We only measure and compare complexity of the central (RNN/CNN+MLP) structure.

We designed RNNs with a many-to-one sequence mapping [219], [285]. That is to produce the DPD output at time t , the RNN processes a block of input samples $x_t - L/2, \dots, x_t, \dots, x_t + L/2$ and hidden (and cell) states at $t - 1$ and $t + 1$, where L is the length of the sequence fed to RNN. It is also possible to design RNNs in a one-to-one sequencing fashion. In our case, we found that this method was not optimal for the tried hyperparameters and needed large hidden state size to achieve comparable performance. For the three RNN-based DPDs, the input feature size of the RNN n_i is 1 and n_h is the number of RNN layers. The complexity of a single RNN layer is given by $n_h(pn_h + q)DL$, where D is 2 for bidirectional case and the values of (p, q) are $(1, 1)$, $(3, 6)$ and $(4, 7)$ for vRNN, GRU and LSTM, respectively.

5.3. TRAINING AND HYPER-PARAMETER SEARCH

To assess the performance, we used auxiliary NN models of a 128 GBaud transmitter obtained from single channel experiments[264]. These models have the CNN+MLP architecture shown in Fig. 5.1 ($\ell_1 = 301, \ell_2 = 101$). Each of these models is trained at different DAC voltages by using probabilistically shaped (PS) 256 QAM symbols with Maxwell-Boltzmann statistics and 7.5 bits of entropy. These models serve as a proxy for a 128 GBaud coherent transmitter for training and evaluation of the considered RNN-based DPDs under investigation. The RNN-based DPDs were trained by cascading them to the front of the auxiliary model and a batch normalization layer in order to learn a pre-inverse of the system. The performance of the RNN-based DPDs is measured in terms of normalized mean square error (NMSE).

The training and test data are generated with different random seeds using the Mersenne Twister algorithm (period $2^{19937} - 1$) and random shuffling to ensure that NNs are not learning the underlying pattern to generate data [286]. To train the considered DPDs, $N = 128$ different sequences of PS 256QAM with length 2^{14} are generated. These 128 sequences are fed in batches of size $128 \times b_l$. For RNN+MLP DPDs, each sequence of length b_l in the batch (with size $128 \times b_l$) is split into segments of length 16, which is joined back to make after RNN processing as done in [220]. Thus at the RNN processing stage the batch size changes to $128 b_l / 16 \times 16$. This helped in reducing the computation time required to process each batch. An optimization of this split ratio is beyond the scope for this paper. A stochastic gradient descent with mean square loss criteria was used for optimization. The training hyper-parameters were obtained via exhaustive search. The learning rate l_r and batch size b_l were changed during the training in steps as per the following rule: The learning rate and batch size (l_r, b_l) are $(10^{-2}, 1024)$ at the start of the training. It is changed to $(10^{-3}, 1024)$ for $-30 \text{ dB} > \text{NMSE} > -32 \text{ dB}$, $(10^{-4}, 2048)$ for $-32 \text{ dB} > \text{NMSE} > -35 \text{ dB}$, and $(10^{-5}, 4096)$ for $\text{NMSE} < -35 \text{ dB}$, respectively.

The hyperparameters of the NN architecture are searched for over a grid. The number of hidden layers in RNN n_h and the sequence length of the RNN input L are fixed to 1 and 11, respectively. The size of the hidden state vector h is varied from 1 to 4. The MLP have neuron counts of hL for the input layer and one for the output layer. The first two intermediate layers of MLP have 21 and 12 neurons. The remaining intermediate layers have eight neurons. The MLP network had layer counts ranging from 1 to 9, where the later intermediate layers are first removed in order to decrease the layer count. The activation function for MLP was chosen as LeakyReLU (with a negative slope of 0.1), which is also used for CNN+MLP configuration. The last MLP layer (with neuron count 1) is without any activation.

5.4. RESULTS

We first measure the performance of the previously proposed CNN+MLP DPD over the auxiliary models for reference as a benchmark. These CNN+MLP NN-DPD models have the same architecture as in [264] (see Fig. 5.1 with $\ell_1 = 101, \ell_2 = 301$). For this work, the training was done over 100 iterations after which the performance is observed to be mostly saturated as shown in Fig. 5.3. In this figure, the CNN+MLP NMSE is obtained by averaging 12 different training realizations. The convergence of vRNN+MLP (conf. 1 in

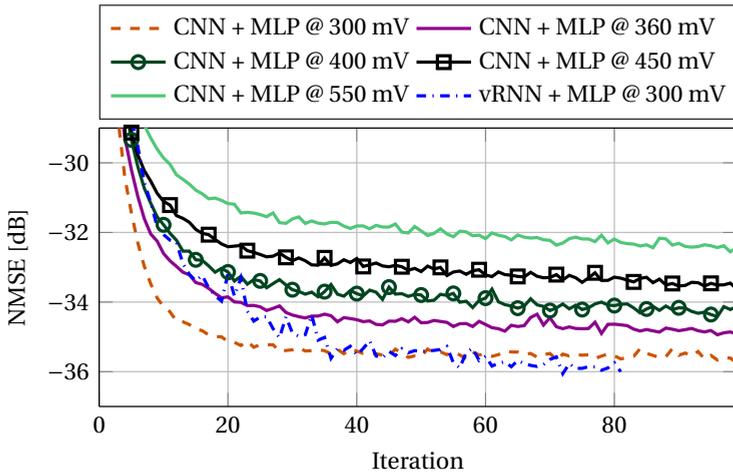


Figure 5.3: Convergence plot of CNN+MLP training at different DAC voltages.

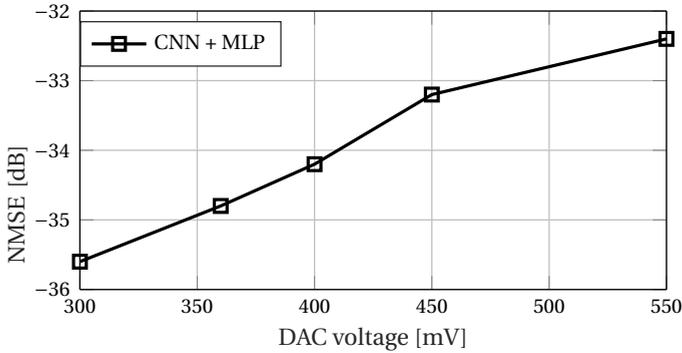


Figure 5.4: The final performance of CNN+MLP-based DPD over different DAC voltages.

Tab. 5.1) is also shown by dashed curve. The final performance of the CNN+MLP based DPD is shown over different DAC voltages in Fig. 5.4.

Among the various configurations of the three RNNs, the DPDs which achieved the target NMSE at all DAC voltages, we list the ones with lowest complexity in Tab. 5.1. The hyperparameters of the RNN and MLP are also tabulated together with the hidden state vector size h and the following MLP structure neuron counts.

We see that the lowest RVMs count is obtained by vRNN configurations, followed by GRU and LSTM configurations. We remark that a high multiplication requirement of LSTMs has also been observed for NN-based equalizers[219]. In our comparative study, the best performing vRNN+MLP configuration had a around 55% higher RVMs count in comparison to the CNN+MLP structure which uses 651 multiplications.

We see that a RNN+MLP configuration obtains performance below the target NMSEs, however, all the tried configuration used higher RVMs count than a simple CNN+MLP

Table 5.1: Hyper-parameters of RNN-based DPDs and their computational complexity per tributary.

RNN type	h	MLP	RVMs
vRNN	2	22, 21, 12, 8, 8, 1	1014
vRNN	2	22, 21, 12, 8, 8, 8, 1	1078
GRU	2	22, 21, 12, 8, 8, 8, 1	1474
GRU	2	22, 21, 12, 8, 8, 8, 8, 1	1538
LSTM	2	22, 21, 12, 8, 8, 8, 1	1606
LSTM	2	22, 21, 12, 8, 8, 8, 8, 1	1670

structure. Thus, in this case study, we find that that a simple MLP is good enough for DPD applications than typical RNN+MLP structure when computation complexity is desired to be low. Note that pruning can reduce the computation in some cases as shown in [89, sec. VI] However, we do not consider pruning of RNN-DPDs for this assessment.

5.5. CONCLUSION

In this paper, we investigated various bidirectional recurrent neural networks (RNNs) for the application of digital pre-distortion for optical coherent transmission. We compared vanilla RNN, LSTM and GRU against an existing previous CNN+MLP structure DPD. For the tried configuration of the different RNN types, we observed that CNN+MLP gave better performance when lower computational complexity is considered. However, there are more configurations that could be explored.

6

BLIND POLARIZATION DEMULPLEXING OF SHAPED QAM SIGNALS ASSISTED BY TEMPORAL CORRELATIONS

While probabilistic constellation shaping (PCS) enables rate and reach adaption with finer granularity [43], it imposes signal processing challenges at the receiver. Since the distribution of PCS-quadrature amplitude modulation (QAM) signals tends to be Gaussian, conventional blind polarization demultiplexing algorithms are not suitable for them [60]. It is known that independently and identically distributed (iid) Gaussian signals, when mixed, cannot be recovered/separated from their mixture. For PCS-QAM signals, there are algorithms such as [63], [64] which are designed by extending conventional blind algorithms used for uniform QAM signals. In these algorithms, an initialization point is obtained by processing only a part of the mixed signal, which have non-Gaussian statistics.

In this paper, we propose an alternative method wherein we add temporal correlations at the transmitter, which are subsequently exploited at the receiver in order to separate the polarizations. We will refer to the proposed method as frequency domain (FD) joint diagonalization (JD) probability aware-multi modulus algorithm (pr-MMA), and it is suited to channels with moderate polarization mode dispersion (PMD) effects. Furthermore, we extend our previously proposed JD-MMA[288] by replacing the standard MMA with a pr-MMA, improving its performance. Both FDJD-pr-MMA and JD-pr-MMA are evaluated for a diverse range of PCS (entropy \mathcal{H}) over a first-order PMD channel that is simulated in

The text in this chapter has previously appeared in modified form in [287].

©2023 IEEE. Reprinted, with permission, from V. Bajaj, R. Van de Plas and S. Wahls, "Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations". In *Journal of Lightwave Technology*, vol. 42, no. 2, pp. 560-571, Jan. 2024, doi: 10.1109/JLT.2023.3315370

a proof-of-concept setup. A MMA initialized with a memoryless constant modulus algorithm (CMA) is used as a benchmark. We show that at a differential group delay (DGD) of 10% of symbol period T_{symp} and 18 dB SNR/pol., JD-pr-MMA successfully demultiplexes the PCS signals, while CMA-MMA fails drastically. Furthermore, we demonstrate that the newly proposed FDJD-pr-MMA is robust against moderate PMD effects by evaluating it over a DGD of up to 40% of T_{symp} . Our results show that the proposed FDJD-pr-MMA successfully equalizes PMD channels with a DGD up to 20% of T_{symp} .

6.1. INTRODUCTION

PROBABILISTIC constellation shaping (PCS) is a tool that allows for rate and reach adaption with finer granularity than a uniformly distributed quadrature amplitude modulation (QAM) signal [43]. It was shown recently that conventional blind digital signal processing (DSP) algorithms, which were originally developed for uniform QAM signals, are not suitable to equalize PCS-QAM signals from distortions such as frequency offset [69], phase noise [98], and polarization impairments [60], [289]. While it is possible to use pilot-aided algorithms, they require additional overhead in the data payload. There is thus a need for blind signal processing algorithms that can deal with PCS-QAM signals.

This paper is concerned with blind algorithms for the compensation of polarization impairments. Already three decades ago, it was known that the commonly used constant modulus algorithm (CMA) is not suited for shaped sources [p. 136][60]. PCS-QAM signals tend to have a Gaussian distribution. However, separating a mixture of two independently and identically distributed (iid) Gaussian signals into the original source signals is not possible. The main reason is that the contrast function that distinguishes the state of separation from the state of mixing, e.g. kurtosis or other higher order statistics (HOS), decreases or even vanishes completely when signals are Gaussian [59]. Therefore, classical methods of blind source separation (BSS), such as independent component analysis [59], CMA [60] which use HOS tend not to work well for PCS-QAM.

In the literature on blind polarization demultiplexing algorithms, this phenomenon is often referred to as the kurtosis problem [60], [289] or described as formation of a sphere instead of a lens-like structure in the Stokes-space [64], [96]. Recently, new algorithms based on the idea of magnitude-based separation of signal constellations (symbols) were proposed [64], [69]. In [64], ‘inner QPSK’ symbols were extracted and transformed into their three-dimensional representation in the Stokes-space, to find an initialization for the multi modulus algorithm (MMA). An extension of this algorithm was presented recently in which symbols beyond the ‘inner QPSK’ were used [96]. In [290], a variational auto-encoder based equalizer for the equalization of a first-order PMD channel was proposed and shown to work for PCS-QAM signals. We note the similarity between the idea of separating constellations in [64], [69], [96] and what is referred to as quantized CMA in the literature [63]. There, the received symbols are separated into two or more categories based on their magnitude. The symbols in separated categories have much better contrast function or optimization landscape than the case where symbols are not categorized, i.e. all symbols are used. Therefore equalization is possible using classical algorithms like CMA. These algorithms extend methods that were originally designed for uniform QAM signals.

Here, we investigate an alternative idea. We exploit that if Gaussian signals exhibit diverse temporal correlations it is often possible to separate them using their second order statistics [59], [291]. An algorithm that is inherently suitable for Gaussian signals is expected to be suitable for PCS-QAM signals as well. Recently, we proposed a first method for blind polarization demultiplexing that exploits this idea [288]. The method is called joint diagonalization (JD) assisted MMA or JD-MMA, because it combines JD and a standard MMA (MMA) [292], [293]. The algorithm requires that temporal correlations are present in the transmitted signal, which were introduced by using additional filters at

the transmitter. A major limitation of the JD-MMA is that it may not work when channel has memory such as PMD channel. Furthermore, it uses a MMA instead of probability aware pr-MMA which has been shown to perform better [64].

In this paper, we present a new blind polarization demultiplexing algorithm based on second order statistics that can also be applied to channels with memory (i.e. with significant PMD). We furthermore improve our previous algorithm for the memoryless case [288] and investigate the performance of these methods in scenarios with different degrees of PMD. In addition, we discuss briefly the impact of pulse shaping, phase noise and frequency offset on the working of the proposed algorithms. Furthermore, the computational complexity and information loss due to the added temporal correlations are presented.

The paper is organized as follows: in Sec. 6.2, we briefly review the basics of PMD channel impairments and describe the first-order PMD model used in this paper. Then, we discuss in Sec. 6.3 our previously proposed JD-MMA together with its improved version JD-pr-MMA, designed for separating signals from instantaneous mixtures. The newly proposed FDJD-pr-MMA algorithm, designed for PMD channels, is presented in Sec. 6.4. Some additional processing steps required by FDJD-pr-MMA are detailed in Sec. 6.5. The simulation setup is explained in the Sec. 6.6, and numerical results are discussed in Sec. 6.7. The impact of pulse shaping, phase noise and frequency offset on the proposed algorithms, computational complexity and information loss due to filtering are briefly discussed in Sec. 6.8. Finally, the paper concludes in Sec. 6.9.

6.2. BRIEF REVIEW OF POLARIZATION MODE DISPERSION

An ideal optical fiber allows propagation of fundamental modes in two orthogonal polarizations without any interference. Real fiber, however, exhibits cylindrical asymmetry caused by factors such as manufacturing defects and mechanical stress, which leads to birefringence. That is, the two orthogonal polarization modes propagate with different group velocities. This phenomena is called polarization mode dispersion (PMD) [9], [294].

To understand PMD, let us consider a chunk of fiber with a constant birefringence along its length. In this chunk of fiber, there exist two axes known as ‘slow’ and ‘fast’ axes. For (linearly polarized) input light whose polarization aligns with either of these axes, the difference in the differential group velocities is maximum and the ‘fast’ and ‘slow’ light pulses thus reach the other end at different times. The difference of arrival times is called differential group delay (DGD). Although actual fiber does not have constant birefringence along its length, PMD can be modeled as a concatenation of several chunks with constant birefringence. We use such a first-order PMD model in our simulations [295], whose implementation is available online [296]. In Jones space, the polarization of light can be represented by a Jones vector. PMD effects that map input Jones vectors to output Jones vectors are given by a 2×2 unitary Jones matrix $\mathbf{U}(\omega)$, where ω is angular frequency. As explained, the PMD of the link can be described by a concatenation of Jones matrices, i.e. $\mathbf{U}(\omega) = \mathbf{U}_{n_c}(\omega)\mathbf{U}_{n_c-1}(\omega)\dots\mathbf{U}_2(\omega)\mathbf{U}_1(\omega)$. Here, $\mathbf{U}_i(\omega)$ is the PMD induced by i^{th} chunk and n_c is the number of chunks. The PMD of a single chunk

is given by

$$\mathbf{U}_i(\omega) = \mathbf{Q}_i(-\theta_i, -\phi_i) \begin{pmatrix} e^{j\omega\Delta\tau_i/2} & 0 \\ 0 & e^{-j\omega\Delta\tau_i/2} \end{pmatrix} \mathbf{Q}_i(\theta_i, \phi_i) \quad (6.1)$$

where $\theta_i \sim \text{Uni}[-\pi/2, \pi/2]$, $\phi_i \sim \text{Uni}[-\pi/4, \pi/4]$, and \mathbf{Q}_i is the rotation matrix of i^{th} chunk,

$$\mathbf{Q}_i(\omega) = \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i)e^{j\phi} \\ -\sin(\theta_i)e^{-j\phi} & \cos(\theta_i) \end{pmatrix}, \quad (6.2)$$

and $\Delta\tau_i$ is the DGD of an individual chunk. In the first-order model, the DGD τ_i is constant over the signal spectrum. The rotation of the state of polarization (RSOP) channel is a special case of (6.1) where the factor $\omega\Delta\tau_i$ is replaced with a phase shift that is constant over the signal bandwidth, i.e. $\omega\Delta\tau_i = \delta$. The RSOP channel can be implemented using a single chunk $n_c = 1$. Using a different notation, we can write the PMD channel propagation as

$$X_i(\omega) = \sum_{k=1}^2 H_{ik}(\omega) S_k(\omega), \quad i, k \in \{1, 2\}, \quad (6.3)$$

$$\text{or } \mathbf{X}(\omega) = \mathbf{H}(\omega)\mathbf{S}(\omega). \quad (6.4)$$

Here, $S_k(\omega)$ and $X_i(\omega)$ represent the discrete Fourier transform (DFT)¹ of the fiber input and output signals $s_k(t)$ and $x_i(t)$, respectively. In (6.4), bold capital letters represent matrices, i.e.

$$\mathbf{S}(\omega) = \begin{bmatrix} S_1(\omega) \\ S_2(\omega) \end{bmatrix}, \quad \mathbf{X}(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix}$$

and

$$\mathbf{H}(\omega) = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) \end{bmatrix}.$$

In the time-domain, (6.3) can be written as convolutive mixing

$$x_i(t) = \sum_{k=1}^2 \sum_{\tau=-\ell_{ch}}^{+\ell_{ch}} h_{ik}(\tau) s_k(t-\tau), \quad i, k \in \{1, 2\} \quad (6.5)$$

with ℓ_{ch} being the one-sided memory of the PMD channel. If $\ell_{ch} = 0$, (6.5) represents instantaneous mixing, i.e. the RSOP channel.

6.3. JD ASSISTED PR-MMA (JD-PR-MMA)

The JD-pr-MMA method for blind demultiplexing of signals assumes that the mixing channel is memoryless [288]. Hence, it is ideally suited for RSOP channel effects. The method is composed of two algorithms working in a sequence. A schematic of the method is shown in Fig. 6.1. First, an approximate joint diagonalization (JD) algorithm estimates an inverse of the instantaneous mixing matrix and applies it to the mixed signal [59, Ch. 7.4.1]. Then, a pr-MMA removes any residual effects to recover the QAM

¹We follow the convention for DFT as $x(\omega) = \sum_k x(k)e^{-j\frac{2\pi\omega k}{N}}$ and IDFT as $x(k) = \frac{1}{N} \sum_w x(\omega)e^{j\frac{2\pi\omega k}{N}}$.

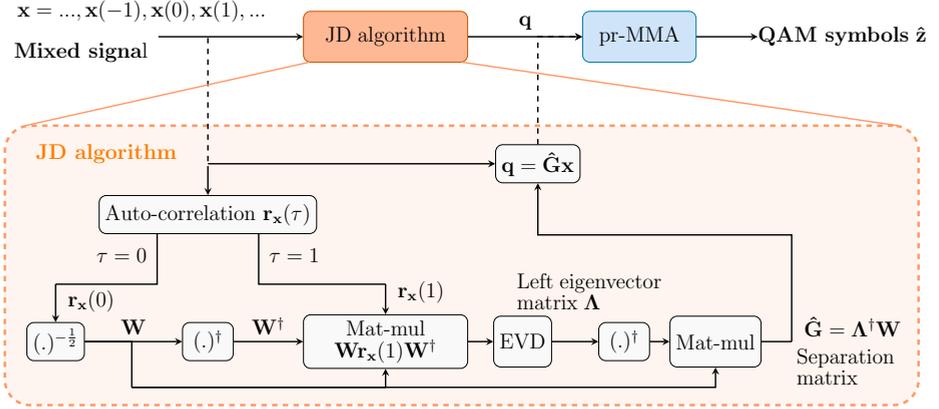


Figure 6.1: Schematic of JD-pr-MMA. The JD algorithm estimates an instantaneous separation matrix $\hat{\mathbf{G}}$, which is applied to the mixed signal. Then, the added temporal correlations and residual distortions are removed by pr-MMA in order to recover PCS-QAM symbols. In the ideal case, the intermediate signal $\mathbf{q} = \mathbf{s}$ for the RSOP channel.

6

symbols. The combined algorithm is referred to as JD assisted pr-MMA or JD-pr-MMA for short. We now explain this method in detail.

Let us denote a dual polarized transmit signal by $\mathbf{s} = [s_1, s_2]^T$, where $s_1(t)$ and $s_2(t)$ are base-band signals, and the mixed signal by $\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t)$. Here, \mathbf{H} is a instantaneous 2×2 unitary channel mixing matrix.

The method assumes that the transmit signals in the two orthogonal polarizations, i.e. s_1 and s_2 are

A1.1 zero-mean, second order ergodic, and wide sense stationary

A1.2 mutually uncorrelated, and

A1.3 have diverse temporal correlation, i.e. the temporal correlations of the sources are pairwise linearly independent.

We refer the reader to [59, Ch.7.3.1] for why the above assumptions are necessary. We meet the assumptions above by applying a filter $\mathbf{f} = [f_1, f_2]^T$ on the transmit iid PCS-QAM symbols on each polarization, $\mathbf{z} = [z_1, z_2]^T$, such that $s_i(t) = z_i(t) \star f_i(t)$ for $i \in 1, 2$. Here, \star denotes the convolution operation. It is sufficient to filter only one of the signals at the transmitter to satisfy assumptions A1.

With the above assumption in place, it can be shown that the auto-correlation of the signals \mathbf{s} , i.e. $\mathbf{r}_s(\tau) = \mathbb{E}\{\mathbf{s}(t)\mathbf{s}^\dagger(t + \tau)\} = \text{diag}(r_{s_1}(\tau), r_{s_2}(\tau))$, forms a sequence of diagonal matrices. Here, $r_{s_i}(\tau)$ is the correlation sequence of the i^{th} polarization and \dagger denotes the Hermitian transpose operation. It can also be shown that the auto-correlation of the mixed signal contains the instantaneous mixing (RSOP channel) matrix:

$$\mathbf{r}_x(\tau) = \mathbb{E}\{\mathbf{x}(t)\mathbf{x}^\dagger(t + \tau)\} = \mathbf{H}\mathbf{r}_s(\tau)\mathbf{H}^\dagger. \quad (6.6)$$

Note that $\mathbf{H}\mathbf{r}_x(0)\mathbf{H}^\dagger = \mathbf{r}_s(0)$ and $\mathbf{H}\mathbf{r}_x(1)\mathbf{H}^\dagger = \mathbf{r}_s(1)$ are both diagonal because of the assumptions A1. We can thus recover \mathbf{H}^\dagger up to rotations and ordering by finding a matrix that jointly diagonalizes $\mathbf{r}_x(0)$ and $\mathbf{r}_x(1)$.

A method for approximate joint diagonalization of correlation matrices $\mathbf{r}_x(\tau)$ for $\tau \in \{\tau_1, \tau_2\}$ is as follows [59, Ch. 7.4.1]:

1. Compute matrix $\mathbf{W} = \mathbf{r}_x(\tau_1)^{-1/2}/N$ by using a received block of length N ;
2. Compute an eigendecomposition $\Lambda \text{diag}(\lambda_1, \lambda_2) \Lambda^{-1} = \mathbf{W}\mathbf{r}_x(\tau_2)\mathbf{W}^\dagger$; and
3. Compute the inverse mixing or the separation matrix $\hat{\mathbf{G}} = \Lambda^\dagger \mathbf{W}$

We remark that we could in principle also jointly diagonalize $\mathbf{r}_x(\tau_1)$ and $\mathbf{r}_x(\tau_2)$ for other values than $\tau_1 = 0$ and $\tau_2 = 1$. Furthermore, note that the transmit filters ensure that $\mathbf{r}_x(1)$ is invertible. The inverse mixing matrix $\hat{\mathbf{G}}$ is an inverse of the RSOP channel mixing matrix up to the ambiguity of complex rotation and order. Note that these ambiguities are also present in other blind algorithms such as CMA.

By applying $\hat{\mathbf{G}}$ to the mixed signal, we recover an intermediate signal $\mathbf{q} = \hat{\mathbf{G}}\mathbf{x}$. Then, a MMA equalizer can recover the transmit PCS-QAM symbols $\hat{\mathbf{z}}$ by removing residual channel distortions as well as the transmit filters \mathbf{f} from \mathbf{q} . It was shown in [64] that a variation of the MMA equalizer that accounts for the probabilistic distribution of the transmit symbols performs better than the standard MMA [292], [293]. From here onwards, we refer to the standard MMA as MMA in the text. For readers who are not familiar with the probability aware-MMA, we explain it briefly below for completeness.

SOURCE PROBABILITY AWARE MMA (PR-MMA)

Similar to MMA, the pr-MMA uses a 2×2 multi-input multi-output (MIMO) structure for adaptive equalization. In each iteration, blocks of input signals are processed to produce the two equalizer output symbols. Then, each of the output symbols is assigned a radius based on the closest QAM symbol. An error is computed as the difference in the absolute squares of the equalizer output and the assigned radius. Finally, the equalizer taps are updated using the computed errors.

In a multi-level constellation, symbols can be categorized based on their magnitudes (i.e., QAM moduli or radii). The pr-MMA in [64] uses the fact that the probabilities of a symbol having a given modulus are not uniformly distributed and should be accounted for in the update equation. This can be achieved by multiplying the error with the probability of the corresponding assigned modulus. Note that these probabilities will vary with the entropy of the PCS formats. Furthermore, assigning a modulus closest to the magnitude of the equalizer output is not optimal in presence of noise [64]. Therefore, the thresholds used to assign the radius can also be optimized. However, that requires knowledge of the signal to noise ratio. In our pr-MMA implementation, only the errors are weighted (i.e. multiplied) by the probability of the corresponding assigned modulus. The equalizer outputs are assigned to the closest modulus of the constellation.

A shortcoming of the JD method is that it is only suitable for a memoryless channel since if \mathbf{H} is not memoryless then (6.6) does not hold true. As a result, for an RSOP channel the JD-pr-MMA works well, while in the presence of PMD effects the algorithm can

fail. In [288], we found that a similar algorithm, JD-MMA, could tolerate weak PMD effects (DGD values up to 10% of the symbol duration). In the next section, we propose a method that can be used to separate PCS-QAM symbols in the presence of stronger PMD impairments.

6.4. PROPOSED FDJD-PR-MMA FOR CONVOLUTIVE MIXING

The frequency-domain description (6.4) of the PMD effects can be interpreted as many parallel channels that all suffer from instantaneous mixing only. A method that uses this fact for blind-identification of a convolutive mixing channel was proposed in [297]. We refer to it here as a FDJD or frequency domain (FD) JD method. Although the original FDJD [297] is suitable only for a two-input-two-output channel that has only cross-mixing (i.e. $H_{11}(\omega) = H_{22}(\omega) = 1$), we can still apply this method to the PMD channel and combine it with other processing steps to propose a method that works for general PMD channels.

The underlying principle of the FDJD algorithm is described in [297]. Similar to JD, the FDJD algorithm makes several assumptions. The assumptions on the transmit signals are the same as for the JD-pr-MMA algorithm in Sec. 6.3, with the only difference being that both of transmit signals need to exhibit temporal correlations. Like JD-pr-MMA, the assumption on transmit signal \mathbf{s} is met in our setup by filtering the PCS-QAM symbols \mathbf{z} before transmission, i.e. $\mathbf{s} = \mathbf{f} \star \mathbf{x}$. Additionally, the FDJD algorithm requires assumptions about the channel:

- A2.1 The channel is a linear time-invariant filter with a finite impulse response (FIR).
- A2.2 The auto-channels are delta pulses, i.e. the diagonal entries satisfy $H_{ii}(\omega) = 1$ for $i \in \{1, 2\}$.
- A2.3 The cross-channels have no common zeros.
- A2.4 Each cross-channel has no zeros in conjugate reciprocal pairs or, equivalently, it does not contain zero-phase convolutional components.

The assumption A2.1 is not restrictive as any stable infinite impulse response (IIR) filter can be approximated by an FIR filter of appropriate length. The assumption A2.2 is not true for a PMD channel in general because the auto-channels are not impulses, i.e. $H_{ii}(\omega) \neq 1$ for $i \in 1, 2$. It was shown in [297] that the FDJD algorithm can still be applied when A2.2 is not satisfied. In that case, the estimated responses will be normalized cross-channels instead of true cross-channels. That is, instead of estimating the true channel responses $\hat{H}_{12}(\omega)$ and $\hat{H}_{21}(\omega)$, the FDJD algorithm estimates

$$\hat{H}'_{12}(\omega) = \frac{\hat{H}_{12}(\omega)}{\hat{H}_{22}(\omega)} \quad \text{and} \quad \hat{H}'_{21}(\omega) = \frac{\hat{H}_{21}(\omega)}{\hat{H}_{11}(\omega)}, \quad (6.7)$$

respectively.

We observed that the assumption A2.4 may not be true for the PMD channel in general. This assumption is required for solving the order ambiguity (explained later in Sec 6.5). Fortunately, it is not necessary when the temporal correlations of the transmit signals are known [297], as it is the case in our setup.

6.4.1. FDJD ALGORITHM

The FDJD algorithm takes a block of the mixed signals and uses the knowledge of the transmit signals' correlations (i.e. transmit filters \mathbf{f}) to provide the normalized cross channel response given in (6.7).

Similar to the JD algorithm in Sec. 6.3, this algorithm is based on the insight that the channel matrix $\mathbf{H}(\omega)$ jointly diagonalizes the matrices $\mathbf{C}_Y(\omega)$ and $\mathbf{C}_Y(\omega, \omega + 1)$, where $\mathbf{C}_Y(\omega_1, \omega_2) = \mathbf{R}_Y(\omega_1, \omega_2)\mathbf{R}_Y(\omega_1, \omega_2)^\dagger$, $\mathbf{R}_Y(\omega_1, \omega_2) = \mathbb{E}\{\mathbf{Y}(\omega_1)\mathbf{Y}(\omega_2)^\dagger\}$ [297, Eq. 21]. Here, $\mathbf{Y}(\omega) = \mathbf{V}(\omega)\mathbf{X}(\omega)$ denotes the whitened signal and $\mathbf{V}(\omega)$ the pre-whitening operator.

Before explaining the algorithm, let us define a $N_f \times N_f$ DFT matrix

$$\mathcal{F} \triangleq e^{-j\frac{2\pi}{N_f}\mathbf{u}\mathbf{u}^\dagger} \quad (6.8)$$

for $\mathbf{u} = [0, 1, \dots, N_f - 1]^T$ and a $N_f \times N_f$ Kaiser window matrix

$$\mathbf{K} = K^\dagger K, \quad (6.9)$$

with K being a Kaiser window of dimension $1 \times N_f$.

The algorithm is as follows (also see Fig. 6.2)

1. Compute the auto-correlation $\mathbf{r}_x(\tau)$ for $\tau \in [-N_f, N_f]$ of a block of the mixed signal $\mathbf{x} = [\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(N-1), \mathbf{x}(N)]$ using (6.6).
2. Compute two frequency domain (FD) correlation matrices $\mathbf{R}_X(\omega, \omega) = \mathbf{R}_X(\omega)$ and $\mathbf{R}_X(\omega, \omega + 1)$ with $\omega \in \{0, 1, 2, \dots, N_f - 1\}$, using

$$R_X^{ij}(\omega_1, \omega_2) = \mathcal{F}(\omega_1, :) \mathbf{K} \mathbf{r}^{x_i, x_j} \mathcal{F}(\omega_2, :)^{\dagger}, \quad (6.10)$$

where \mathbf{r}^{x_i, x_j} denotes the Toeplitz cross-correlation matrix of size $N_f \times N_f$ derived from $\mathbf{r}_x(\tau)$ and $\mathcal{F}(\omega_1, :)$ denotes the ω_1^{th} row of the matrix \mathcal{F} .

3. Compute the pre-whitening operator

$$\mathbf{V}(\omega) = \mathbf{R}_X^{-1/2}(\omega). \quad (6.11)$$

4. Next, compute the FD correlation matrix of the whitened signal $\mathbf{y}(t)$ given by $\mathbf{Y}(\omega) = \mathbf{V}(\omega)\mathbf{X}(\omega)$ as

$$\mathbf{R}_Y(\omega, \omega + 1) = \mathbf{V}(\omega) \mathbf{R}_X(\omega, \omega + 1) \mathbf{V}^\dagger(\omega + 1). \quad (6.12)$$

5. Compute the matrix

$$\mathbf{C}_Y(\omega, \omega + 1) \triangleq \mathbf{R}_Y(\omega, \omega + 1)\mathbf{R}_Y^\dagger(\omega, \omega + 1). \quad (6.13)$$

6. The eigenvalue decomposition $\mathbf{C}_Y(\omega, \omega + 1)$

$$\mathbf{C}_Y(\omega, \omega + 1) = \mathbf{W}(\omega)\hat{\mathbf{D}}_2(\omega, \omega + 1)\mathbf{W}^\dagger(\omega + 1) \quad (6.14)$$

provides the column eigenvector matrix $\mathbf{W}(\omega)$.

7. The normalized MIMO channel response is given by

$$\hat{\mathbf{H}}'(\omega) = \mathbf{F}(\omega) \text{diag}(\mathbf{F}(\omega))^{-1}, \quad (6.15)$$

where

$$\mathbf{F}(\omega) \triangleq \mathbf{V}(\omega)^{-1} \mathbf{W}(\omega). \quad (6.16)$$

By applying the inverse of the normalized MIMO filters $\hat{\mathbf{H}}'(\omega)$ to the mixed signal $\mathbf{X}(\omega)$, the cross mixing components will be removed. However, a filter within each trans-

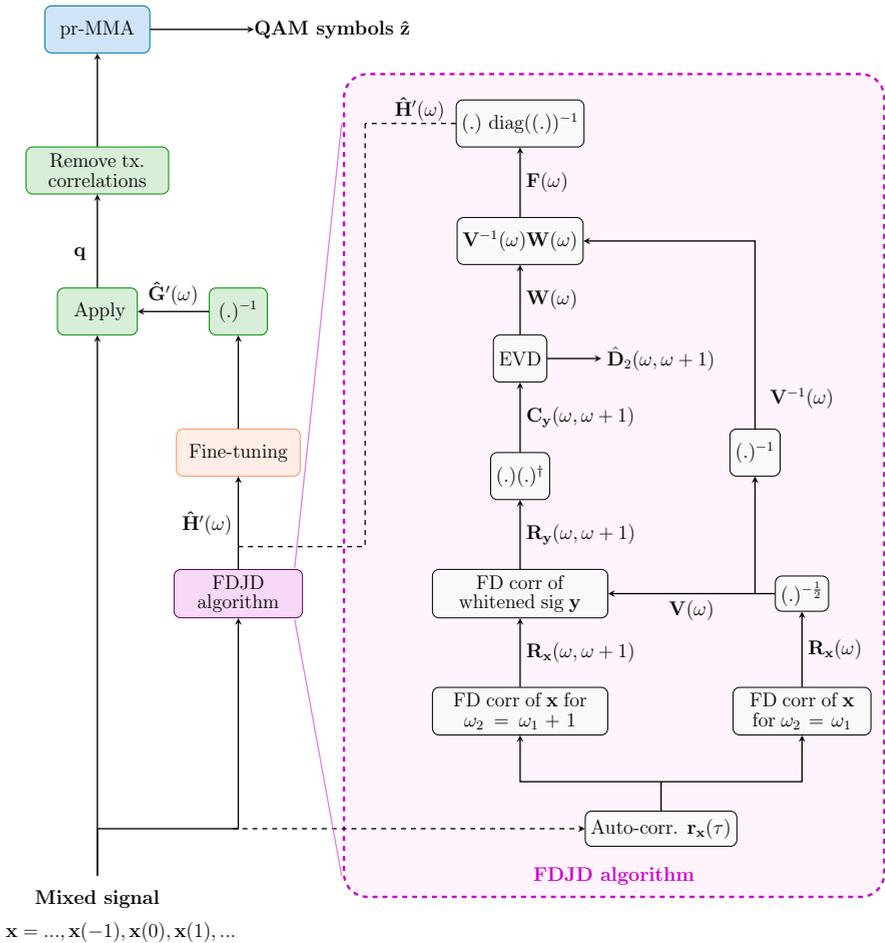


Figure 6.2: Schematic of FDJD-pr-MMA. The FDJD algorithm blindly estimates the PMD channel's normalized response $\hat{\mathbf{H}}'(\omega)$. After ambiguity removal and reconstruction, an inverse of this response $\hat{\mathbf{G}}'(\omega)$ is applied, which removes the cross mixing components of the PMD channel. Finally, the temporal correlations are removed and PCS-QAM symbols are recovered using pr-MMA.

mit signals remains (see [297, Sec. IV]), i.e.

$$\mathbf{Q}(\omega) = \hat{\mathbf{H}}'(\omega)^{-1} \mathbf{X}(\omega) = \begin{bmatrix} \hat{H}_{11}(\omega) s_1(\omega) \\ \hat{H}_{22}(\omega) s_2(\omega) \end{bmatrix}. \quad (6.17)$$

In order to recover the PCS-QAM symbols $\hat{\mathbf{z}}$, the filters $\hat{H}_{11}(\omega)$ and $\hat{H}_{22}(\omega)$ and the transmit filters $\mathbf{f}(t)$ need to be removed. As we assume that the transmit filters are known, their inverse can be applied to the signal right after removing the PMD channel's cross-component using (6.17). Finally, a MMA equalizer can be used to remove filters remaining from the FDJD algorithm as well as any distortions.

The FDJD-pr-MMA requires three additional processing blocks to resolve the order ambiguity and to improve FDJD estimates and pr-MMA convergence. We explain them in the next section.

6.5. ADDITIONAL PROCESSING BLOCKS

The three blocks of processing pertain to ambiguity removal, reconstruction of the normalized PMD channel response, and candidate phase response search (CPRS). The first block is needed since the estimated normalized response $\hat{\mathbf{H}}'(\omega)$ from the FDJD algorithm is not necessarily equal to the true normalized channel, and can contain ambiguities similar to polarization switching in CMA. This block is described in subsection 6.5.1. After ambiguity removal, the next block performs reconstruction of the normalized PMD channel as described in subsection 6.5.2. This block discards frequency bins that contain relatively large estimation errors and reconstructs the FDJD estimated response. These two blocks are applied post-FDJD algorithm, and the last block, CPRS, is applied right before pr-MMA. The CPRS block was added because, during evaluation, we found that in some cases pr-MMA would fail in equalizing for the PMD auto-channels, i.e. $H_{11}(\omega)$ and $H_{22}(\omega)$ as given in (6.17), despite FDJD estimates being quite good. The CPRS block improved convergence of the pr-MMA, which will be elaborated on in the Results section. The working of the CPRS block is provided in subsection 6.5.3.

6.5.1. AMBIGUITY REMOVAL

The ambiguity in FDJD estimated responses tends to be more severe than that present in algorithms such as CMA or JD-(pr-)MMA. As the FDJD algorithm is applied in the frequency domain, the order ambiguity is present in every frequency component. Because the PMD channel is a two-input-two-output system, at any given ω the estimated normalized response will be either [297]

$$\hat{\mathbf{H}}'(\omega) = \begin{bmatrix} 1 & \hat{H}'_{12}(\omega) \\ \hat{H}'_{21}(\omega) & 1 \end{bmatrix} \quad (6.18)$$

or

$$\hat{\mathbf{H}}'(\omega) = \begin{bmatrix} 1 & 1/\hat{H}'_{21}(\omega) \\ 1/\hat{H}'_{12}(\omega) & 1 \end{bmatrix}. \quad (6.19)$$

This frequency component switching problem is visualized in Fig. 6.3, where we plot 128 point DFTs of normalized PMD channel response. In Fig. 6.3(a)-(b), the curve in black

shows the underlying true response $H'_{12}(\omega)$. The blue dashed line as well as the green circles show the FDJD estimated response $\hat{H}'_{12}(\omega)$. It is clear that at some frequency bins the estimated response is switched/flipped, which is just a change of sign on the log-scale of vertical axis in Fig. 6.3(a).

This ambiguity on the frequency component can be resolved by using assumption A2.4 or by the knowledge on the transmit signals correlations $\mathbf{r}_s(\tau)$ [297]. We utilize the latter. We know that PCS-QAM symbols \mathbf{z} are iid. By assuming \mathbf{z} is normalized, it can be shown that the transmit signal correlations $\mathbf{r}_s(\tau) = \mathbf{r}_f(\tau)$, are known at the receiver since the transmit filters \mathbf{f} are known. From that, the FD correlations of \mathbf{f} , i.e. $\mathbf{R}_f(\omega)$ and $\mathbf{R}_f(\omega, \omega + 1)$, can be obtained. Finally, the matrix $\mathbf{D}_2(\omega, \omega + 1)$ can be computed using $\mathbf{R}_y = \mathbf{R}_f$ in (6.13) and (6.14). The diagonal entries of the matrix $\hat{\mathbf{D}}_2(\omega, \omega + 1)$ are the real and positive eigenvalues of $\mathbf{C}_f(\omega, \omega + 1)$. By comparing the true order of eigenvalues in \mathbf{D}_2 against the estimated order in $\hat{\mathbf{D}}_2$, we can rearrange the estimated $\mathbf{W}(\omega)$ and thus determine if $\hat{\mathbf{H}}'(\omega)$ at the given ω is from relation (6.18) or (6.19).

We observed that it was not straightforward to resolve the frequency components' switching in the presence of estimation errors. Therefore, we select a set of reliable frequencies ω_r and correct the frequency components' switching of them. The selection of the reliable frequencies ω_r is based on a quantity called a normalized ratio of singular values (NRSV) [297], defined as

$$\text{NRSV}_{\text{tx}}(\omega) = \frac{|\mathbf{D}_2^{11}(\omega) - \mathbf{D}_2^{22}(\omega)|}{\max(\mathbf{D}_2^{11}(\omega), \mathbf{D}_2^{22}(\omega))}, \quad (6.20)$$

$$\text{NRSV}_{\text{rx}}(\omega) = \frac{|\hat{\mathbf{D}}_2^{11}(\omega) - \hat{\mathbf{D}}_2^{22}(\omega)|}{\max(\hat{\mathbf{D}}_2^{11}(\omega), \hat{\mathbf{D}}_2^{22}(\omega))}. \quad (6.21)$$

For a given threshold NRSV_{th} , frequencies for which $\text{NRSV}_{\text{rx}}(\omega) > \text{NRSV}_{\text{th}}$ are added to the set of reliable frequencies. Choosing a large threshold NRSV_{th} seems to require the transmit filters $\mathbf{f}(t)$ to be more frequency selective, while choosing a small threshold requires an accurate estimation of the statistics of the mixed signal. In our simulations, we chose not to make the transmit filter more frequency selective and used a small value of 0.02 for NRSV_{th} . In addition to NRSV_{th} , more restrictions specific to the PMD channel are imposed for selecting reliable frequencies. From the PMD model described in Sec. 6.2, it can be inferred that $|H_{11}(\omega)| = |H_{22}(\omega)|$ and $|H_{12}(\omega)| = |H_{21}(\omega)|$ and thus, $|\hat{H}'_{21}(\omega)| = |\hat{H}'_{12}(\omega)|$. Their ratio $|I_1(\omega)|$, defined as

$$\hat{I}_1(\omega) = \frac{\hat{H}'_{21}(\omega)}{\hat{H}'_{12}(\omega)}, \quad (6.22)$$

should be 1 at all ω . The frequencies for which $|I_1(\omega) - 1|$ is larger than a set threshold $I_{1,\text{tol}}$ (chosen as 0.2 in our case) are also discarded. The $\text{NRSV}_{\text{tx}}(\omega)$ and $\text{NRSV}_{\text{rx}}(\omega)$ are shown in Fig. 6.3(c). There values at reliable frequencies ω_r are indicated by markers. At frequency bins where $\text{NRSV}_{\text{tx}}(\omega_r)$ and $\text{NRSV}_{\text{rx}}(\omega_r)$ are not in agreement (i.e. positive/negative), the frequency components are switched. In other words, at those ω values, FDJD estimated the solution from (6.19), which can be corrected to (6.18) by a reciprocal operation followed by a swap operation on $\hat{H}'_{12}(\omega)$ and $\hat{H}'_{21}(\omega)$.

An illustration of the corrected spectrum is shown in Fig. 6.3(d)-(e). The frequency switching problem shown in Fig. 6.3(a)-(b) is corrected at the reliable ω_r bins.

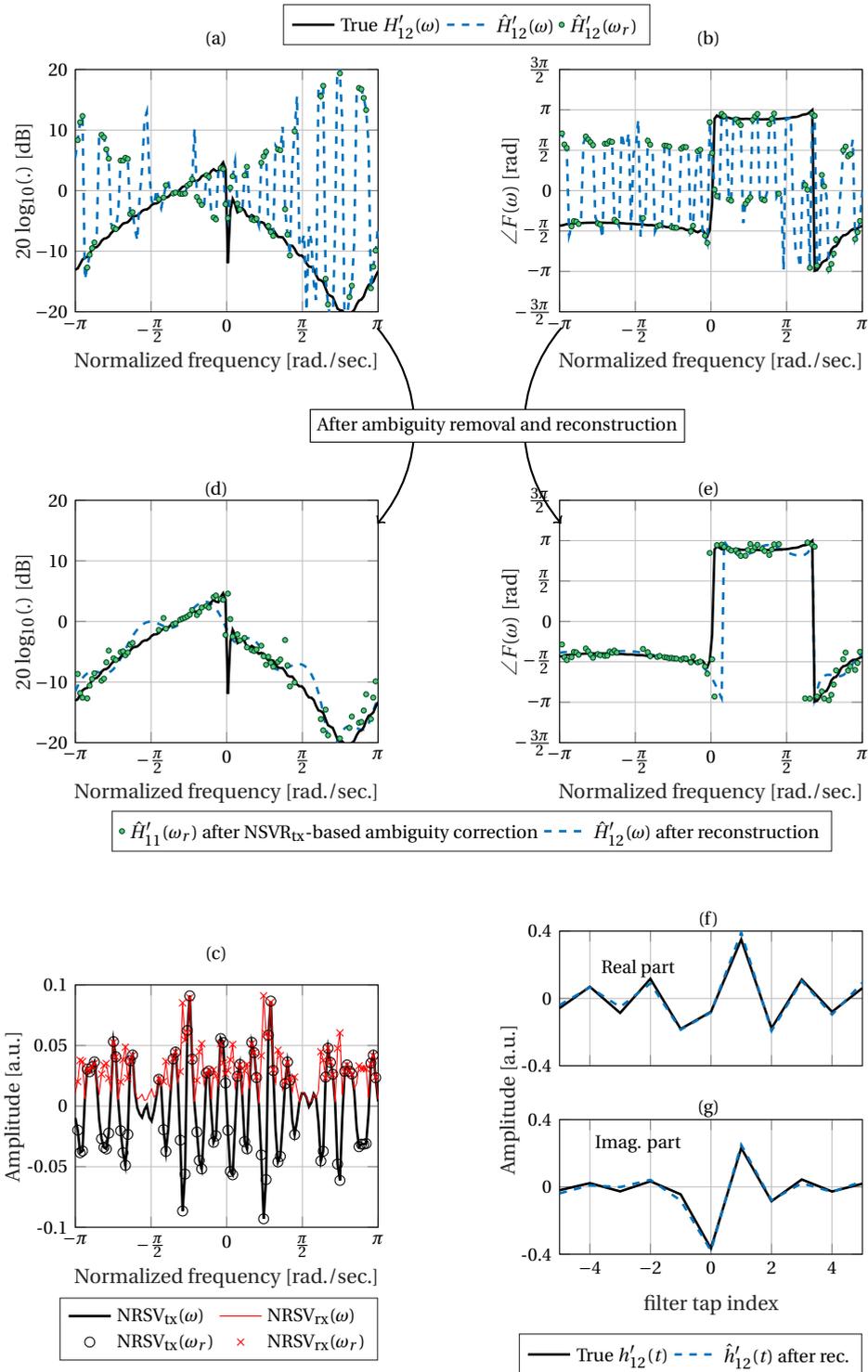


Figure 6.3: An example of ambiguity removal and reconstruction of the normalized PMD channel's response estimates obtained through the FDJD algorithm: (a-b) raw estimates, (c) NRSV based selection and decision on reliable frequencies ω_r , (d-e) reconstructed channel responses after removing ambiguity, and (f) true and estimated impulse responses.

6.5.2. RECONSTRUCTION OF THE NORMALIZED PMD CHANNEL RESPONSE

After removing ambiguity at reliable frequencies, the channel response is reconstructed using the following iterative procedure.

Let us assume that the memory-length of the normalized cross-channel impulse responses $\hat{h}'_{12}(n)$ and $\hat{h}'_{21}(n)$ is roughly known and referred to as estimator memory. We carry out multiple iterations of the following four steps: First, an IDFT operation is applied on $\hat{H}'_{12}(\omega)$ and $\hat{H}'_{21}(\omega)$. Then, memory of the obtained time domain impulse responses $\hat{h}'_{12}(n)$ and $\hat{h}'_{21}(n)$ is limited to the estimator memory length, i.e. taps beyond the estimator memory are forced to zero. In the third step, a DFT operation is applied to get frequency responses. Finally, the magnitude responses are constrained to be equal at ω_r , i.e. $|\hat{H}'_{21}(\omega_r)| = |\hat{H}'_{12}(\omega_r)|$. After a few iterations, an improved estimate of the normalized channel responses $\hat{H}'_{12}(\omega)$ and $\hat{H}'_{21}(\omega)$ is obtained. An example of the reconstructed spectrum is shown in Fig. 6.3(d)-(e) along with their corresponding impulse response in Fig. 6.3(f)-(g). The MIMO inverse of the normalized channel response is finally given by $\hat{\mathbf{G}}(\omega) = \hat{\mathbf{H}}'^{-1}(\omega)$.

6.5.3. CANDIDATE PHASE RESPONSE SEARCH

Ideally, the task of the pr-MMA in the FDJD-pr-MMA is reduced to finding the phase response of $G_{11}(\omega) = H_{11}^{-1}(\omega)$ and $G_{22}(\omega) = H_{22}^{-1}(\omega)$ because the magnitude responses $|G_{11}(\omega)| = |H_{22}^{-1}(\omega)|$ can be calculated from the normalized cross channel responses $\hat{H}_{21}(\omega)$ and $\hat{H}_{12}(\omega)$ estimated by the FDJD algorithm. The PMD channel matrix $\mathbf{H}(\omega)$ is a 2×2 unitary matrix for every ω . Then, algebraic manipulation shows that $|G_{11}(\omega)| = |H_{11}^{-1}(\omega)| = \sqrt{1 - \hat{H}_{12}(\omega)\hat{H}_{21}(\omega)}$.

As pr-MMA failed often for larger DGD, we added an additional processing block before the pr-MMA, called the candidate phase response search (CPPS). CPRS takes a phase response $\angle G_{11}(\omega)$ from 27 pre-chosen phase responses, combines it with $|G_{11}(\omega)|$ to make a candidate $G_{11}(\omega)$ and applies this to the signal. Then, the cost function of the pr-MMA cost is measured. The candidate that achieves the minimum cost function is applied on the signal prior to pr-MMA. This filter was applied in the time domain, on the first polarization, by doing an IDFT operation and limiting memory to (estimator memory)/2. Similarly, another filter was derived and applied to the signal in the second polarization. We observed that after applying these inverse filters to the signal, the PCS-QAM constellation starts appearing. Feeding the filtered signal to the pr-MMA recovers the PCS-QAM symbols in most cases where pr-MMA failed previously.

6.6. SIMULATION SETUP

The block diagram of the simulation setup is shown in the Fig. 6.4. At the transmitter, a PCS-64QAM mapper is used to generate two 32 GBaud sequences z_1 and z_2 of length N_{symp} according the Maxwell-Boltzmann distribution for the given entropy \mathcal{H} . Next, the sequences of symbols are passed through transmit filters that add time correlations to the symbols. The filtered signals s_1 and s_2 are multiplexed and transmitted through the channel. The channel was simulated as a cascade of the PMD/RSOP model (explained in Sec. 6.2), followed by addition of noise from an additive white Gaussian noise (AWGN) source to the mixed signal. The DGD parameter is provided as a percentage of the sym-

bol duration. At the receiver, candidate DSP algorithms are applied to recover the transmit PCS-64QAM symbols. Then, the performance is computed in terms of 2-D mutual information (MI) bits.

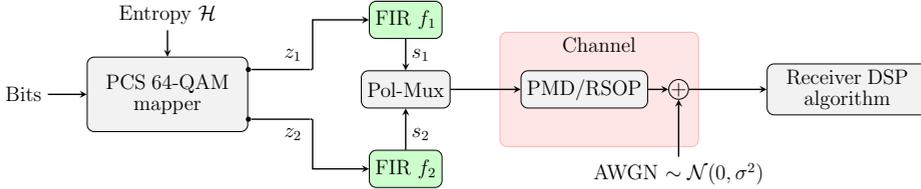


Figure 6.4: The simulation setup. Receiver DSP algorithm candidates are JD-(pr-)MMA, the proposed FDJD-pr-MMA, and CMA-(pr-)MMA to benchmark against. The FIR filters f_1 and f_2 are only used for JD-(pr-)MMA and FDJD-pr-MMA. In other words, for CMA-(pr-)MMA, $f_1(t)$ and $f_2(t)$ are Dirac impulses $\delta(t)$.

The receiver DSP algorithm candidates are our previously proposed JD-MMA, the newly proposed JD-pr-MMA and FDJD-pr-MMA, and CMA-(pr-)MMA, which serves as a reference for our benchmark. The transmit filters are introduced to satisfy the assumptions involved in JD-(pr-)MMA and FDJD-pr-MMA algorithms. For CMA-(pr-)MMA, these filters are not needed and therefore omitted from the signal path or rather treated as Dirac impulses $f_1(t) = f_2(t) = \delta(t)$.

It should be noted that simulation setup as well as algorithms under investigation are implemented at 1 sample per symbol (sps).

6.7. RESULTS

We carried out a comparative study on the performance of the reference CMA-(pr-)MMA, the improved JD-(pr-)MMA for memoryless channels described in Sec. 6.3, and the new FDJD-MMA from Sec. 6.4. We first present results for the JD-(pr-)MMA, and then show the performance of the new FDJD-pr-MMA.

6.7.1. JD-PR-MMA (RSOP CHANNEL AND LOW PMD)

Here, we compare the gains obtained by replacing MMA by pr-MMA. We use the same configuration that was used in our previous paper [288]. That is, we chose $N_{\text{symb}} = 2^{16}$ and $f_1(t) = [1, 0.05 + 0.05j]$, $f_2(t) = \delta(t)$ for the JD-(pr-)MMA.

The number of taps for (pr-)MMA is fixed to 11. The learning rate of CMA and MMA was 5×10^{-3} and 5×10^{-4} , respectively. The learning rate of pr-MMA was scaled up according to the probability distribution of the transmit signal constellation such that on average the step size of pr-MMA matches that of MMA. The length of the training sequence is fixed at 2^{14} , and 101 different random realizations of the channel were studied. In the figures, the average of the MI values calculated for each realization and each polarization is reported, unless stated otherwise.

First, we compare the CMA-(pr-)MMA and JD-(pr-)MMA on the RSOP channel. In Fig. 6.5(a), the average MI for CMA-(pr-)MMA is plotted against the SNR per polarization. For reference, the AWGN channel spectral efficiency (SE) is plotted as well. We observe that CMA-MMA performs close to the AWGN bound for SNRs up to 18 and then

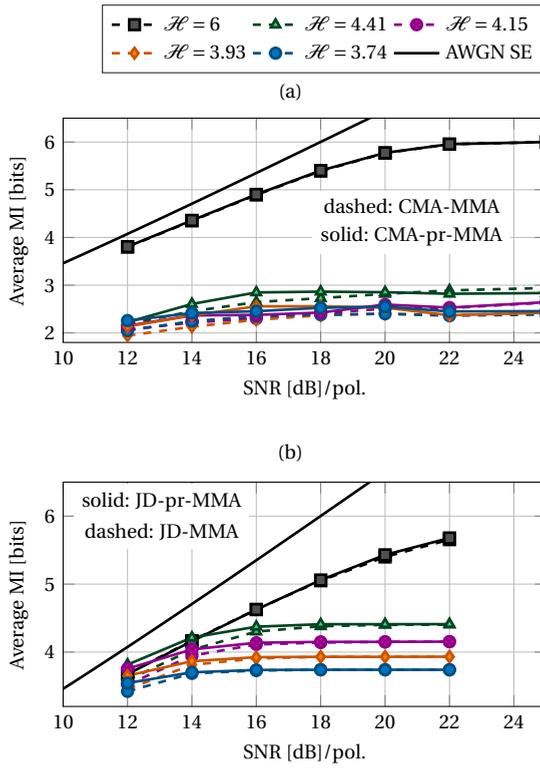


Figure 6.5: Average MI of (a) CMA-(pr-)MMA and (b) JD-(pr-)MMA for the RSOP channel. \mathcal{H} denotes the entropy of the PCS-64QAM format. Average over 101 realizations and two polarizations.

saturates at 6 bits for entropy $\mathcal{H} = 6$, i.e. uniform 64-QAM signal, which is expected. Next, we look at the performance for shaped QAMs of various entropy ranges. We observe that the average MI performance for entropies $\mathcal{H} \neq 6$ is worse compared to $\mathcal{H} = 6$. For example, at higher SNR > 20 dB, the average MIs are not reaching the corresponding entropy value, but instead stay below 3 bits. By looking into the MI for each realization, we found that for only around half of the realizations of shaped 64-QAM reach closer to their entropy value at SNR 20 dB. The rest of the realizations produced MI < 3 bits with some producing MI of only around 1 bit. These realizations were not successfully equalized by CMA-(pr-)MMA.

Now, we look at the performance of JD-MMA, which is shown in Fig. 6.5(b). An average MI > 3 bits is observed for all the entropies in the considered SNR range. We also see that average MI values reach their corresponding entropy values at higher SNRs (> 18 dB), indicating successful equalizations in all the cases. A noticeable difference in the average MI for uniform 64-QAM is observed (for e.g., at SNR = 20 dB) when compared with CMA-MMA. This difference can be reduced by running more training iterations of MMA. We also observe that using JD-pr-MMA instead of JD-MMA increases the average

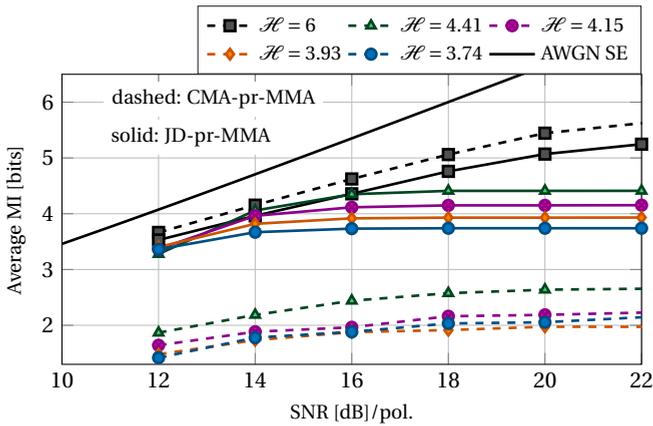


Figure 6.6: Average MI of CMA-pr-MMA and JD-pr-MMA for DGD = 10% of symbol duration T_{symb} , averaged over 101 realizations of two polarizations.

MI in the smaller SNR range (< 15 dB). These results show that using pr-MMA helped in improving convergence of both CMA-MMA and JD-MMA. Therefore, we use the pr-MMA instead of MMA for the results in the following part of the paper.

The RSOP channel is an instantaneous mixing case for which JD-(pr-)MMA should always work in the absence of noise. Next, we test these algorithms for a PMD channel. We set a DGD value of $0.1 T_{\text{symb}}$ and evaluate the performance. The results are shown in Fig. 6.6. A similar trend is observed in average MI performance of CMA-pr-MMA and JD-pr-MMA. Though when compared with the RSOP channel, both CMA-pr-MMA and JD-pr-MMA achieve slightly lower MI because the channel memory is increased with DGD.

When we further increase the DGD parameter of the PMD channel, we found that a memory of 11 taps is no longer sufficient. Therefore, we increased the MMA memory to

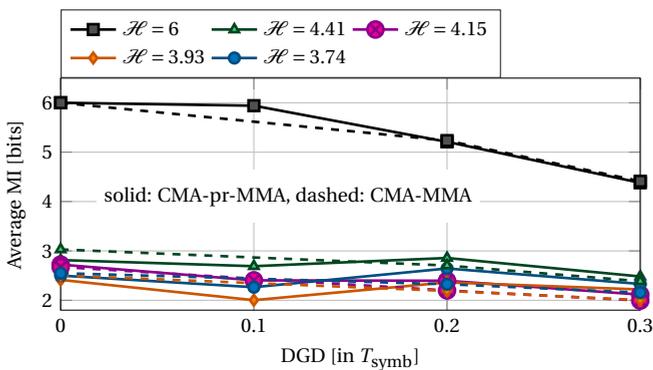


Figure 6.7: Average MI of CMA-(pr-)MMA for various DGD values for SNR = 30 dB.

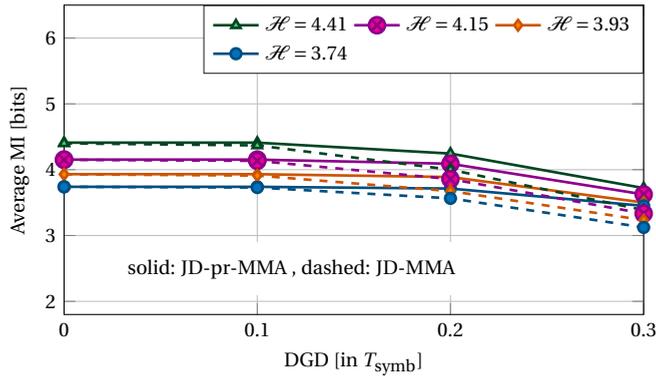


Figure 6.8: Average MI of JD-(pr-)MMA for various DGD values for SNR = 22 dB.

31 taps. The average MI for CMA-(pr-)MMA for a fixed SNR of 30 dB and varying DGD values are shown in Fig. 6.7. We see that CMA-MMA only achieves average MI > 3 bits for the case of uniform QAM signals. For shaped QAMs, average MI is < 3 bits. On observing MI for each PMD channel realization, we found more cases of unsuccessful equalization relative to the RSOP channel.

The average MI of JD-(pr-)MMA for a fixed SNR of 22 dB and varying DGD values is shown in Fig. 6.8. We observe that the average MI obtained with JD-pr-MMA is reducing with increasing DGD values but it is more than 3 bits in all cases. We found that only around 2% of channel realizations gave MI < 3 bits at DGD = 0.2 T_{symb} and $\mathcal{H} = 4.15$ bits. On the other hand, this percentage was around 50% for CMA-MMA even for DGD $\leq 0.2 T_{\text{symb}}$. We note that at higher DGD, JD-pr-MMA performance is not acceptable. But, for low DGD of = 0.1 T_{symb} , JD-pr-MMA obtained MI > 3 bits for all the channel realizations and entropies. It shows that JD-pr-MMA is quite effective in compensating polarization impairments in the weak PMD distortions regime. For the strong PMD channel, we propose FDJD-pr-MMA, which is investigated next.

6.7.2. FDJD-PR-MMA (MEDIUM PMD)

For evaluation of the FDJD-pr-MMA, the first task was to design transmit filters f_1 and f_2 . A transmit filter is considered well-fit if it produces reliable frequencies (explained in Sec 6.5), well-distributed over the spectrum, while not being too frequency selective. We found that if both filters are conjugate of each other, better channel estimations are obtained. We chose a 21 taps long transmit filter f_1 and accordingly $f_2 = f_1^\dagger$. The inverse of filter \mathbf{f} was truncated to 51 taps. We remark that the chosen filters and memory length parameters are hand-tuned.

The frequency responses of the transmit filters are plotted in Fig. 6.9. It is visible that the magnitude responses have many fluctuations but most of them are within ± 2 dB and nearly all of them are within -4.5 dB and +3 dB. Furthermore, the phase responses are also well within $\pm \frac{\pi}{4}$. We also found that an accurate estimation of the normalized cross-channel impulse responses is required by the FDJD algorithm or the subsequent steps may fail. A straightforward way is to increase the data length for the auto-correlation

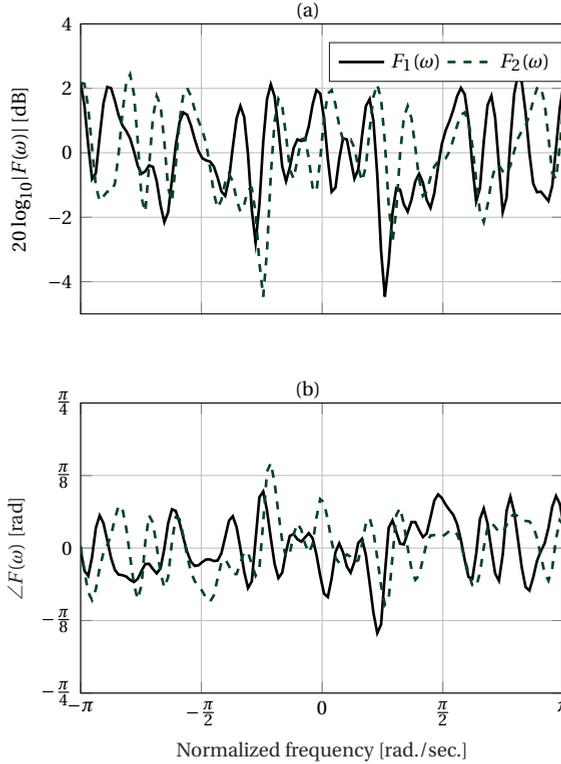


Figure 6.9: (a) Magnitude and (b) phase responses of the transmit filters \mathbf{f} used for FDJD-pr-MMA results.

measurement. Therefore, we use a data length of 2^{18} symbols for FDJD, while the following MMA uses only 2^{14} symbols. During the evaluation of the FDJD-pr-MMA, we fix the FDJD estimator memory to 11 taps, $\text{NSVR}_{\text{th}} = 0.02$, $I_{1,\text{tol.}} = 0.2$, $N_f = 128$, and the Kaiser window parameter $\beta = 24$.

The average MI of FDJD-pr-MMA are shown in Fig. 6.10 for 30 dB SNR and varying DGD values. The dashed curves show the average MI for the FDJD-pr-MMA. We see that for DGD values $< 0.2T_{\text{symb}}$, an average MI close to the corresponding entropy values is obtained. The average MI decreases with increasing DGD values. This is because for more number of realizations, FDJD-pr-MMA failed to equalize. On investigating why the FDJD-pr-MMA algorithm failed for some random realizations, we found that despite the FDJD estimation of modified channel responses being quite good in those cases, the subsequent pr-MMA did not perform well. Therefore, we add the CPRS block, whose working is explained in 6.5.3, right before pr-MMA.

In Fig. 6.10, we also plot the performance of FDJD-pr-MMA after including CPRS block. We observe that average MI is quite steady for DGD values up to 40% of symbol duration and a comparatively much slower decrease in MI is observed when the CPRS block is used. Furthermore, for DGD values $\leq 0.2T_{\text{symb}}$, none of the channel realization gave MI < 3 bits for $\mathcal{H} \neq 6$ bits. We also plot the average MI obtained by the FDJD-pr-

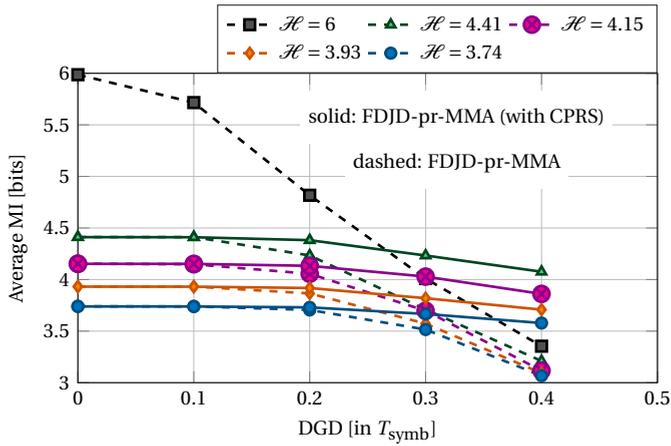


Figure 6.10: Average MI of FDJD-pr-MMA for various DGD values and SNR = 30 dB. Adding a CPRS stage before pr-MMA helps in equalizing PCS-QAMs.

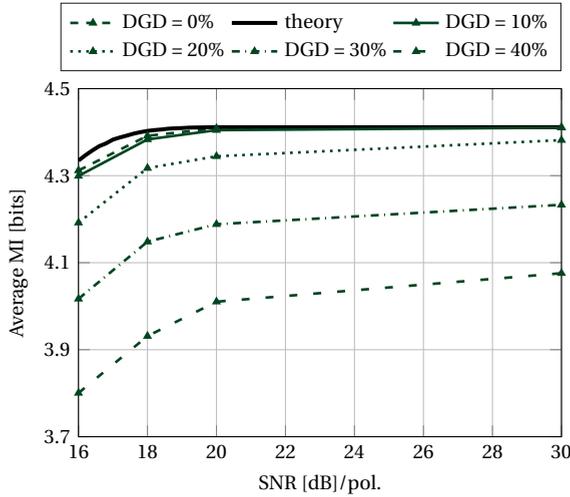


Figure 6.11: FDJD-pr-MMA (with CPRS) performance: Average MI against SNR over various DGD values in % of symbol duration.

MMA (with CPRS) in Fig. 6.11 for $\mathcal{H} = 4.41$ bits. A curve with MI obtained with an ideal demultiplexing is also plotted for reference. The average MI obtained for $\text{DGD} \leq 10\%$ of T_{symb} is quite close to the theoretically expected value. We also observed that for all of the tested entropies, not plotted here for brevity, the performance is quite close to theory for low DGD ($\leq 0.1T_{\text{symb}}$). An increase in the DGD causes decrease in the average MI. At a DGD of $0.4T_{\text{symb}}$, a few (~ 7) out of 101 random realizations achieved $\text{MI} < 3$ bits as shown in Fig. 6.12. We observe that MI values around 4 bits are obtained in most of the realizations; 3 out of 101 realizations gave MI less than 1 bit showing those realizations

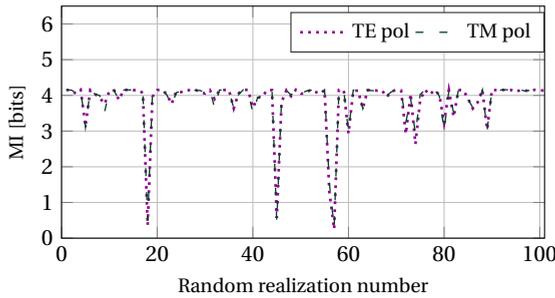


Figure 6.12: FDJD-pr-MMA (with CPRS) performance: MI obtained for 101 random realizations (DGD = 0.4 T_{sym} , SNR = 30 dB, $\mathcal{H} = 4.15$ bits).

failed to equalize completely.

6.8. DISCUSSION

In this study, we proposed two algorithms and demonstrated them on a simplified setup. The algorithms were evaluated at 1 sample per symbol and impairments such as phase noise and frequency offset were not included. In contrast, in a typical coherent receiver DSP, polarization demultiplexing algorithms usually operate at 2 sample per symbol in the presence of phase noise and frequency offset. Furthermore, the algorithms proposed in this work require filters at the transmitter, whose impact on the information rate is not yet clear. In this section, we discuss on these aspects together with the computational complexity of the algorithms.

6.8.1. PULSE SHAPING

The proposed algorithms (and setup) require a few modifications for oversampling scenarios. There are two possibilities for adding temporal correlations at the transmitter: before or after the pulse shaping stage. The former approach seems preferable because it is easier to implement as the filter is applied to an iid sequence. The added correlations can be extracted at the receiver by feeding a down-sampled sequence (at 1 sps) to the JD and FDJD algorithm.

The subsequent stages of FDJD algorithm such as applying MIMO inverse $\hat{\mathbf{G}}(\omega)$ and inverse of the transmit filters and candidate phase response search (CPRS) need to account for the change in the sample rate. The inverse of modified channel response $\mathbf{G}^{-1}(\omega)$ and filters that remove temporal correlations \mathbf{f}^{-1} are designed for the critically sampled sequence. Therefore, these filters need to be upsampled before applying them to the original oversampled sequence. After removing temporal correlations, a down-sampled sequence can be fed to the CPRS stage. Then, the candidate response filter can be upsampled and applied to the original oversampled sequence. The MMA equalizer can still be implemented with more than 1 sps and perform matched filtering.

6.8.2. EFFECT OF PHASE NOISE AND FREQUENCY OFFSET

The proposed algorithms require the computation of cross-correlations of the mixed signal. When phase noise and frequency offset are present, the correlation measurement can be erroneous. For lasers of 100kHz linewidth, which is often used in current coherent transceivers, We find that phase noise varies slowly over the considered correlation time-lags. Therefore, its impact is expected to be negligible on estimation. This was verified by quick simulations not presented here. The effect of a frequency offsets on the JD algorithm is also observed to be minimal. On the other hand, the effect of a frequency offset as small as 1 GHz is detrimental to FDJD algorithm. Therefore, a frequency offset correction is necessary for the FDJD algorithm.

6.8.3. EFFECT OF TRANSMIT FILTERS

Due to the presence of inverse filters f^{-1} at the receiver, the white Gaussian noise present in the received signal becomes colored. To quantify the impact of the transmit filter, we considered a simpler setup without any polarization impairments. In such a case, the end-to-end communication channel is a cascade of the transmit filters, an additive white Gaussian noise (AWGN) channel and the inverse transmit filters. This channel is equivalent to a additive colored Gaussian noise (ACGN) channel, where the colored noise arises from the applying the inverse filter at the receiver to the AWGN. In order to quantify the change in MI compared to AWGN channel, we run single polarization transmission simulations for two cases: one with and other without a transmit filter. The transmit filter used in this test is same as the one used in horizontal polarization of FDJD setup. The maximum MI achieved at given SNR is plotted in Fig 6.13. We see that maximum MI is reduced due to the transmit filters. In Fig. 6.14, we plot the difference in maximum MI achieved over ACGN channels and AWGN channels. We see that on average 0.21 bits of information are lost because of the filters. The maximum MI loss is 0.27 bits at 12 dB SNR per polarization. We note that these results are only for one polarization. The other polarization has a similar magnitude response and expected to have similar reduction in maximum MI. Moreover, we believe a detection scheme that could exploit the correlation present in the noise samples could improve MI performance.

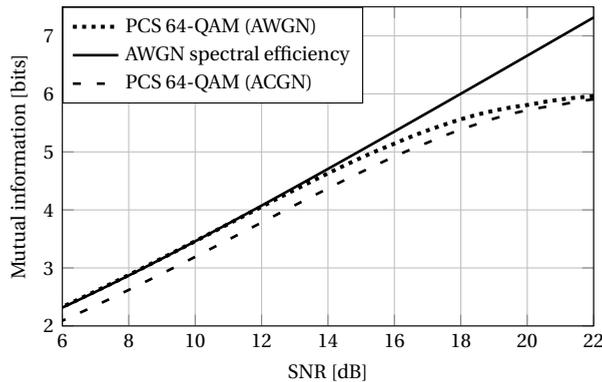


Figure 6.13: Estimated MI over AWGN channel and ACGN channel (i.e. with FDJD filters).

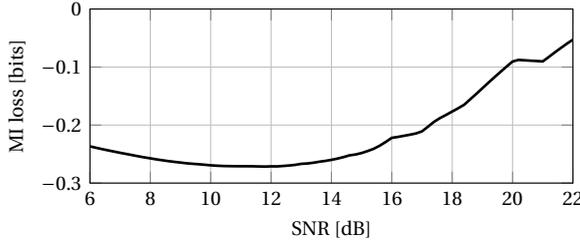


Figure 6.14: Estimated MI loss: difference in MI for ACGN and AWGN channel for PCS-64QAM.

6.8.4. COMPUTATIONAL COMPLEXITY

The JD algorithm processes $N_t = 2^{14}$ (training length) received mixed symbols to determine a 2×2 instantaneous de-mixing matrix G . The dominant computation in this algorithm is the estimation of the correlations $r_x[\tau]$ ($\tau \in \{0, 1\}$), which requires $2 \times 4 \times N_t$ complex multiplications. Therefore, the time complexity of JD algorithm as $\mathcal{O}(8N_t)$. The FDJD algorithm has many computational stages. Here, the major computational cost is due to step 1, 2 and CPRS stage. The overall complexity of these three steps scales with $\mathcal{O}(N_t \log_2 N_t + N_f^3 + N_{mma})$, where N_{mma} is the training length of the MMA.

6.9. CONCLUSION

We investigated algorithms for blind polarization demultiplexing of PCS-QAM signals by inserting temporal correlations at the transmitter and exploiting them at the receiver. We proposed a novel algorithm called frequency domain joint diagonalization (FDJD) pr-MMA (FDJD-pr-MMA), which combines two algorithms, namely FDJD and pr-MMA. The first algorithm, FDJD, removes the cross mixing components of PMD channel and subsequently, the pr-MMA removes residual effects. A proof-of-concept simulation study is presented for a first-order PMD channel. We show that for moderate PMD effects, the FDJD-pr-MMA is well-suited, evaluating it on PMD channels with DGD up to 40% of the symbol duration (T_{symb}). The newly proposed FDJD-pr-MMA algorithm successfully equalizes PMD channels for DGD up to 20% of T_{symb} . Furthermore, we extended our previously proposed JD-MMA [288], which works ideally for RSOP channels, to improve its robustness to PMD effects.

We believe that FDJD estimates can be further improved by correcting the frequency-switching on non-reliable frequencies and by including them in the reconstruction of normalized channel responses. These algorithms' merit should be assessed in a more sophisticated setup using a higher-order PMD model and including other impairments that are often present in coherent transmission system. Finally, the effects of added temporal correlations lend themselves to further investigation as well.

7

CONCLUSIONS AND RECOMMENDATIONS

We assess the achievement of the work completed in this dissertation in aspects of the current state-of-the-art optical coherent communication systems. In addition, some recommendations for further research are discussed.

7.1. CONCLUSIONS

The main objective of this dissertation was to develop novel digital signal-processing techniques to increase data rates of modern fiber-optic communication systems (FOCSs). The throughput of these systems is limited by nonlinear effects such as Kerr effects and transceiver nonlinearity which must be compensated to improve it further. Furthermore, a reduction in the effective data rate of the system caused by the use of training overheads should be avoided. We propose solutions for each of these problems of great practical interest. We addressed a well-known practical challenge in implementing a Kerr-nonlinearity immune transmission technique, known as nonlinear frequency division multiplexing (NFDM). NFDM systems require a lossless fiber model to work with which is not a practical case. Therefore, the path-average approximation was used to design NFDM systems in the literature. We proposed a design of NFDM systems that makes them exact even in the presence of fiber loss, avoiding any errors associated with the path-average approximation. Another practical challenge for NFDM systems is due to the distortions from transceivers. We developed a neural network(NN)-based digital pre-distortion (DPD) technique which when applied on a high symbol rate coherent transmitter improved data rate significantly. Finally, we developed two algorithms that can demultiplex probabilistically shaped signals without needing any training overhead thus avoiding a reduction in the net data rate. Below we explain briefly the motivation, achievements and impact of the work carried out in this dissertation.

NFDM is a transmission technique based on nonlinear Fourier transform (NFT) that linearizes an ideal lossless fiber channel. In NFDM systems, the information is encoded on the nonlinear Fourier spectral components among which cross-talk is absent even though the signal propagates through a nonlinear fiber. NFT applies to a certain class of evolution equations, known as integrable evolution equations. The lossless nonlinear Schrodinger equation (NLSE) that models pulse-propagation an ideal lossless optical fiber, is one of them. However, fibers are not lossless in practice. The presence of loss breaks the integrability property of NLSE therefore the NFT is not exactly applicable. In the literature, a path-average approximation is used to work around wherein a lossy fiber model is approximated to a lossless model by averaging the nonlinear parameter of fiber. However, the propagation described by the path-average approximation deviates from the actual one. Therefore, a perfect nonlinearity compensation cannot be achieved even in the absence of noise. Moreover, the path-average approximation has errors associated with it which scales with link parameters like power, span length and bandwidth. In Chapter 2, we investigated these errors in two well-known NFDM systems from the literature. Furthermore, to avoid the approximation and make the NFDM systems exact, we proposed a solution that uses suitably designed fiber together with a modified NFT. For the first time, we numerically demonstrated an exact NFDM transmission that can achieve a perfect compensation of nonlinearity in noiseless scenarios. We further demonstrate the performance gains obtained by using the proposed approach to design the NFDM systems by comparing them against the ones designed with the path-average approximation. Since the publication of this method, another study was published [173] where other practical challenges for NFDM systems from third-order dispersion and the Raman effect were included. The study concludes that noise and fiber-loss are the limiting factors in practical systems with lumped amplification.

In Chapter 3, we address the nonlinearity of transmitter hardware components. Digital pre-distortion (DPD) techniques are often used to pre-compensate for non-ideal responses of transmitter components. Often, a linear DPD is used in coherent transceivers. However, it only mitigates the limited bandwidth response of the components. Furthermore, signal amplitude swing is kept small to not excite the nonlinear region of the components. This in turn reduces signal power and signal-to-noise (SNR) ratio. A larger swing can improve SNR but may require a nonlinear DPD to pre-compensate the increased nonlinear distortions. A nonlinear DPD technique such as based on memory polynomials and look-up tables was commonly demonstrated over coherent optical transmitters. Another type of DPD technique based on neural networks received attention recently. However, their exploration was limited to a single component or simulated models. Moreover, transceivers of the next generation, which will operate at high symbol rates, were not investigated. These impairments could be stronger in such transceivers because of the large memory effects. For the application of DPD, we investigated a cutting-edge transmitter operating at 128 G symbol rate and probabilistic constellation shaping (PCS)-256 quadrature amplitude modulation (QAM) signaling. We proposed a neural network (NN)-based DPD technique tailored to it. By applying the NN-based DPD, we improved transmit SNR by 1.2 dB. We demonstrated a record net data rate of 1.6 Terabits per second over a single wavelength for an 80 km length of fiber. Furthermore, we compared the performance of the proposed NN-based DPD with that of other popular DPDs. The computational complexity of the NN-DPD was compared against a Volterra series-based DPD. Furthermore, a pruning method was applied to reduce the complexity of the proposed NN-DPD. This work showed that transceiver distortions cannot be neglected in the next-generation transceivers.

In Chapter 4, we investigated joint I-Q compensation at the transmitter DSP by modifying our previously proposed NN-DPD. It can be favorable to compensate transmitter I-Q impairments at the transmitter DSP instead of at the receiver DSP. The latter often performs worse because the signal is noisier at the receiver and contains more impairments due to propagation and demodulation. Furthermore, we evaluated the new joint I-Q NN-DPD over a field-deployed fiber with a WDM configuration. In this demonstration, we achieved a total C-band capacity of 54.5 Terabits over 48 km of field-deployed fiber. Moreover, we evaluated the stability of the NN-DPD over time. We found that the NN-DPD could maintain its peak performance for at least three days.

To further simplify the complexity of NN-DPD, we explored recurrent neural networks (RNNs). RNNs are capable of modeling nonlinearities with very long memory in time-series data. Furthermore, RNNs have shown better performance over feed-forward architecture-based NNs in the application of nonlinear equalization and end-to-end learning. Therefore, we evaluated RNNs for the DPD application in Chapter 5. We evaluated a vanilla RNN, gated recurrent units (GRUs)-based and long short-term memory (LSTM)-based NN against our previously proposed NN-based DPD of Chapter 3 that uses feed-forward architecture. A comparison was carried out in terms of their computational complexity and effectiveness in pre-compensation of an experimentally obtained model of the transmitter presented in Chapter 3. Our comparison showed that to achieve an acceptable performance our previously proposed FFNN-based DPD uses fewer computations compared to the DPDs designed with RNNs. Since the publication of our pro-

posed NN-based DPD, many applications of NN for DPD have been investigated. In [298], a DPD inspired by our proposed NN architecture was investigated for an electrical back-to-back experimental system. The authors found that their NN-DPD trained with indirect learning outperforms Volterra series-based DPD while having similar computational complexity. In follow-up work, we also demonstrated that our NN architecture proposed for DPD could also be used as an equalizer in the receiver DSP [299]. Our approach for training NN-DPD required modeling the channel using an auxiliary NN which was eliminated by employing reinforcement learning in [281]. NNs were also used for a symbol-based DPD in [90], [300], where it was shown to be generalized for DPD of different modulation formats. Furthermore, NN-based DPD was experimentally demonstrated to achieve similar performance to a reduced-size look-up table-based DPD but with lower memory requirements[300].

Finally, in Chapter 6, we propose two novel algorithms that do not require data overhead for polarization demultiplexing of probabilistically shaped signals. Probabilistically shaped QAM signals are used in the new generation of coherent transceivers to maximize the data rate and transmission reach over given channel conditions. But with probabilistic shaping, the distribution of signal tends to be Gaussian which is problematic for the conventional blind polarization demultiplexing algorithms. It is well known that independently and identically distributed (iid) Gaussian signals, when mixed, cannot be recovered/separated from their mixture. To achieve blind-demultiplexing of PCS signals, there exist extended approaches wherein only a part of the signal with non-Gaussian distribution is processed with the conventional algorithm. We proposed a non-conventional method wherein temporal correlations are added to the signal at the transmitter, which are subsequently leveraged at the receiver in order to separate the polarization. Our first proposed algorithm uses a joint diagonalization (JD) stage that performs instantaneous de-mixing of temporally colored signals. Then, a probability-aware (pr) multi-modulus algorithm(MMA) removes any other channel distortions as well as the temporal correlations inserted at the transmitter. This algorithm, which we refer to as JD-pr-MMA, is ideally suitable for memoryless channels. However, our evaluation of JD-pr-MMA over a proof-of-concept simulation found that it is robust to weak polarization mode dispersion (PMD) channels.

The other proposed algorithm uses a frequency domain (FD) JD of the signal followed by some additional processing steps and a pr-MMA. It is referred to as FDJD-pr-MMA. This algorithm can tackle the convolutional mixing (i.e., with memory mixing) of the shaped signals. The FDJD stage of the algorithm removes the cross-mixing between the signals and the pr-MMA stage removes auto-channel responses. We found that the FDJD-pr-MMA is suitable for channels with moderate PMD i.e., We also addressed briefly the impact of impairments like phase noise and frequency offset on the proposed algorithms. Furthermore, a discussion on the computational complexity of the algorithms and integration with matched filtering is also provided. We demonstrate the effectiveness of this idea and believe it opens up new research avenues that can be useful for designing innovative algorithms.

7.2. RECOMMENDATIONS FOR FUTURE WORK

In this section, we discuss some of the possible research directions on the topic of digital signal processing for coherent fiber-optic communication systems.

Exact NFDN Transmission in the Presence of Fiber-Loss We assessed one of the practical challenges in implementing NFDN systems which is the fiber-loss. We compared the performance of our exact NFDN system against the one designed with the path-average model in simulations. The comparative study of gains on an experimental system could be interesting. Our proposed exact NFDN systems require dispersion-decreasing fibers, which are difficult to fabricate. An approximation of dispersion decreasing profile by using fibers with constant dispersion but of different values and an evaluation of such a NFDN system could be interesting. Additionally, in our framework of comparison, we chose to keep transmitter requirements identical for the two systems under investigation. However, a comparison framework solely designed to maximize data rate or spectral efficiency could also be useful. Finally, in our study, we only considered fiber-loss and amplifier noise as impairments. The impact of fiber-loss and noise is found as limiting factors for the performance of NFDN systems [173]. We already proposed a solution for fiber-loss problem. However, the impact of noise is still a critical challenge for NFDN systems which is a potential topic for further research.

NN-based DPD of a coherent transmitter Our proposed NN-based DPD technique provided significant gains in SNR compared to other DPDs. It required a much larger number of computations. To address this, we applied a pruning technique to reduce the computational complexity. However, it is still complex to be implemented in real systems. Therefore, there is an interest in exploring other NN-DPD architectures and deep learning techniques that are computationally efficient.

Blind polarization-demultiplexing of shaped signals, The algorithms that we proposed were tested in proof-of-concept numerical simulations that utilized a first-order PMD model. We briefly discussed the impact of impairments like phase noise and frequency offset on them. We believe that conducting an experimental assessment of these algorithms in the presence of impairments such as higher-order PMD effects, sampling clock drift, and residual dispersion, is important. Furthermore, the proposed algorithms are computationally intensive compared to existing ones and need to be simplified. Finally, the algorithms work by adding temporal correlations at the transmitter. This concept is unconventional in optical fiber communication systems. Investigating the impact of these correlations on other subsystems and algorithms is necessary.

BIBLIOGRAPHY

- [1] “Digital Economy: Value Creation and Capture - Implications for Developing Countries”, in *United Nations Conference on Trade and Development*, United Nations Publications, 2019.
- [2] D. Richardson, “New optical fibres for high-capacity optical communications”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2062, p. 20 140 441, 2016.
- [3] K. C. Kao and G. A. Hockham, “Dielectric-fibre surface waveguides for optical frequencies”, in *Proceedings of the Institution of Electrical Engineers*, IET, vol. 113, 1966, pp. 1151–1158.
- [4] J. Lytollis, “Optical communication systems—a survey”, *Optics Technology*, vol. 1, no. 1, pp. 1–10, 1968.
- [5] F. Kapron, D. B. Keck, and R. D. Maurer, “Radiation losses in glass optical waveguides”, *Applied Physics Letters*, vol. 17, no. 10, pp. 423–425, 1970.
- [6] R. D. Maurer, “Glass fibers for optical communications”, *Proceedings of the IEEE*, vol. 61, no. 4, pp. 452–462, 1973.
- [7] E. Basch, R. Beaudette, and H. Carnes, “Optical transmission for interoffice trunks”, *IEEE Transactions on Communications*, vol. 26, no. 7, pp. 1007–1014, 1978.
- [8] *Fiber-Optic Technologies*, <https://www.ciscopress.com/articles/article.asp?p=170740>, Accessed: 2023-04-02.
- [9] G. P. Agrawal, *Fiber-Optic Communication Systems*. John Wiley & Sons, 2012.
- [10] T. Li, “Advances in optical fiber communications: An historical perspective”, *IEEE Journal on Selected Areas in Communications*, vol. 1, no. 3, pp. 356–372, 1983.
- [11] R. Stolen and A. Ashkin, “Optical kerr effect in glass waveguide”, *Applied Physics Letters*, vol. 22, no. 6, pp. 294–296, 1973.
- [12] D. N. Payne and W. Gambling, “Zero material dispersion in optical fibres”, *Electronics Letters*, vol. 8, no. 11, pp. 176–178, 1975.
- [13] Y. Suematsu, “Long-wavelength optical fiber communication”, *Proceedings of the IEEE*, vol. 71, no. 6, pp. 692–721, 1983.
- [14] F. Kapron and D. Keck, “Pulse transmission through a dielectric optical waveguide”, *Applied optics*, vol. 10, no. 7, pp. 1519–1523, 1971.
- [15] P. Urquhart, “Review of rare earth doped fibre lasers and amplifiers”, *IEE Proceedings J (Optoelectronics)*, vol. 135, no. 6, pp. 385–407, 1988.

- [16] N. Edagawa, "904 km, 1.2 Gbit/s non-regenerative optical transmission experiment using 12 Erbium-doped fiber amplifiers", *ECOC'89, Gothenburg, Sweden, post deadline paper PDA-8*, 1989.
- [17] T. Okoshi and K. Kikuchi, *Coherent optical fiber communications*. Springer Science & Business Media, 1988, vol. 4.
- [18] S. Saito, T. Imai, and T. Ito, "An over 2200-km coherent transmission experiment at 2.5 Gb/s using erbium-doped-fiber in-line amplifiers", *Journal of Lightwave Technology*, vol. 9, no. 2, pp. 161–169, 1991. DOI: [10.1109/50.65873](https://doi.org/10.1109/50.65873).
- [19] E. Desurvire, "The golden age of optical fiber amplifiers.", *Physics Today*, vol. 47, no. 1, pp. 20–27, 1994.
- [20] M. J. Yadlowsky, E. M. Deliso, and V. L. Da Silva, "Optical fibers and amplifiers for wdm systems", *Proceedings of the IEEE*, vol. 85, no. 11, pp. 1765–1779, 1997.
- [21] J.-M. Beaufils, "How do submarine networks web the world?", *Optical fiber technology*, vol. 6, no. 1, pp. 15–32, 2000.
- [22] G. R. Hill, "Wavelength domain optical network techniques", *Proceedings of the IEEE*, vol. 78, no. 1, pp. 121–132, 1990.
- [23] G. E. Keiser, "A review of WDM technology and applications", *Optical Fiber Technology*, vol. 5, no. 1, pp. 3–39, 1999.
- [24] S. Aisawa, T. Sakamoto, M. Fukui, J. Kani, M. Jinno, and K. Oguchi, "Ultra-wideband, long distance WDM demonstration of 1 Tbit/s (50 × 20 Gbit/s), 600 km transmission using 1550 and 1580 nm wavelength bands", *Electronics Letters*, vol. 34, no. 11, pp. 1127–1129, 1998.
- [25] A. Hasegawa, "Soliton-based optical communications: An overview", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 6, pp. 1161–1172, 2000.
- [26] A. Pratt, P. Harper, S. Alleston, *et al.*, "5,745 km DWDM transcontinental field trial using 10 Gbit/s dispersion managed solitons and dynamic gain equalization", in *Optical Fiber Communication Conference*, Optica Publishing Group, 2003, PD26.
- [27] Y. Yamamoto, "Receiver performance evaluation of various digital optical modulation-demodulation systems in the 0.5-10 μm wavelength region", *IEEE Journal of Quantum Electronics*, vol. 16, no. 11, pp. 1251–1259, 1980. DOI: [10.1109/JQE.1980.1070400](https://doi.org/10.1109/JQE.1980.1070400).
- [28] M. Grant, W. Michie, and M. Fletcher, "The performance of optical phase-locked loops in the presence of nonnegligible loop propagation delay", *Journal of lightwave technology*, vol. 5, no. 4, pp. 592–597, 1987.
- [29] R. Noe, "Phase noise tolerant synchronous QPSK receiver concept with digital I and Q baseband processing", in *OECC/COIN 2004: 9th OptoElectronics and Communications Conference and 3rd International Conference on Optical Internet*, Institute of Electronics, Information and Communication Engineers Kanagawa, 2004, pp. 818–819.

- [30] S. Tsukamoto, D.-S. Ly-Gagnon, K. Katoh, and K. Kikuchi, “Coherent demodulation of 40-Gbit/s polarization-multiplexed QPSK signals with 16-GHz spacing after 200-km transmission”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2005, PDP29.
- [31] D.-S. Ly-Gagnon, S. Tsukamoto, K. Katoh, and K. Kikuchi, “Coherent detection of optical quadrature phase-shift keying signals with carrier phase estimation”, *Journal of lightwave technology*, vol. 24, no. 1, p. 12, 2006.
- [32] S. Tsukamoto, Y. Ishikawa, and K. Kikuchi, “Optical homodyne receiver comprising phase and polarization diversities with digital signal processing”, in *2006 European Conference on Optical Communications*, IEEE, 2009, pp. 1–2.
- [33] J. McNicol, M. O’sullivan, K. Roberts, A. Comeau, D. McGhan, and L. Strawczynski, “Electrical domain compensation of optical dispersion [optical fibre communication applications]”, in *OFC/NFOEC Technical Digest. Optical Fiber Communication Conference, 2005.*, IEEE, vol. 4, 2005, 3–pp.
- [34] E. Ip and J. M. Kahn, “Digital equalization of chromatic dispersion and polarization mode dispersion”, *Journal of Lightwave Technology*, vol. 25, no. 8, pp. 2033–2043, 2007.
- [35] H. Sun, K.-T. Wu, and K. Roberts, “Real-time measurements of a 40 Gb/s coherent system”, *Opt. Express*, vol. 16, no. 2, pp. 873–879, Jan. 2008. DOI: [10.1364/OE.16.000873](https://doi.org/10.1364/OE.16.000873).
- [36] P. J. Winzer and D. T. Neilson, “From scaling disparities to integrated parallelism: A decathlon for a decade”, *Journal of Lightwave Technology*, vol. 35, no. 5, pp. 1099–1115, 2017.
- [37] J.-X. Cai, H. Batshon, M. V. Mazurczyk, *et al.*, “70.4 Tb/s capacity over 7,600 km in C+ L band using coded modulation with hybrid constellation shaping and nonlinearity compensation”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2017, Th5B–2.
- [38] R. Emmerich, M. Sena, R. Elschner, *et al.*, “Enabling SCL-band systems with standard C-band modulator and coherent receiver using coherent system identification and nonlinear predistortion”, *Journal of Lightwave Technology*, vol. 40, no. 5, pp. 1360–1368, 2021.
- [39] B. J. Puttnam, G. Rademacher, and R. S. Luís, “Space-division multiplexing for optical fiber communications”, *Optica*, vol. 8, no. 9, pp. 1186–1203, 2021.
- [40] A. Leven, N. Kaneda, and S. Corteselli, “Real-time implementation of digital signal processing for coherent optical digital communication systems”, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1227–1234, 2010.
- [41] “Forward error correction for submarine systems (2018–2023)”, ITU-T, Tech. Rep., 2000. [Online]. Available: <https://www.itu.int/rec/T-REC-G.975>.
- [42] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, “Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration”, *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1599–1609, 2016.

- [43] J. Cho and P. J. Winzer, “Probabilistic Constellation Shaping for Optical Fiber Communications”, *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.
- [44] T. M. Cover and J. A. Thomas, “Information theory and statistics”, *Elements of information theory*, vol. 1, no. 1, pp. 279–335, 1991.
- [45] F. R. Kschischang and S. Pasupathy, “Optimal nonuniform signaling for Gaussian channels”, *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 913–929, 1993.
- [46] P. Schulte and G. Böcherer, “Constant composition distribution matching”, *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, 2016. DOI: [10.1109/TIT.2015.2499181](https://doi.org/10.1109/TIT.2015.2499181).
- [47] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic shaping and forward error correction for fiber-optic communication systems”, *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 230–244, 2019.
- [48] A. Das, *Digital Communication: Principles and system modelling*. Springer Science & Business Media, 2010.
- [49] I. Fatadin, S. J. Savory, and D. Ives, “Compensation of quadrature imbalance in an optical QPSK coherent receiver”, *IEEE Photonics Technology Letters*, vol. 20, no. 20, pp. 1733–1735, 2008.
- [50] P.-O. Löwdin, “On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals”, *The Journal of Chemical Physics*, vol. 18, no. 3, pp. 365–375, 1950.
- [51] I. Mayer, “On löwdin’s method of symmetric orthogonalization”, *International Journal of Quantum Chemistry*, vol. 90, no. 1, pp. 63–65, 2002.
- [52] T. Tanimura, S. Oda, T. Tanaka, T. Hoshida, Z. Tao, and J. C. Rasmussen, “A simple digital skew compensator for coherent receiver”, in *2009 35th European Conference on Optical Communication*, IEEE, 2009, pp. 1–2.
- [53] J. C. M. Diniz, E. P. da Silva, M. Piels, and D. Zibar, “Joint IQ skew and chromatic dispersion estimation for coherent optical communication receivers”, in *Signal Processing in Photonic Communications*, Optica Publishing Group, 2016, SpTu2F–2.
- [54] M. S. Faruk and K. Kikuchi, “Compensation for in-phase/quadrature imbalance in coherent-receiver front end for optical quadrature amplitude modulation”, *IEEE Photonics Journal*, vol. 5, no. 2, pp. 7 800 110–7 800 110, 2013.
- [55] M. Paskov, D. Lavery, and S. J. Savory, “Blind equalization of receiver in-phase/quadrature skew in the presence of nyquist filtering”, *IEEE Photonics Technology Letters*, vol. 25, no. 24, pp. 2446–2449, 2013.
- [56] S. Savory, “Digital equalization in coherent optical transmission systems”, *Enabling Technologies for High Spectral-Efficiency Coherent Optical Communication Networks*, pp. 311–332, 2016.

- [57] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems", *IEEE transactions on communications*, vol. 28, no. 11, pp. 1867–1875, 1980.
- [58] K. Kikuchi, "Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers", *Optics Express*, vol. 19, no. 10, pp. 9868–9880, 2011.
- [59] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic press, 2010.
- [60] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review", *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, 1998.
- [61] 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.
- [62] B. Szafraniec, B. Nebendahl, and T. Marshall, "Polarization demultiplexing in stokes space", *Optics express*, vol. 18, no. 17, pp. 17 928–17 939, 2010.
- [63] S. A. Athuraliya and L. M. Garth, "Quantized CMA equalization for shaped signal constellations", *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 67–70, 2004.
- [64] S. Dris, S. Alreesh, and A. Richter, "Blind Polarization Demultiplexing and Equalization of Probabilistically Shaped QAM", in *Optical Fiber Communication Conference*, Optica Publishing Group, 2019, W1D–2.
- [65] F. Gardner, "A BPSK/QPSK timing-error detector for sampled receivers", *IEEE Transactions on communications*, vol. 34, no. 5, pp. 423–429, 1986.
- [66] K. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers", *IEEE transactions on communications*, vol. 24, no. 5, pp. 516–531, 1976.
- [67] M. Oerder and H. Meyr, "Digital filter and square timing recovery", *IEEE Transactions on communications*, vol. 36, no. 5, pp. 605–612, 1988.
- [68] A. Leven, N. Kaneda, U.-V. Koc, and Y.-K. Chen, "Frequency estimation in intradyne reception", *IEEE Photonics Technology Letters*, vol. 19, no. 6, pp. 366–368, 2007.
- [69] Q. Yan, L. Liu, and X. Hong, "Blind Carrier Frequency Offset Estimation in Coherent Optical Communication Systems With Probabilistically Shaped M-QAM", *Journal of Lightwave Technology*, vol. 37, no. 23, pp. 5856–5866, 2019.
- [70] J. R. Barry and E. A. Lee, "Performance of coherent optical receivers", *Proceedings of the IEEE*, vol. 78, no. 8, pp. 1369–1394, 1990.
- [71] 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.

- [72] T. Pfau, S. Hoffmann, and R. Noé, “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.
- [73] A. Spalvieri and L. Barletta, “Pilot-aided carrier recovery in the presence of phase noise”, *IEEE transactions on communications*, vol. 59, no. 7, pp. 1966–1974, 2011.
- [74] M. Seimetz, “Laser linewidth limitations for optical systems with high-order modulation employing feed forward digital carrier phase estimation”, in *optical fiber communication conference*, Optica Publishing Group, 2008, OTuM2.
- [75] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks”, *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.
- [76] C. E. Shannon, “Communication in the presence of noise”, *Proceedings of the IEEE*, vol. 72, no. 9, pp. 1192–1201, 1984.
- [77] E. Ip and J. M. Kahn, “Compensation of dispersion and nonlinear impairments using digital backpropagation”, *Journal of Lightwave Technology*, vol. 26, no. 20, pp. 3416–3425, 2008.
- [78] F. P. Guiomar, J. D. Reis, A. L. Teixeira, and A. N. Pinto, “Digital postcompensation using Volterra series transfer function”, *IEEE Photonics Technology Letters*, vol. 23, no. 19, pp. 1412–1414, 2011.
- [79] A. Mecozzi and R.-J. Essiambre, “Nonlinear shannon limit in pseudolinear coherent systems”, *Journal of Lightwave Technology*, vol. 30, no. 12, pp. 2011–2024, 2012. DOI: [10.1109/JLT.2012.2190582](https://doi.org/10.1109/JLT.2012.2190582).
- [80] M. Malekiha, I. Tselniker, and D. V. Plant, “Efficient nonlinear equalizer for intra-channel nonlinearity compensation for next generation agile and dynamically reconfigurable optical networks”, *Optics express*, vol. 24, no. 4, pp. 4097–4108, 2016.
- [81] M. I. Yousefi and F. R. Kschischang, “Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools”, *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4312–4328, 2014.
- [82] S. K. Turitsyn, J. E. Prilepsky, S. T. Le, *et al.*, “Nonlinear Fourier transform for optical data processing and transmission: Advances and perspectives”, *Optica*, vol. 4, no. 3, pp. 307–322, 2017.
- [83] H. Bülow, “Experimental demonstration of optical signal detection using nonlinear Fourier transform”, *Journal of Lightwave Technology*, vol. 33, no. 7, pp. 1433–1439, 2015.
- [84] S. Civelli, E. Forestieri, and M. Secondini, “Why Noise and Dispersion May Seriously Hamper Nonlinear Frequency-Division Multiplexing”, *IEEE Photonics Technology Letters*, vol. 29, no. 16, pp. 1332–1335, 2017.
- [85] V. Oliari, S. Goossens, C. Häger, *et al.*, “Revisiting efficient multi-step nonlinearity compensation with machine learning: An experimental demonstration”, *Journal of Lightwave Technology*, vol. 38, no. 12, pp. 3114–3124, 2020. DOI: [10.1109/JLT.2020.2994220](https://doi.org/10.1109/JLT.2020.2994220).

- [86] B. I. Bitachon, A. Ghazisaeidi, M. Eppenberger, B. Baeuerle, M. Ayata, and J. Leuthold, “Deep learning based digital backpropagation demonstrating snr gain at low complexity in a 1200 km transmission link”, *Optics Express*, vol. 28, no. 20, pp. 29 318–29 334, 2020.
- [87] P. W. Berenguer, M. Nölle, L. Molle, *et al.*, “Nonlinear Digital Pre-Distortion of Transmitter Components”, *J. Lightw. Technol.*, vol. 34, no. 8, pp. 1739–1745, 2015.
- [88] G. Paryanti, H. Faig, L. Rokach, and D. Sadot, “A Direct Learning Approach for Neural Network Based Pre-Distortion for Coherent Nonlinear Optical Transmitter”, *Journal of Lightwave Technology*, vol. 38, no. 15, pp. 3883–3896, 2020.
- [89] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, “Deep neural network-based digital pre-distortion for high baudrate optical coherent transmission”, *Journal of Lightwave Technology*, vol. 40, no. 3, pp. 597–606, 2022.
- [90] Z. He, J. Song, C. Häger, *et al.*, “Symbol-Based Supervised Learning Predistortion for Compensating Transmitter Nonlinearity”, in *2021 European Conference on Optical Communication (ECOC)*, IEEE, 2021, pp. 1–4.
- [91] W. Shieh and K.-P. Ho, “Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing”, *Optics Express*, vol. 16, no. 20, pp. 15 718–15 727, 2008.
- [92] M. Qiu, Q. Zhuge, M. Chagnon, *et al.*, “Digital subcarrier multiplexing for fiber nonlinearity mitigation in coherent optical communication systems”, *Optics Express*, vol. 22, no. 15, pp. 18 770–18 777, 2014.
- [93] H. Sun, M. Torbatian, M. Karimi, *et al.*, “800G DSP ASIC Design Using Probabilistic Shaping and Digital Sub-Carrier Multiplexing”, *Journal of Lightwave Technology*, 2020.
- [94] D. Welch, A. Napoli, J. Bäck, *et al.*, “Point-to-multipoint optical networks using coherent digital subcarriers”, *Journal of Lightwave Technology*, vol. 39, no. 16, pp. 5232–5247, 2021.
- [95] T. Duthel, C. R. Fludger, B. Liu, *et al.*, “Dsp design for coherent optical point-to-multipoint transmission”, *Journal of Lightwave Technology*, 2023.
- [96] P. Zhang, Q. Yan, and X. Hong, “Probability-Aware Stokes Space Blind Polarization Demultiplexing for Probabilistically Shaped Signals”, *Journal of Lightwave Technology*, vol. 39, no. 19, pp. 6120–6129, 2021.
- [97] V. Lauinger, F. Buchali, and L. Schmalen, “Blind equalization for coherent optical communications based on variational inference”, in *Signal Processing in Photonic Communications*, Optica Publishing Group, 2021, SpTh1D–6.
- [98] F. A. Barbosa, S. M. Rossi, and D. A. Mello, “Phase and Frequency Recovery Algorithms for Probabilistically Shaped Transmission”, *Journal of Lightwave Technology*, vol. 38, no. 7, pp. 1827–1835, 2019.
- [99] L. Zhang, K. Tao, W. Qian, *et al.*, “Real-time FPGA investigation of interplay between probabilistic shaping and forward error correction”, *Journal of Lightwave Technology*, vol. 40, no. 5, pp. 1339–1345, 2022.

- [100] J. Man, W. Chen, X. Song, and L. Zeng, “A low-cost 100GE optical transceiver module for 2km SMF interconnect with PAM4 modulation”, in *OFC 2014*, 2014, pp. 1–3. DOI: [10.1364/OFC.2014.M2E.7](https://doi.org/10.1364/OFC.2014.M2E.7).
- [101] K. Zhong, X. Zhou, J. Huo, C. Yu, C. Lu, and A. P. T. Lau, “Digital signal processing for short-reach optical communications: A review of current technologies and future trends”, *Journal of Lightwave Technology*, vol. 36, no. 2, pp. 377–400, 2018.
- [102] J. Estaran, R. Rios-Müller, M. Mestre, *et al.*, “Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IM/DD systems”, in *ECOC 2016; 42nd European Conference on Optical Communication*, VDE, 2016, pp. 1–3.
- [103] B. Karanov, M. Chagnon, F. Thouin, *et al.*, “End-to-end deep learning of optical fiber communications”, *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018.
- [104] A. Mecozzi, C. Antonelli, and M. Shtaif, “Kramers–Kronig coherent receiver”, *Optica*, vol. 3, no. 11, pp. 1220–1227, 2016.
- [105] D. Che, A. Li, X. Chen, Q. Hu, Y. Wang, and W. Shieh, “Stokes vector direct detection for linear complex optical channels”, *Journal of Lightwave Technology*, vol. 33, no. 3, pp. 678–684, 2014.
- [106] D. Che, C. Sun, and W. Shieh, “Optical field recovery in Stokes space”, *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 451–460, 2019.
- [107] M. Matsumoto, “Optical signal phase reconstruction based on temporal transport-of-intensity equation”, *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4722–4729, 2020.
- [108] R. Nagarajan, I. Lyubomirsky, and O. Agazzi, “Low power DSP-based transceivers for data center optical fiber communications (invited tutorial)”, *Journal of Lightwave Technology*, vol. 39, no. 16, pp. 5221–5231, 2021.
- [109] T. Kupfer, A. Bisplinghof, T. Duthel, C. Fludger, and S. Langenbach, “Optimizing power consumption of a coherent DSP for metro and data center interconnects”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2017, Th3G–2.
- [110] E. Börjeson and P. Larsson-Edefors, “Energy-efficient implementation of carrier phase recovery for higher-order modulation formats”, *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 505–510, 2021. DOI: [10.1109/JLT.2020.3027781](https://doi.org/10.1109/JLT.2020.3027781).
- [111] F. N. Hauske, M. Kuschnerov, B. Spinnler, and B. Lankl, “Optical performance monitoring in digital coherent receivers”, *Journal of Lightwave Technology*, vol. 27, no. 16, pp. 3623–3631, 2009.
- [112] Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, “Optical performance monitoring: A review of current and future technologies”, *Journal of Lightwave Technology*, vol. 34, no. 2, pp. 525–543, 2015.
- [113] D. Wang, H. Jiang, G. Liang, *et al.*, “Optical performance monitoring of multiple parameters in future optical networks”, *Journal of Lightwave Technology*, vol. 39, no. 12, pp. 3792–3800, 2020.

- [114] T. Tanimura, S. Yoshida, K. Tajima, S. Oda, and T. Hoshida, "Fiber-longitudinal anomaly position identification over multi-span transmission link out of receiver-end signals", *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2726–2733, 2020.
- [115] M. Lonardi, J. Pesic, P. Jennevé, *et al.*, "Optical nonlinearity monitoring and launch power optimization by artificial neural networks", *Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2637–2645, 2020.
- [116] F. Boitier, V. Lemaire, J. Pesic, *et al.*, "Proactive fiber damage detection in real-time coherent receiver", in *2017 European Conference on Optical Communication (ECOC)*, IEEE, 2017, pp. 1–3.
- [117] W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Machine learning techniques for optical performance monitoring and modulation format identification: A survey", *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2839–2882, 2020.
- [118] V. Zakharov and A. Shabat, "Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media", *Sov. Phys. JETP*, vol. 34, no. 1, p. 62, 1972.
- [119] D. D. J. Korteweg and D. G. de Vries, "Xli. on the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 39, no. 240, pp. 422–443, 1895. DOI: [10.1080/14786449508620739](https://doi.org/10.1080/14786449508620739).
- [120] C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Method for solving the Korteweg-deVries equation", *Physical review letters*, vol. 19, no. 19, p. 1095, 1967.
- [121] P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves", *Communications on pure and applied mathematics*, vol. 21, no. 5, pp. 467–490, 1968.
- [122] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, "The inverse scattering transform-fourier analysis for nonlinear problems", *Studies in applied mathematics*, vol. 53, no. 4, pp. 249–315, 1974.
- [123] *What is integrability*. Springer Science & Business Media, 1991.
- [124] M. Wadati, "The modified korteweg-de vries equation", *Journal of the Physical Society of Japan*, vol. 34, no. 5, pp. 1289–1296, 1973.
- [125] J. Rubinstein, "Sine-gordon equation", *Journal of Mathematical Physics*, vol. 11, no. 1, pp. 258–266, 1970.
- [126] A. Hasegawa and F. Tappert, "Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. i. anomalous dispersion", *Applied Physics Letters*, vol. 23, no. 3, pp. 142–144, 1973.
- [127] K. Tajima, "Compensation of soliton broadening in nonlinear optical fibers with loss", *Optics letters*, vol. 12, no. 1, pp. 54–56, 1987.
- [128] H. Kuehl, "Solitons on an axially nonuniform optical fiber", *JOSA B*, vol. 5, no. 3, pp. 709–713, 1988.

- [129] S. V. Chernikov and P. V. Mamyshev, “Femtosecond soliton propagation in fibers with slowly decreasing dispersion”, *JOSA B*, vol. 8, no. 8, pp. 1633–1641, 1991.
- [130] A. Hasegawa and Y. Kodama, “Guiding-center soliton in optical fibers”, *Optics letters*, vol. 15, no. 24, pp. 1443–1445, 1990.
- [131] J. Gordon and H. Haus, “Opt. lett.”, *Opt. Lett.*, vol. 11, no. 662, 1986.
- [132] A. Mecozzi, J. D. Moores, H. A. Haus, and Y. Lai, “Soliton transmission control”, *Optics letters*, vol. 16, no. 23, pp. 1841–1843, 1991.
- [133] L. F. Mollenauer, J. P. Gordon, and S. G. Evangelides, “The sliding-frequency guiding filter: An improved form of soliton jitter control”, *Optics letters*, vol. 17, no. 22, pp. 1575–1577, 1992.
- [134] M. I. Yousefi and F. R. Kschischang, “Information Transmission Using the Non-linear Fourier Transform, Part III: Spectrum Modulation”, *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4346–4369, 2014.
- [135] L. F. Mollenauer, S. G. Evangelides, and J. P. Gordon, “Wavelength division multiplexing with solitons in ultra-long distance transmission using lumped amplifiers”, *Journal of lightwave technology*, vol. 9, no. 3, pp. 362–367, 1991.
- [136] Y. Matsuda, H. Terauchi, and A. Maruta, “Design of eigenvalue-multiplexed multi-level modulation optical transmission system”, in *2014 OptoElectronics and Communication Conference and Australian Conference on Optical Fibre Technology*, 2014, pp. 1016–1018.
- [137] A. Maruta, “Eigenvalue modulated optical transmission system”, in *2015 Opto-Electronics and Communications Conference (OECC)*, 2015, pp. 1–3. DOI: [10.1109/OECC.2015.7340112](https://doi.org/10.1109/OECC.2015.7340112).
- [138] S. Hari, M. I. Yousefi, and F. R. Kschischang, “Multieigenvalue communication”, *Journal of Lightwave Technology*, vol. 34, no. 13, pp. 3110–3117, 2016.
- [139] H. Bülow, V. Aref, and L. Schmalen, “Modulation on Discrete Nonlinear Spectrum: Perturbation Sensitivity and Achievable Rates”, *IEEE Photonics Technology Letters*, vol. 30, no. 5, pp. 423–426, 2018.
- [140] T. Gui, Z. Dong, C. Lu, P.-K. A. Wai, and A. P. T. Lau, “Phase modulation on nonlinear discrete spectrum for nonlinear frequency division multiplexed transmissions”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2016, W3A–2.
- [141] V. Aref and H. Buelow, “Design of 2-Soliton Spectral Phase Modulated Pulses Over Lumped Amplified Link”, in *ECOC 2016; 42nd European Conference on Optical Communication*, VDE, 2016, pp. 1–3.
- [142] H. Buelow, V. Aref, and W. Idler, “Transmission of Waveforms Determined by 7 Eigenvalues with PSK-Modulated Spectral Amplitudes”, in *ECOC 2016; 42nd European Conference on Optical Communication*, VDE, 2016, pp. 1–3.
- [143] T. Gui, T. H. Chan, C. Lu, A. P. T. Lau, and P.-K. A. Wai, “Alternative decoding methods for optical communications based on nonlinear fourier transform”, *Journal of Lightwave Technology*, vol. 35, no. 9, pp. 1542–1550, 2017.

- [144] S. T. Le, V. Aref, and H. Buelow, “High speed precompensated nonlinear frequency-division multiplexed transmissions”, *Journal of Lightwave Technology*, vol. 36, no. 6, pp. 1296–1303, 2018. DOI: [10.1109/JLT.2017.2787185](https://doi.org/10.1109/JLT.2017.2787185).
- [145] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, “Nonlinear inverse synthesis technique for optical links with lumped amplification”, *Optics express*, vol. 23, no. 7, pp. 8317–8328, 2015.
- [146] X. Yangzhang, S. T. Le, V. Aref, H. Buelow, D. Lavery, and P. Bayvel, “Experimental demonstration of dual-polarisation nfdm transmission with b-modulation”, in *2019 IEEE Photonics Conference (IPC)*, IEEE, 2019, pp. 1–4.
- [147] M. Balogun, S. Derevyanko, and L. Barry, “Dual-Polarization Hermite-Gaussian-Based NFDM Transmission System”, *IEEE Access*, vol. 11, pp. 125 692–125 697, 2023.
- [148] I. Tavakkolnia and M. Safari, “Signalling over nonlinear fibre-optic channels by utilizing both solitonic and radiative spectra”, in *2015 European Conference on Networks and Communications (EuCNC)*, IEEE, 2015, pp. 103–107.
- [149] V. Aref, S. T. Le, and H. Buelow, “Demonstration of fully nonlinear spectrum modulated system in the highly nonlinear optical transmission regime”, in *ECOC 2016-Post Deadline Paper; 42nd European Conference on Optical Communication*, VDE, 2016, pp. 1–3.
- [150] V. Aref, S. T. Le, and H. Buelow, “Modulation over nonlinear fourier spectrum: Continuous and discrete spectrum”, *Journal of Lightwave Technology*, vol. 36, no. 6, pp. 1289–1295, 2018.
- [151] V. Aref, H. Bülow, K. Schuh, and W. Idler, “Experimental demonstration of nonlinear frequency division multiplexed transmission”, in *2015 European Conference on Optical Communication (ECOC)*, IEEE, 2015, pp. 1–3.
- [152] A. Geisler and C. G. Schaeffer, “Experimental nonlinear frequency division multiplexed transmission using eigenvalues with symmetric real part”, in *ECOC 2016; 42nd European Conference on Optical Communication*, VDE, 2016, pp. 1–3.
- [153] A. Span, V. Aref, H. Bülow, and S. Ten Brink, “On time-bandwidth product of multi-soliton pulses”, in *2017 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2017, pp. 61–65.
- [154] T. Gui, C. Lu, A. P. T. Lau, and P. Wai, “High-order modulation on a single discrete eigenvalue for optical communications based on nonlinear fourier transform”, *Optics express*, vol. 25, no. 17, pp. 20 286–20 297, 2017.
- [155] A. Span, V. Aref, H. Bülow, and S. ten Brink, “Time-bandwidth product perspective for nonlinear fourier transform-based multi-eigenvalue soliton transmission”, *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5544–5557, 2019.
- [156] G. Zhou, T. Gui, C. Lu, A. P. T. Lau, and P.-K. A. Wai, “Improving soliton transmission systems through soliton interactions”, *Journal of Lightwave Technology*, vol. 38, no. 14, pp. 3563–3572, 2019.

- [157] S. T. Le, J. E. Prilepsky, and S. K. Turitsyn, “Nonlinear inverse synthesis for high spectral efficiency transmission in optical fibers”, *Optics express*, vol. 22, no. 22, pp. 26 720–26 741, 2014.
- [158] S. Le, S. Wahls, D. Lavery, J. Prilepsky, and S. Turitsyn, “Reduced complexity nonlinear inverse synthesis for nonlinearity compensation in optical fiber links”, in *The European Conference on Lasers and Electro-Optics*, Optica Publishing Group, 2015, CI_3_2.
- [159] S. Wahls, “Generation of Time-Limited Signals in the Nonlinear Fourier Domain via b-Modulation”, in *2017 European Conference on Optical Communication (ECOC)*, IEEE, 2017, pp. 1–3.
- [160] T. Gui, G. Zhou, C. Lu, A. P. T. Lau, and S. Wahls, “Nonlinear frequency division multiplexing with b-modulation: Shifting the energy barrier”, *Optics express*, vol. 26, no. 21, pp. 27 978–27 990, 2018.
- [161] S. T. Le, V. Aref, and H. Buelow, “Nonlinear signal multiplexing for communication beyond the kerr nonlinearity limit”, *Nature Photonics*, vol. 11, no. 9, pp. 570–576, 2017.
- [162] S. Gaiarin, A. M. Perego, E. P. da Silva, F. Da Ros, and D. Zibar, “Dual-polarization nonlinear fourier transform-based optical communication system”, *Optica*, vol. 5, no. 3, pp. 263–270, 2018.
- [163] W. A. Gemechu, T. Gui, J.-W. Goossens, *et al.*, “Dual polarization nonlinear frequency division multiplexing transmission”, *IEEE Photonics Technology Letters*, vol. 30, no. 18, pp. 1589–1592, 2018.
- [164] X. Yangzhang, V. Aref, S. T. Le, H. Buelow, D. Lavery, and P. Bayvel, “Dual-Polarization Non-Linear Frequency-Division Multiplexed Transmission With b-Modulation”, *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1570–1578, 2019.
- [165] F. Da Ros, S. Civelli, S. Gaiarin, *et al.*, “Dual-polarization nfdm transmission with continuous and discrete spectral modulation”, *Journal of Lightwave Technology*, vol. 37, no. 10, pp. 2335–2343, 2019.
- [166] J. Wei, L. Xi, X. Zhang, *et al.*, “Improvement for a full-spectrum modulated nonlinear frequency division multiplexing transmission system”, *Optics Express*, vol. 30, no. 17, pp. 31 195–31 208, 2022.
- [167] S. Wahls, “Shortening solitons for fiber-optic transmission”, in *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, IEEE, 2021, pp. 1–6.
- [168] S. Wahls, “Full spectrum b-modulation of time-limited signals using linear programming”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2022, Th2A–38.
- [169] F. Da Ros, S. Gaiarin, and D. Zibar, “Impact of transmitter phase noise on NFDm transmission with discrete spectral modulation”, *IEEE Photonics Technology Letters*, vol. 31, no. 22, pp. 1767–1770, 2019.

- [170] Y. Wang, R. Xin, S. Fu, M. Tang, and D. Liu, "Laser linewidth tolerance for nonlinear frequency division multiplexing transmission with discrete spectrum modulation", *Optics Express*, vol. 28, no. 7, pp. 9642–9652, 2020.
- [171] M. Kamalian, J. E. Prilepsky, S. T. Le, and S. K. Turitsyn, "On the design of NFT-based communication systems with lumped amplification", *Journal of Lightwave Technology*, vol. 35, no. 24, pp. 5464–5472, 2017.
- [172] M. Kamalian, J. E. Prilepsky, S. ThaiLe, and S. Turitsyn, "Optimal Launch and Detection Points for the NFT-based Communication System with Lumped Amplification", in *2017 European Conference on Optical Communication (ECOC)*, IEEE, 2017, pp. 1–3.
- [173] T. D. De Menezes, C. Tu, V. Besse, *et al.*, "Nonlinear Spectrum Modulation in the Anomalous Dispersion Regime Using Second- and Third-Order Solitons", in *Photonics*, MDPI, vol. 9, 2022, p. 748.
- [174] A. M. Turing, *Computing machinery and intelligence*. Springer, 2009.
- [175] R. Anyoha, *The history of artificial intelligence*, 2017. [Online]. Available: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- [176] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [177] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms", Cornell Aeronautical Lab Inc Buffalo NY, Tech. Rep., 1961.
- [178] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [179] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, 2014.
- [180] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: [1512.03385 \[cs.CV\]](https://arxiv.org/abs/1512.03385).
- [181] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning", in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.
- [182] N. Kitaev, L. Kaiser, and A. Levskaya, "Reformer: The efficient transformer", *arXiv preprint arXiv:2001.04451*, 2020.
- [183] A. N. Gomez, M. Ren, R. Urtasun, and R. B. Grosse, "The reversible residual network: Backpropagation without storing activations", *Advances in neural information processing systems*, vol. 30, 2017.
- [184] G. Litjens, T. Kooi, B. E. Bejnordi, *et al.*, "A survey on deep learning in medical image analysis", *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2017.07.005>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841517301135>.

- [185] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review", *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [186] Z. Wu, T. Yao, Y. Fu, and Y.-G. Jiang, "Deep learning for video classification and captioning", in *Frontiers of multimedia research*, 2017, pp. 3–29.
- [187] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various frameworks and libraries of machine learning and deep learning: A survey", *Archives of computational methods in engineering*, pp. 1–24, 2019.
- [188] D. Zibar, M. Piels, R. Jones, and C. G. Schäeffler, "Machine learning techniques in optical communication", *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442–1452, 2015.
- [189] F. Musumeci, C. Rottondi, A. Nag, *et al.*, "An overview on application of machine learning techniques in optical networks", *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2018.
- [190] F. N. Khan, Q. Fan, J. Lu, G. Zhou, C. Lu, and P. T. Lau, "Applications of machine learning in optical communications and networks", in *Optical Fiber Communication Conference*, Optical Society of America, 2020, M1G–5.
- [191] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications", *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, 2019.
- [192] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems", *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.
- [193] I. Macaluso, D. Finn, B. Ozgul, and L. A. DaSilva, "Complexity of spectrum activity and benefits of reinforcement learning for dynamic channel selection", *IEEE journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2237–2248, 2013.
- [194] Q. Zhuge, X. Zeng, H. Lun, *et al.*, "Application of machine learning in fiber non-linearity modeling and monitoring for elastic optical networks", *Journal of Lightwave Technology*, vol. 37, no. 13, pp. 3055–3063, 2019.
- [195] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer", *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.
- [196] C. Häger and H. D. Pfister, "Deep learning of the nonlinear Schrödinger equation in fiber-optic communications", in *2018 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2018, pp. 1590–1594.
- [197] V. Neskornik, A. Carnio, V. Bajaj, *et al.*, "End-to-end deep learning of long-haul coherent optical fiber communications via regular perturbation model", in *2021 European Conference on Optical Communication (ECOC)*, IEEE, 2021, pp. 1–4.
- [198] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems", *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 11, pp. 2503–2516, 2019.

- [199] B. Karanov, V. Oliari, M. Chagnon, *et al.*, “End-to-end learning in optical fiber communications: Experimental demonstration and future trends”, in *2020 European Conference on Optical Communications (ECOC)*, IEEE, 2020, pp. 1–4.
- [200] T. Uhlemann, S. Cammerer, A. Span, S. Dörner, and S. ten Brink, “Deep-learning autoencoder for coherent and nonlinear optical communication”, in *Photonic Networks; 21th ITG-Symposium*, VDE, 2020, pp. 1–8.
- [201] R. T. Jones, M. P. Yankov, and D. Zibar, “End-to-end learning for gmi optimized geometric constellation shape”, in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4.
- [202] O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, “End-to-end learning of a constellation shape robust to channel condition uncertainties”, *Journal of Lightwave Technology*, vol. 40, no. 10, pp. 3316–3324, 2022.
- [203] V. Oliari, B. Karanov, S. Goossens, *et al.*, “High-cardinality hybrid shaping for 4D modulation formats in optical communications optimized via end-to-end learning”, *arXiv preprint arXiv:2112.10471*, 2021.
- [204] V. Aref and M. Chagnon, “End-to-end learning of joint geometric and probabilistic constellation shaping”, in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2022, pp. 1–3.
- [205] O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, “Gradient-free training of autoencoders for non-differentiable communication channels”, *Journal of Lightwave Technology*, vol. 39, no. 20, pp. 6381–6391, 2021.
- [206] O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, “End-to-end learning of a constellation shape robust to variations in SNR and laser linewidth”, in *2021 European Conference on Optical Communication (ECOC)*, IEEE, 2021, pp. 1–4.
- [207] A. Rode, B. Geiger, and L. Schmalen, “Geometric constellation shaping for phase-noise channels using a differentiable blind phase search”, in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2022, pp. 1–3.
- [208] A. Rode, B. Geiger, S. Chimmalgi, and L. Schmalen, “End-to-end Optimization of Constellation Shaping for Wiener Phase Noise Channels with a Differentiable Blind Phase Search”, *Journal of Lightwave Technology*, 2023.
- [209] M. Kuschnerov, M. Schaedler, C. Bluemm, and S. Calabro, “Advances in deep learning for digital signal processing in coherent optical modems”, in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2020, pp. 1–3.
- [210] J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk, and S. J. Savory, “Machine learning for optical fiber communication systems: An introduction and overview”, *APL Photonics*, vol. 6, no. 12, 2021.
- [211] K. Mishina, S. Sato, Y. Yoshida, D. Hisano, and A. Maruta, “Eigenvalue-domain neural network demodulator for eigenvalue-modulated signal”, *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4307–4317, 2021.

- [212] R. T. Jones, S. Gaiarin, M. P. Yankov, and D. Zibar, “Time-domain neural network receiver for nonlinear frequency division multiplexed systems”, *IEEE Photonics Technology Letters*, vol. 30, no. 12, pp. 1079–1082, 2018.
- [213] S. Yamamoto, K. Mishina, and A. Maruta, “Demodulation of optical eigenvalue modulated signals using neural networks”, *IEICE Communications Express*, vol. 8, no. 12, pp. 507–512, 2019.
- [214] K. Mishina, S. Yamamoto, T. Kodama, Y. Yoshida, D. Hisano, and A. Maruta, “Experimental demonstration of neural network based demodulation for on-off encoded eigenvalue modulation”, in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4.
- [215] K. Mishina, S. Sato, S. Yamamoto, Y. Yoshida, D. Hisano, and A. Maruta, “Demodulation of eigenvalue modulated signal based on eigenvalue-domain neural network”, in *Optical Fiber Communication Conference*, Optica Publishing Group, 2020, W3D–1.
- [216] Y. Wu, L. Xi, X. Zhang, *et al.*, “Robust neural network receiver for multiple-eigenvalue modulated nonlinear frequency division multiplexing system”, *Optics Express*, vol. 28, no. 12, pp. 18 304–18 316, 2020.
- [217] O. Kotlyar, M. Pankratova, M. Kamalian-Kopae, A. Vasylychenkova, J. E. Prilepsy, and S. K. Turitsyn, “Combining nonlinear Fourier transform and neural network-based processing in optical communications”, *Optics letters*, vol. 45, no. 13, pp. 3462–3465, 2020.
- [218] P. Ge, L. Xi, J. Deng, J. Wei, S. Du, and X. Zhang, “Autoencoder assisted subcarrier optimization for nonlinear frequency division multiplexing”, *Optics Express*, vol. 31, no. 21, pp. 34 443–34 458, 2023.
- [219] P. J. Freire, Y. Osadchuk, B. Spinnler, *et al.*, “Performance Versus Complexity Study of Neural Network Equalizers in Coherent Optical Systems”, *Journal of Lightwave Technology*, vol. 39, no. 19, pp. 6085–6096, 2021.
- [220] B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, “End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks”, *Optics express*, vol. 27, no. 14, pp. 19 650–19 663, 2019.
- [221] V. Bajaj, S. Chimmalgi, V. Aref, and S. Wahls, “Exact NFDm Transmission in the Presence of Fiber-Loss”, *Journal of Lightwave Technology*, vol. 38, no. 11, pp. 3051–3058, 2020. DOI: [10.1109/JLT.2020.2984041](https://doi.org/10.1109/JLT.2020.2984041).
- [222] *Creative Commons Attribution 4.0 International (CC BY 4.0) License, WebPage*. <https://creativecommons.org/licenses/by/4.0/>, Accessed: 2024-07-31.
- [223] E. Bidaki and S. Kumar, “Nonlinear Fourier transform using multistage perturbation technique for fiber-optic systems”, *JOSA B*, vol. 35, no. 9, pp. 2286–2293, 2018.
- [224] J. Koch, S. Li, and S. Pachnicke, “Transmission of higher order solitons created by optical multiplexing”, *Journal of Lightwave Technology*, vol. 37, no. 3, pp. 933–941, 2018.

- [225] J. E. Prilepsky, S. A. Derevyanko, K. J. Blow, I. Gabitov, and S. K. Turitsyn, “Non-linear Inverse Synthesis and Eigenvalue Division Multiplexing in Optical Fiber Channels”, *Physical review letters*, vol. 113, no. 1, p. 013 901, 2014.
- [226] W. Forsyiaik, N. Doran, F. Knox, and K. Blow, “Average soliton dynamics in strongly perturbed systems”, *Optics communications*, vol. 117, no. 1-2, pp. 65–70, 1995.
- [227] V. Bajaj, S. Chimmalgi, V. Aref, and S. Wahls, “Exact nonlinear frequency division multiplexing in lossy fibers”, in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4.
- [228] V. A. Bogatyrev, M. M. Bubnov, E. M. Dianov, *et al.*, “A single-mode fiber with chromatic dispersion varying along the length”, *Journal of lightwave technology*, vol. 9, no. 5, pp. 561–566, 1991.
- [229] E. Dianov, P. Mamyshev, A. Prokhorov, and S. Chernikov, “Generation of a train of fundamental solitons at a high repetition rate in optical fibers”, *Optics letters*, vol. 14, no. 18, pp. 1008–1010, 1989.
- [230] W. Forsyiaik, F. Knox, and N. Doran, “Average soliton propagation in periodically amplified systems with stepwise dispersion-profiled fiber”, *Optics letters*, vol. 19, no. 3, pp. 174–176, 1994.
- [231] V. N. Serkin and T. L. Belyaeva, “Optimal control of optical soliton parameters: Part 1. the lax representation in the problem of soliton management”, *Quantum Electronics*, vol. 31, no. 11, p. 1007, 2001.
- [232] R. S. Burington, N. A. Lange, *et al.*, “Handbook of Mathematical Tables and Formulas”, 1940.
- [233] G. H. van Tartwijk, R.-J. Essiambre, and G. P. Agrawal, “Dispersion-tailored active-fiber solitons”, *Optics letters*, vol. 21, no. 24, pp. 1978–1980, 1996.
- [234] M. Brehler, C. Mahnke, S. Chimmalgi, and S. Wahls, “NFDMLab: Simulating Non-linear Frequency Division Multiplexing in Python”, in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2019, pp. 1–3.
- [235] S. Wahls, S. Chimmalgi, and P. J. Prins, “FNFT: A Software Library for Computing Nonlinear Fourier Transforms”, *Journal of Open Source Software*, vol. 3, no. 23, p. 597, 2018.
- [236] S. Chimmalgi and S. Wahls, “Theoretical analysis of maximum transmit power in a b-modulator”, in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4.
- [237] F. Buchali, K. Schuh, R. Dischler, *et al.*, “DCI Field Trial demonstrating 1.3-Tb/s Single-Channel and 50.8-Tb/s WDM Transmission Capacity”, *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 9, pp. 2710–2718, 2020.
- [238] F. Buchali, V. Aref, M. Chagnon, *et al.*, “1.52 Tb/s Single Carrier Transmission Supported by a 128 GSa/s SiGe DAC”, in *Optical Fiber Communication Conference*, Optical Society of America, 2020, Th4C–2.
- [239] *Ciena unveils WaveLogic 5: 800G and so much more*, <https://www.ciena.com/insights/articles/Ciena-unveils-WaveLogic-5-800G-and-so-much-more.html>. (visited on 02/20/2019).

- [240] D. Rafique, A. Napoli, S. Calabro, and B. Spinnler, "Digital Preemphasis in Optical Communication Systems: On the DAC Requirements for Terabit Transmission Applications", *Journal of Lightwave Technology*, vol. 32, no. 19, pp. 3247–3256, 2014.
- [241] J. Zhang, H.-C. Chien, Y. Xia, Y. Chen, and J. Xiao, "A novel adaptive digital pre-equalization scheme for bandwidth limited optical coherent system with dac for signal generation", in *OFC 2014*, IEEE, 2014, pp. 1–3.
- [242] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture", *IEEE transactions on signal processing*, vol. 45, no. 1, pp. 223–227, 1997.
- [243] Y. H. Lim, Y. S. Cho, I. W. Cha, and D. H. Youn, "An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems", *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1726–1730, 1998.
- [244] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory", *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, 2001.
- [245] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers", *IEEE Transactions on signal processing*, vol. 54, no. 10, pp. 3852–3860, 2006.
- [246] D. Zhou and V. E. DeBrunner, "Novel Adaptive Nonlinear Predistorters Based on the Direct Learning Algorithm", *IEEE transactions on signal processing*, vol. 55, no. 1, pp. 120–133, 2006.
- [247] H. Faig, Y. Yoffe, E. Wohlgemuth, and D. Sadot, "Dimensions-Reduced Volterra Digital Pre-Distortion Based On Orthogonal Basis for Band-Limited Nonlinear Opto-Electronic Components", *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–13, 2019.
- [248] R. Elschner, R. Emmerich, C. Schmidt-Langhorst, F. Frey, P. W. Berenguer, *et al.*, "Improving Achievable Information Rates of 64-GBd PDM-64QAM by Nonlinear Transmitter Predistortion", in *2018 Optical Fiber Communications Conference (OFC)*, 2018, pp. 1–3.
- [249] G. Khanna, B. Spinnler, S. Calabrò, E. De Man, and N. Hanik, "A Robust Adaptive Pre-Distortion Method for Optical Communication Transmitters", *IEEE Photonics Technology Letters*, vol. 28, no. 7, pp. 752–755, 2015.
- [250] Y. Yoffe, G. Khanna, E. Wohlgemuth, E. de Man, B. Spinnler, *et al.*, "Low-Resolution Digital Pre-Compensation Enabled by Digital Resolution Enhancer", *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1543–1551, 2019.
- [251] A. Napoli, P. W. Berenguer, T. Rahman, G. Khanna, M. M. Mezghanni, *et al.*, "Digital Pre-Compensation Techniques Enabling Cost-Effective High-Order Modulation Formats Transmission", *Opt. Comm.*, vol. 409, pp. 52–65, 2018.
- [252] D. Psaltis, A. Sideris, and A. A. Yamamura, "A multilayered neural network controller", *IEEE control systems magazine*, vol. 8, no. 2, pp. 17–21, 1988.

- [253] A. Bernardini, M. Carrarini, and S. De Fina, "The use of a Neural Net for Copeing with Nonlinear Distortions", in *1990 20th European Microwave Conference*, IEEE, vol. 2, 1990, pp. 1718–1723.
- [254] T. Gotthans, G. Baudoin, and A. Mbaye, "Digital predistortion with advance/delay neural network and comparison with Volterra derived models", in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, IEEE, 2014, pp. 811–815.
- [255] Y. Wu, U. Gustavsson, A. G. i Amat, and H. Wymeersch, "Residual Neural Networks for Digital Predistortion", 2020. arXiv: [2005.05655](https://arxiv.org/abs/2005.05655) [eess.SP].
- [256] C. Tarver, A. Balatsoukas-Stimming, and J. R. Cllaro, "Design and Implementation of a Neural Network Based Predistorter for Enhanced Mobile Broadband", in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, IEEE, 2019, pp. 296–301.
- [257] M. Abu-Romoh, S. Sygletos, I. D. Phillips, and W. Forysiak, "Neural-network-based pre-distortion method to compensate for low resolution DAC nonlinearity", in *45th European Conference on Optical Communication (ECOC 2019)*, IET, 2019, pp. 1–4.
- [258] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers", *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 146–148, 2019.
- [259] X. Hu, Z. Liu, X. Yu, *et al.*, "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers", *arXiv preprint arXiv:2005.09848*, 2020.
- [260] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral Modeling and Linearization of Wideband RF Power Amplifiers Using BiLSTM Networks for 5G Wireless Systems", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 348–10 356, 2019.
- [261] S. Boumaiza and F. Mkadem, "Wideband RF power amplifier predistortion using real-valued time-delay neural networks", in *2009 European Microwave Conference (EuMC)*, IEEE, 2009, pp. 1449–1452.
- [262] M. Schaedler, M. Kuschnerov, S. Calabrò, F. Pittalà, C. Bluemm, and S. Pachnicke, "AI-Based Digital Predistortion for IQ Mach-Zehnder Modulators", in *2019 Asia Communications and Photonics Conference (ACP)*, IEEE, 2019, pp. 1–3.
- [263] T. Sasai, M. Nakamura, E. Yamazaki, *et al.*, "Wiener-Hammerstein model and its learning for nonlinear digital pre-distortion of optical transmitters", *Opt. Express*, vol. 28, no. 21, pp. 30 952–30 963, 2020. DOI: [10.1364/OE.400605](https://doi.org/10.1364/OE.400605).
- [264] V. Bajaj, F. Buchali, M. Chagon, S. Wahls, and V. Aref, "Single-channel 1.61 Tb/s Optical Coherent Transmission Enabled by Neural Network-Based Digital Pre-Distortion", in *European Conference in Optical Communications (ECOC), 2020, Tu1D5*, 2020.

- [265] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, “54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion”, in *Optical Fiber Communication (OFC) conference*, 2021.
- [266] L. Schmalen, V. Aref, J. Cho, D. Suikat, D. Rösener, and A. Leven, “Spatially Coupled Soft-Decision Error Correction for Future Lightwave Systems”, *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1109–1116, 2014.
- [267] D. R. Morgan, Z. Ma, and L. Ding, “Reducing measurement noise effects in digital predistortion of RF power amplifiers”, in *IEEE International Conference on Communications, 2003. ICC’03.*, IEEE, vol. 4, 2003, pp. 2436–2439.
- [268] T. Ogunfunmi, *Adaptive Nonlinear System Identification: The Volterra and Wiener Model Approaches*. Springer Science & Business Media, 2007, ch. 1, ISBN: ISBN 978-0-387-26328-1.
- [269] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, and K. Keutzer, “A Survey of Quantization Methods for Efficient Neural Network Inference”, *CoRR*, vol. abs/2103.13630, 2021. arXiv: 2103 . 13630. [Online]. Available: <https://arxiv.org/abs/2103.13630>.
- [270] Z. Cai, X. He, J. Sun, and N. Vasconcelos, “Deep Learning with Low Precision by Half-Wave Gaussian Quantization”, *CoRR*, vol. abs/1702.00953, 2017. arXiv: 1702.00953.
- [271] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *arXiv e-prints*, arXiv:1412.6980, arXiv:1412.6980, Dec. 2014. arXiv: 1412.6980 [cs.LG].
- [272] H. Paaso and A. Mammela, “Comparison of direct learning and indirect learning predistortion architectures”, in *2008 IEEE International Symposium on Wireless Communication Systems*, 2008, pp. 309–313.
- [273] S. Zhalehpour, J. Lin, W. Shi, and L. A. Rusch, “Reduced-size lookup tables enabling higher-order QAM with all-silicon IQ modulators”, *Optics express*, vol. 27, no. 17, pp. 24 243–24 259, 2019.
- [274] J. H. Ke, Y. Gao, and J. C. Cartledge, “400 Gbit/s single-carrier and 1 Tbit/s three-carrier superchannel signals using dual polarization 16-QAM with look-up table correction and optical pulse shaping”, *Opt. Express*, vol. 22, no. 1, pp. 71–84, Jan. 2014. DOI: 10 . 1364 / OE . 22 . 000071. [Online]. Available: <http://www.osapublishing.org/oe/abstract.cfm?URI=oe-22-1-71>.
- [275] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, “A Comparative Analysis of the Complexity/Accuracy Tradeoff in Power Amplifier Behavioral Models”, *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 6, pp. 1510–1520, 2010.
- [276] M. Zhu and S. Gupta, *To prune, or not to prune: Exploring the efficacy of pruning for model compression*, 2017. arXiv: 1710.01878 [stat.ML].
- [277] F. Buchali, V. Aref, M. Chagnon, *et al.*, “52.1 Tb/s C-Band DCI Transmission over DCI Distances at 1.49 Tb/s/λ.”, in *2020 European Conference on Optical Communications (ECOC)*, IEEE, 2020, pp. 1–4.

- [278] F. Buchali, V. Aref, R. Dischler, *et al.*, “128 GSa/s SiGe DAC Implementation Enabling 1.52 Tb/s Single Carrier Transmission”, *Journal of Lightwave Technology*, vol. 39, no. 3, pp. 763–770, 2020.
- [279] V. Bajaj, V. Aref, and S. Wahls, “Performance analysis of recurrent neural network-based digital pre-distortion for optical coherent transmission”, in *2022 European Conference on Optical Communication (ECOC)*, IEEE, 2022, pp. 1–4.
- [280] V. Bajaj, M. Changon, S. Wahls, and V. Aref, “Efficient Training of Volterra Series-Based Pre-distortion Filter Using Neural Networks”, in *Optical Fiber Communication (OFC) conference*, 2022.
- [281] J. Song, Z. He, C. Häger, *et al.*, “Over-the-fiber Digital Predistortion Using Reinforcement Learning”, in *2021 European Conference on Optical Communication (ECOC)*, IEEE, 2021, pp. 1–4.
- [282] A. Paszke, S. Gross, S. Chintala, *et al.*, “Automatic differentiation in PyTorch”, in *NIPS-W*, 2017.
- [283] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [284] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the Properties of Neural Machine Translation: Encoder-Decoder Approaches”, *arXiv preprint arXiv:1409.1259*, 2014.
- [285] S. Deligiannidis, A. Bogris, C. Mesaritakis, and Y. Kopsinis, “Compensation of Fiber Nonlinearities in Digital Coherent Systems Leveraging Long Short-Term Memory Neural Networks”, *Journal of Lightwave Technology*, vol. 38, no. 21, pp. 5991–5999, 2020.
- [286] T. A. Eriksson, H. Bülow, and A. Leven, “Applying Neural Networks in Optical Communication Systems: Possible Pitfalls”, *IEEE Photonics Technology Letters*, vol. 29, no. 23, pp. 2091–2094, 2017.
- [287] V. Bajaj, R. Van de Plas, and S. Wahls, “Blind polarization demultiplexing of shaped qam signals assisted by temporal correlations”, *Journal of Lightwave Technology*, 2023.
- [288] V. Bajaj, R. Van de Plas, V. Aref, and S. Wahls, “Blind Polarization Demultiplexing of Probabilistically Shaped Signals”, in *2022 IEEE Photonics Conference (IPC)*, IEEE, 2022, pp. 1–2.
- [289] Q. Yan, C. Guo, and X. Hong, “The Impact of Probabilistic Constellation Shaping on Channel Equalization with Constant Modulus Algorithm”, in *Optoelectronics and Communications Conference*, Optical Society of America, 2021, T1B–3.
- [290] V. Lauinger, F. Buchali, and L. Schmalen, “Blind Equalization and Channel Estimation in Coherent Optical Communications Using Variational Autoencoders”, *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2529–2539, 2022.
- [291] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics”, *IEEE Transactions on signal processing*, vol. 45, no. 2, pp. 434–444, 1997.

- [292] M. J. Ready and R. P. Gooch, "Blind equalization based on radius directed adaptation", in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 1990, pp. 1699–1702.
- [293] F. P. Guiomar, S. B. Amado, A. Carena, *et al.*, "Fully Blind Linear and Nonlinear Equalization for 100G PM-64QAM Optical Systems", *J. Lightwave Technol.*, vol. 33, no. 7, pp. 1265–1274, Apr. 2015. [Online]. Available: <https://opg.optica.org/jlt/abstract.cfm?URI=jlt-33-7-1265>.
- [294] A. Galtarossa and C. R. Menyuk, *Polarization Mode Dispersion*. Springer, 2005, vol. 296.
- [295] C. Xie and L. Möller, "The accuracy assessment of different polarization mode dispersion models", *Optical Fiber Technology*, vol. 12, no. 2, pp. 101–109, 2006.
- [296] P. Serena, M. Bertolini, and A. Vannucci. "Optilux: An Open Source of Light". (2009), [Online]. Available: <http://www.optilux.sourceforge.net>.
- [297] K. I. Diamantaras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on frequency domain second-order statistics", *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 534–542, 2000.
- [298] H. Imtiaz, Z. Zheng, R. H. Nejad, L. A. Rusch, and M. Zeng, "Performance vs. complexity in NN pre-distortion for a nonlinear channel", *Optics Express*, vol. 31, no. 23, pp. 38 513–38 528, 2023.
- [299] V. Neskornik, F. Buchali, V. Bajaj, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, "Neural-Network-Based Nonlinearity Equalizer for 128 GBaud Coherent Transceivers", in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, IEEE, 2021, pp. 1–3.
- [300] Z. He, J. Song, K. Vijayan, *et al.*, "Periodicity-Enabled Size Reduction of Symbol Based Predistortion for High-Order QAM", *Journal of Lightwave Technology*, vol. 40, no. 18, pp. 6168–6178, 2022.

LIST OF PUBLICATIONS

7.3. JOURNAL ARTICLES

3. **V. Bajaj**, R. Van de Plas and S. Wahls, "Blind Polarization Demultiplexing of Shaped QAM Signals Assisted by Temporal Correlations," in *IEEE/OSA Journal of Lightwave Technology*, Vol. 42, No. 2, pp. 560-571, 2024.
2. **V. Bajaj**, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "Deep Neural Network-Based Digital Pre-distortion for High Baudrate Optical Coherent Transmission," in *IEEE/OSA Journal of Lightwave Technology*, Vol. 40, No. 3, pp. 597-606, Feb. 2022..
1. **V. Bajaj**, S. Chimmalgil, V. Aref and S. Wahls, "Exact NFDM Transmission in the Presence of Fiber-Loss," in *IEEE/OSA Journal of Lightwave Technology*, Vol. 38, No. 11, pp. 3051-3058, Jun. 2020.

7.4. CONFERENCE PROCEEDINGS

8. **V. Bajaj**, R. V. d. Plas, V. Aref and S. Wahls, "Blind Polarization Demultiplexing of Probabilistically Shaped Signals, in *Proc. IEEE Photonics Conference (IPC)*, Vancouver, Canada, Nov. 2022.
7. **V. Bajaj**, V. Aref and S. Wahls, "Performance Analysis of Recurrent Neural Network-based Digital Pre-Distortion for Optical Coherent Transmission," in *Proc. European Conference on Optical Communication (ECOC)*, Basel, Switzerland, Sep. 2022.
6. **V. Bajaj**, M. Chagnon, S. Wahls and V. Aref, "Efficient Training of Volterra Series-Based Pre-distortion Filter Using Neural Networks," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, USA, Apr. 2022.
5. V. Neskorniuk, A. Carnio, **V. Bajaj**, D. Marsella, S. K. Turitsyn, J. E. Prilepsky, V. Aref, "End-to-End Deep Learning of Long-Haul Coherent Optical Fiber Communications via Regular Perturbation Model," in *Proc. European Conference on Optical Communication (ECOC)*, Bordeaux, France, Sep. 2021.
4. **V. Bajaj**, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "54.5 Tb/s WDM Transmission over Field Deployed Fiber Enabled by Neural Network-Based Digital Pre-Distortion," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, San Francisco, USA, Jun. 2021.
3. V. Neskorniuk, F. Buchali, **V. Bajaj**, S. K. Turitsyn, J. E. Prilepsky and V. Aref, "Neural-Network-Based Nonlinearity Equalizer for 128 GBaud Coherent Transceivers," in *Proc. Optical Fiber Communication Conference (OFC)*, San Francisco, USA, Jun. 2021
2. **V. Bajaj**, F. Buchali, M. Chagnon, S. Wahls and V. Aref, "Single-channel 1.6 Tb/s Optical Coherent Transmission Enabled by Neural Network-Based Digital Predistortion," in *Proc. European Conference on Optical Communication (ECOC)*, Brussels, Belgium, Dec. 2020.

1. **V. Bajaj**, S. Chimmalgi, V. Aref and S. Wahls, "Exact Nonlinear Frequency Division Multiplexing in Lossy Fibers," in *Proc. European Conference on Optical Communication (ECOC)*, Dublin, Ireland, Sep. 2019.

CURRICULUM VITÆ

Vinod BAJAJ

09-12-1992 Born in Jodhpur, India.

EDUCATION

2010–2014 Bachelor of Technology in Electronics & Communication Engineering
Engineering College, Ajmer
Ajmer, India

2014–2016 Master of Technology in Electrical Engineering
Indian Institute of Technology Madras
Chennai, India

WORK EXPERIENCE

2016–2018 Associate Manager
Center of Excellence, Sterlite Technologies
Aurangabad, India

2023–today Research Engineer
Optical System and Device Research, Nokia Bell Labs
Paris, France

HONORS AND AWARDS

- | | |
|------|------------------------------------------------------------------------------------------------------------------------------------------------|
| 2022 | Invited talk at the European Conference on Optical Communication (ECOC) |
| 2021 | Finalist for 2021 Corning Outstanding Student Paper Competition Grand Prize at the Optical Fiber Communication Conference and Exhibition (OFC) |
| 2020 | Bell Labs Summer Intern Award for Outstanding Innovation at Nokia Bell-Labs |
| 2018 | Marie Skłodowska-Curie Actions fellowship for three years of PhD |
| 2017 | Star Performer award at Sterlite Technologies, India |
| 2013 | First prize in Robotics and Circuit Designing Competition in technical festival Tarangini 2013 at Government Engineering College Ajmer, India |