Shaping Strategies to Embrace Nonlinear Effects in Optical Fiber Communications

Jingtian Liu

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2024IPPAT007

Thèse de doctorat

Shaping Strategies to Embrace Nonlinear Effects in Optical Fiber Communications

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Électronique et Optoélectronique

Thèse présentée et soutenue à Palaiseau, le 29/02/2024, par

JINGTIAN LIU

Composition du Jury :

Mme Ghaya REKAYA Full Professor, Télécom Paris	Présidente/Examinatrice
Alex ALVARADO Full Professor, Eindhoven University of Technology	Rapporteur
Marco SECONDINI Full Professor, Sant'Anna School of Advanced Studies	Rapporteur
Gabriella BOSCO Full Professor, Politecnico di Torino	Examinatrice
Christophe PEUCHERET Full professor, Université de Rennes Enssat	Examinateur
Yves JAOUËN Full Professor, Télécom Paris	Directeur de thèse
Élie AWWAD Assistant Professor, Télécom Paris	Co-directeur de thèse
Hartmut HAFERMANN Research Engineer, Huawei Technologies	Invité

"Die Philosophen haben die Welt nur verschieden interpretiert; es kommt aber darauf an, sie zu verändern."
"The philosophers have only interpreted the world, in various ways. The point, however, is to change it."

— Karl Marx, Eleven Theses on Feuerbach

Remerciements

I would like to extend my most profound appreciation to Professors Élie AWWAD and Yves JAOUËN, whose invaluable guidance and steadfast support have been pivotal throughout my academic journey. As mentors, they illuminated the path into the mesmerizing world of fiber optic communications. Professor Élie AWWAD, in particular, was instrumental from my master's internship onwards, demystifying the complexities of this domain and igniting my passion for future research endeavors.

In the subsequent doctoral phase, Professors Élie AWWAD and Yves JAOUËN consistently demonstrated remarkable seriousness, responsibility, and enthusiasm. Their patience and accessibility were key in surmounting numerous challenges. Even in the face of setbacks, their unwavering belief in my potential and their ceaseless encouragement were invaluable. I am profoundly indebted for their sage advice and motivation, which have been crucial to my growth and the successful culmination of my doctoral studies. Indeed, standing on the shoulders of such giants, I have gained a deeper understanding of this adage.

My gratitude extends beyond my immediate academic environment. I am deeply thankful to researcher Djalal Falih BENDIMERAD of Infinera, researcher Hartmut HAFERMANN, and Professor Yann FRIGNAC at Huawei Technology. Our collaborative discussions and partnerships during my doctorate played an instrumental role in shaping my research. I am also deeply indebted to the members of my supervisory committee, especially Professor Alex ALVARADO and Professor Marco SECONDIN. Their papers and lectures have been a fountain of inspiration and learning. I would also like to express my profound gratitude to Professor Ghaya REKAYA and Professor Luiz ANETNETO for their invaluable guidance and encouragement at a pivotal stage in my PhD journey.

Special thanks are reserved for Professor CHEN Bin, whose dedication to scientific inquiry has profoundly influenced my research approach.

I am also immensely grateful to my peers and friends - ZHAO Shiyuan, DING Shihao, and WANG Yibo - for their friendship and stimulating discussions. Their collaboration has greatly enriched my journey and expanded my perspectives on scientific re-

search. Similarly, my classmates from Telecom Paris - ZHANG Yarui, WANG Xiaolin, LIU Yi, YAO Yuan, WANG Yinghao, and CUI Di- have broadened my academic horizons with their diverse and interdisciplinary insights.

I wish to extend my heartfelt gratitude to my dear friends, particularly WANG Xiaoyuan, GUO Qianyu, LIANG Yangzhe, and GAO Yan, and also my after-school mentors - my guitar teacher Ge ZI, and my powerlifting instructor YAN Ning, whose companionship and support have been invaluable during my doctoral journey.

Lastly, but most importantly, I express my deepest gratitude to my parents and my grandparents. Their unwavering love, steadfast belief in my capabilities, enduring patience during trying times, and exuberant celebrations in moments of success have been the bedrock of my strength.

As I conclude this dissertation, I recognize that this is not the end, but the beginning of a new chapter in my academic pursuit. The wisdom, knowledge, and support I have received as I step into the next phase of my career have significantly shaped the work presented here—my heartfelt thanks to all who have been part of this incredible journey.

Abstract

In the 21st century, the demand for data communications has surged, thrusting the advancement of faster and more efficient optical fiber transmissions into the spotlight. The main impediment in long-distance communications is nonlinear interference (NLI), stemming from nonlinear effects in optical fibers. While Digital signal processing algorithms offer partial mitigation, the inherent nonlinear nature of the fiber, coupled with predominant dispersion effects, continues to challenge the increase of transmission throughputs. Addressing nonlinearity at the information source through signal modulation technology is at the heart of our research. Traditional modulation schemes, as spectral efficiency climbs, such as Quadrature Amplitude Modulation, become increasingly susceptible to nonlinear interference while their Mean Squared Euclidean Distance (MSED) diminishes. While multi-dimensional (MD) modulation yields improved linear and partial nonlinear gains, it has not yet demonstrated tangible benefits. On the other hand, the emergence of probabilistic constellation shaping (PCS), preferred for its enhanced linear gain and compatibility with conventional modulation hardware and software, introduces additional nonlinear interference. Consequently, the design of nonlinear-tolerant probabilistic shaping is emerging as a pivotal research direction. Our thesis begins with a novel MD modulation for uniformly distributed signals. Then, we propose a novel approach combining MD with PCS to examine performance variations. Delving into PCS, we investigate the enumerative sphere shaping distribution matcher (DM), initially from an MD stance, and design a distribution matcher optimized for nonlinear tolerance over shorter distances. Subsequently, we introduce a new NLI measurement technique, accounting for dispersion effects. Integrating this with the sequence selection framework of PCS, we achieve successful long-distance transmission with notable nonlinear gains.

Keywords: Optical Fiber Communications, Polarization Division Multiplexing, Nonlinear Kerr effects, Probabilistic constellation shaping, Multi-dimensional modulations, Sequence selection.

Résumé

Le principal obstacle aux communications longue distance par fibre optique est l'interférence non-linéaire (NLI), résultant des effets non linéaires des fibres optiques. Bien que les algorithmes de traitement numérique des signaux offrent une réduction partielle de la NLI, la nature non-linéaire inhérente à la fibre, associée à des effets de dispersion prédominants, continue à entraver l'accroissement des débits de transmission. Le traitement de ce problème à la source par la conception de nouvelles modulations intrinsèquement moins sensibles aux distorsions non-linéaires est au cœur de notre recherche. À mesure que l'efficacité spectrale augmente, les schémas de modulation traditionnels, tels que les formats QAM, deviennent de plus en plus sensibles à la NLI alors que leur distance Euclidienne minimale diminue. Bien que les modulations multidimensionnelles (MD) permettent d'améliorer les gains linéaires et d'apporter des gains non linéaires partiels, elles n'ont pas encore démontré d'avantages tangibles. D'autre part, l'émergence de la mise en forme de constellation probabiliste (PCS pour Probabilistic Constellation Shaping), préférée pour son gain linéaire accru et sa compatibilité avec la chaîne de traitement numérique utilisée pour les formats de modulation conventionnels, introduit une NLI supplémentaire. Par conséquent, la conception de schémas PCS tolérants à la non-linéarité apparaît comme un axe de recherche crucial.

Notre thèse commence par la proposition d'une nouvelle modulation MD pour les signaux uniformément distribués. Nous avons conçu une nouvelle constellation 4D, appelée 4D-2A-RS64, avec une entropie de 6 bits/4D, qui a montré une meilleure performance linéaire et une plus grande tolérance à la non-linéarité de la fibre que les formats de modulation conventionnels PDM-8QAM-star ou un format 4D récent, 4D-2A-8PSK, optimisé pour le canal non-linéaire. Nous avons démontré par des simulations numériques que les distances de transmission maximales sont augmentées pour les deux formats pour différents taux de symboles. Bien que la modulation proposée n'ait pas battu les performances d'un autre concurrent 4D, le format 4D-64PRS, nous avons démontré que le format 4D-2A-RS64 impose des contraintes moins strictes à l'émetteur que les autres formats 4D. Par conséquent, sa génération à l'aide de composants non idéaux (déséquilibres extrêmes dans les MZM et les DAC à résolution limitée) présente une dégradation des performances moins importante que les formats 4D de pointe tels que 4D-2A-8PSK et 4D-64PRS dans les régimes de transmission linéaires et non linéaires. Dans la conception de futurs schémas de modulation d'ordre

élevé avec une efficacité spectrale plus élevée, nous pouvons encore optimiser les performances linéaires en augmentant le MSED et en adaptant la distribution de toutes les distances euclidiennes au carré. Pour la partie non linéaire, il est crucial de maintenir les fluctuations d'énergie temporelle aussi faibles que possible. En outre, une efficacité spectrale plus élevée nécessite souvent plus de symboles et une cartographie plus précise, ce qui signifie que la sensibilité au déséquilibre MZM et aux variations DAC augmentera. L'extension de la solution proposée pour atteindre une efficacité spectrale encore plus élevée en utilisant plus d'anneaux reste un problème ouvert à explorer à l'avenir. En outre, la conception de schémas de signalisation optimisés pour la transmission sur une fibre non linéaire utilisant à la fois des modulations multidimensionnelles et une mise en forme probabiliste pourrait permettre d'obtenir des gains linéaires et non linéaires supplémentaires, comme cela a été suggéré dans [1, 2].

Dans la deuxième partie de la thèse, nous proposons une nouvelle approche combinant les modulations MD et la PCS pour examiner les variations de performance. En ce qui concerne les schémas PCS, nous étudions différentes solutions de transformation de la distribution des signaux (DM pour Distribution Matcher), en tenant compte des dimensions multiples, et concevons un DM optimisé pour la tolérance aux effets non-linéaires sur les liaisons de courte distance. Nous avons présenté deux nouvelles conceptions de PCS. La première, PCS-4D set-partitioned QAM, exploite le PCS dans des modulations multidimensionnelles, illustrant une légère amélioration des performances dans les régimes linéaires et non linéaires. Cette approche a été illustrée en utilisant la modulation PCS-4D 88SP-16QAM, qui montre des améliorations de performance grâce à l'utilisation de moins de niveaux d'énergie élevés. Toutefois, ces avantages se manifestent principalement à des taux d'erreur plus faibles et nécessitent des techniques DSP plus sophistiquées. La deuxième conception, l'ESS (Enumerative Sphere Shaping) à bande d'énergie limitée 4D, opère sur quatre dimensions (I et Q de deux états de polarisaiton orthogonaux) et vise à gérer les fluctuations d'énergie pour stimuler les gains de performance linéaires ou non linéaires. Ce nouveau schéma surpasse les méthodes ESS existantes, en particulier pour les liaisons à courte distance. La limitation des gains de performance aux courtes distances est attribuée à la complexité croissante de l'équilibrage des gains linéaires et non linéaires sur de plus longues distances, un défi aggravé par l'accumulation de la dispersion chromatique qui intensifie les variations d'énergie dans les signaux optiques modulés à haut débit. Notre analyse a également révélé que les métriques 4D sont plus efficaces pour prédire les performances non linéaires des systèmes ESS que les métriques 1D et 2D. Cependant, les métriques 4D actuelles manquent encore de précision, ce qui souligne

la nécessité de développer des métriques 4D plus précises comme une direction clé pour la recherche future et la réalisation de gains non linéaires pour la transmission à tronçons multiples.

Dans la troisième partie de la thèse, nous définissons une nouvelle technique de mesure de la NLI, qui tient compte des effets de dispersion accumulés. En intégrant cette technique au cadre de sélection des séquences de symboles PCS, nous parvenons à une transmission longue distance réussie avec des gains non-linéaires notables. Nous avons proposé une nouvelle mesure dépendante du signe : D-EDI (Dispersive Energy Dispersion Index), qui peut servir de mesure de prédiction pour la génération de NLI car elle a montré une corrélation négative avec les performances mesurées des systèmes de transmission de longue distance. En appliquant le D-EDI (resp. l'EDI) à la sélection des séquences (SS), nous avons défini le schéma D-SS (resp. E-SS). Ces deux schémas ont permis d'obtenir des gains plus importants avec l'ESS utilisé pour l'adaptation de la distribution que les schémas ESS de pointe sur de courtes distances de transmission. Notamment, le schéma D-SS a montré des gains même pour des distances de transmission plus longues et des débits plus élevés sur des systèmes à canal unique et WDM, même lorsqu'une correction de la phase absolue est utilisé pour réduire l'impact des rotations de phase non linéaires, atteignant la même performance que la sélection de séquences basée sur une simulation SSFM complète [3]. En outre, le D-SS a conservé une performance robuste pour une complexité modérée (nombre inférieur d'opérations de dispersion numérique pour calculer la métrique) et a surpassé l'ESS conventionnel pour différentes longueurs de blocs.

Ces résultats ouvrent la voie à de nouvelles optimisations des schémas de mise en forme pour les transmissions longue distance et confirment la nécessité d'utiliser des mesures dépendantes du signe et du canal pour évaluer efficacement les distorsions non linéaires. Néanmoins, ces résultats nous indiquent qu'il y a encore de la place pour l'amélioration. Pour les transmissions basées sur la DMB, dans lesquelles le taux de transmission par sous-porteuse est inférieur à celui des transmissions à porteuse unique, l'impact de la mémoire du canal est limité à quelques symboles par tronçon de fibre, voire à moins d'un symbole par tronçon. Ainsi, l'impact de la mémoire du canal peut être pris en compte en concaténant plusieurs DM pour minimiser l'influence mutuelle des séquences adjacentes. Au fur et à mesure que les taux de transmission augmentent, tant pour les transmissions multiporteuses que pour les transmissions monoporteuses, la mémoire du canal peut s'étendre sur des centaines ou des milliers de symboles. Parallèlement, l'augmentation de la longueur des séquences candidates accroît également la complexité du système. Par conséquent, la

question des modifications et des optimisations de D-EDI en vue d'obtenir d'excellentes prévisions de performances pour les transmissions longue distance à haut débit, tout en maintenant une complexité réduite, reste ouverte. D'autres améliorations de la métrique pourraient concerner différents schémas de gestion de la dispersion ou prendre en compte les effets non linéaires inter-canaux. Dans ce travail, nous n'avons considéré que les systèmes longue distance sans gestion de la dispersion où les contributions NLI des différents tronçons peuvent être considérées comme indépendantes et où le D-EDI donne une bonne prédiction de performance en additionnant les EDI de la séquence au début de chaque portée. Enfin, un objectif plus large consiste à effectuer une analyse approfondie des "bonnes" séquences sélectionnées. En comprenant bien leurs propriétés, nous visons à concevoir des solutions de génération de séquences plus simples.

Mots-clés : Communications sur Fibres Optiques, Multiplexage en Polarisation, Non-linéaire Effets Kerr, Mise en forme de constellation probabiliste, Modulations multi-dimensionnelles, Sélection de séquence.

Contents

Glossary

4D Four-Dimension

AIR Achievable Information Rates

B-ESS Band-Trellis ESS

BER Bit Error Rate

BL-ESS Band-Limited Enumerative Sphere Shaping

BPSK Binary-Phase-Shift-Keying

BS-SS Bit Scrambling Sequence Selection

CAZAC Constant-Amplitude Zero-Autocorrelation

CCDM Constant Composition Distribution Matcher

CD Chromatic Dispersion

CFO Carrier Frequency Offset

CPE Carrier Phase Estimation

CUT Channel Under Test

D-EDI EDI of Dispersed Sequences

DGD Differential Group Delay

DM Distribution Matcher

DM systems Dispersion-managed systems

DMB Digital Multi-Bands

DSCM Digital Sub-Carrier Multiplexing

DSP Digital Signal Processing

DUM systems Dispersion-unmanaged Systems

EDFA Erbium-Doped Fiber Amplifier

EDI Energy Dispersion Index

EEPN Equalization Enhanced Phase Noise

ENOB Effective Number of Bits

ESS Enumerative Sphere Shaping

FEC Forward Error Correct-ion/-ing

FWM Four Wave Mixing

GA Genie aided

Gb 1 Gigabit = 10^9 bits

GS Geometric Shaping

iid Independent Identically Distributed

ISI Inter-symbol Interference

K-ESS Kurtosis-Limited ESS

LDPC Low-Density Parity-Check

LSAS Lowpass-Filtered Symbol-Amplitude Sequence

LUT Look-up Table

MB Maxwell-Boltzmann

MCI Multi-Channel Interference

MIMO Multiple-Input Multiple-Output

MMSE Minimum-Mean-Square-Error

MPDM Multi-set partition distribution matching

MZM Mach-Zehnder Modulator

NLI Nonlinear Interference

NLPN Nonlinear Phase Noise

PAPR Peak-to-Average Power Ratio

Pb 1 Petabit = 10^6 Gigabits

PCS Probabilistic Constellation Shaping

PDC Pre-Dispersion Compensation

PDM/Pol-Mux Polarization Division Multiplexing

PMD Polarization Mode Dispersion

PRBS Pseudo-Random Binary Sequence

RMS Root Mean Square

RRC Root-Raised-Cosine

RX Receiver

SC Single Carrier

SCI Self-Channel Interference

SE Spectral Efficiency

SI-SS Symbol Interleaving Sequence Selection

SOP State-of-Polarization

SP Set Partitioned

SPM Self Phase Modulation

SSFM Split-Step Fourier Method

SSMF Standard Single Mode Fiber

Tb 1 Terabit = 10^3 Gigabits

TX Transmitter

WDM Wavelength Division Multiplexing

XCI Cross-Channel Interference

XPM Cross Phase Modulation

Notations & Symbols

Notation	Definition	unit
x	scalar	
$\mathbf{x}$	vector	
$\mathbf{X}$	matrix	
$\|\cdot\|_F$	Frobenius norm	
$\lfloor\cdot\rfloor$	floor function	
$\lceil\cdot\rceil$	ceiling function	
$(\cdot)^T$	transpose	
$(\cdot)^*$	conjugate	
$(\cdot)^\dagger$	transpose conjugate	
$\mathbb{E}[\cdot]$	Expectation	
j	imaginary unit, $j^2 = -1$	
$\Im(\cdot)$	Imaginary part of a complex number	
$p\{\cdot\}$	Probability	
$p\{\cdot\|\cdot\}$	Conditional probability	
$\Re(\cdot)$	Real part of a complex number	
α	Fiber loss coefficient	km^{-1}
α_{dB}	Fiber loss coefficient	dB/km
A_{eff}	Fiber effective area	m^2
β_1	Fiber group velocity inverse	ps/km
β_2	Fiber group velocity dispersion	ps^2/km
B_{ref}	Reference spectral bandwidth	nm
c	light velocity in vacuum	m/s
C	Channel capacity	bit/s/Hz
$\Delta\tau$	Differential group delay	ps
d_{min}	minimal distance of a code or a constellation	
D	Dispersion coefficient	ps/nm/km
E_b	Energy per information bit	J
E_s	Energy per modulated symbol	J
$\hbar$	Planck constant	eV.s
γ	Non-linearity coefficient	$km^{-1}W^{-1}$
κ_{dB}	Polarization dependent loss	dB
I, Q	In-phase and quadrature field components	
k_c	Modulation spectral efficiency	
λ_c	Laser wavelength	m
L	Fiber length	m
L_{eff}	Effective length of the fiber	m
q	number of states in a constellation	
M	number of channels in an SDM system	
$\mathbf{n}(\omega)$	Fiber refractive index	
n_{sp}	Spontaneous emission factor	
N_{ASE}	ASE spectral density	W/Hz
$N_0/2$	Noise spectral density per real dimension	W/Hz

ω	Pulsation (angular frequency)	rad/s
Φ	Phase	rad
R	Photodiode responsivity	A/W
R_c	FEC rate	
R_b	Bit rate	bit/s
R_s	Symbol rate	Baud
t	time	s
T_s	Symbol duration	s

List of Figures

List of Tables

Introduction

In the 21st century, the need to transmit, daily, vast quantities of information between cities and countries has made communication essential. Low-latency, minimal-blocking, high-speed communication systems are paramount. Spanning from access networks and metropolitan area networks to long-distance terrestrial and transoceanic connections, optical fiber stands as the unrivaled conduit for global data traffic. Its prominence is largely due to its capacity for low-attenuation and high-bandwidth transmission of modulated light over considerable distances. Consequently, leveraging fiber optics for communications has become increasingly vital.

The integration of wavelength division multiplexing (WDM), coherent reception, high-order modulation formats, and polarization multiplexing (PDM) has significantly accelerated the capacity of optical fiber transmission networks over the past decade. These advances have escalated the bit rate per wavelength from 2.5 Gb/s to 10 Gb/s and now beyond 400 Gb/s. According to Shannon's theorem, the theoretical capacity of a linear transmission system is constrained by a limit, called the Shannon limit. As communication system capacities expand, signal power must also increase to satisfy signal-to-noise ratio demands. However, given that the optical fiber is a nonlinear channel, i.e. the optical refractive index depends not only on the frequency of the incoming signal but also on the intensity of the light (Kerr effect), the maximum capacity of an optical fiber transmission network is ultimately limited by these nonlinear effects (as well as other nonlinear effects such as Raman scattering when considering ultra wideband transmissions).

Current signaling schemes predominantly utilized in fiber optic communications often adopt standard design principles suited for additive white Gaussian noise (AWGN) channels, even when applied to nonlinear fiber optic channels. Consequently, the achievable information rate peaks at a certain optical power threshold and then diminishes with any further increase. This trend starkly contrasts with linear channels, where capacity typically grows logarithmically with the signal-to-noise ratio (SNR). This work endeavors to devise practical and efficient signaling schemes that more effectively address the nonlinear effects encountered in long-distance coherent optical fiber transmission links. Promising approaches include multidimensional modulation and probabilistic constellation shaping (PCS), both of which offer substantial benefits in the linear regime over traditional two-dimensional constellations. The synergy between these methodologies warrants deeper

investigation. Furthermore, within the framework of PCS, our primary goal is to finely adjust the signal through sequence selection to curtail the generation of Kerr-induced Nonlinear Interference (NLI). Our sequence selection process necessitates careful consideration of the waveform transformations caused by dispersion, and the nonlinearity that light waves experience throughout transmission.

Goals and contributions of the thesis

In this thesis, we start with the development of a multidimensional modulation design and then propose three distinct designs based on probabilistic shaping techniques.

Our research journey commenced with the establishment of a MATLAB-based simulation framework, utilizing the Optilux open source code library. As our investigation deepened, we progressively customized and expanded our simulation system by replacing and adding code blocks to better fit our specific research needs.

The initial phase of our study concentrated on multidimensional modulation, where we crafted a novel four-dimensional modulation format tailored for long-distance communication. Building upon this foundation, we then introduced a method that combines multidimensional modulation with probabilistic shaping. Then, we included a strategy for mitigating NLI in PCS by constraining energy fluctuations within multiple dimensions.

Finally, we devised an innovative NLI measurement technique specifically for long-distance communication applications. This method was integrated into a sequence selection strategy for PCS modulations, enhancing the overall performance and efficiency of our proposed communication system.

Thesis Outline

Accordingly, the thesis is organized as follows: in Chapter 1, we briefly review the evolution of optical transmission systems and introduce our simulation system.

Chapter 2 is dedicated to the basic knowledge of signaling technologies, from conventional constellation shaping to probabilistic constellation shaping.

In Chapter 3, we focus on the evolution of multidimensional modulation and introduce our first multidimensional shaping design, called the 4D-2A-RS64 format.

Chapter 4 is dedicated to the evolution of nonlinear tolerant PCS and to our designs for PCS schemes: first, a combination of multi-dimensional modulation and PCS, named PCS-4D-128SP-16QAM, and second, a four-dimensional energy-limited PCS, named 4D-BL-ESS.

Chapter 5 is dedicated to the evolution of sequence selection for enhanced nonlinear transmission performance. In this chapter, we introduce our new sign-dependent performance metric

named D-EDI (Energy dispersion index of dispersed sequences), as well as the application of this metric to sequence selection.

Chapter 1

Evolution of Optical Fiber Transmission Systems

1.1 Introduction

In the 21st century, the relentless exchange of vast quantities of data between cities and countries highlights the critical importance of telecommunications. Systems that provide instantaneous connectivity, seamless data flow, and rapid transmission are not just desired but essential. Consequently, the significance of optical fiber communications in meeting these demands has become increasingly pronounced.

This chapter delves into the evolution of optical communication, spotlighting the advancements and methodologies that have shaped its trajectory. A significant emphasis is placed on the entire coherent simulation transmission system, marking the transition from early optical technologies to the sophisticated, high-capacity systems of today. In section 1.2, we emphasize on the evolution of the transmission system. Subsequently, in section 1.3, we introduce our simulation framework, detailing the methodologies and tools used to model and analyze optical fiber systems. The chapter then dissects the three main components of the transmission link. We start with the transmitter side in section 1.4, where the signal's genesis and initial shaping occur. We then progress to the channel model in section 1.5, where the signal's journey through the optical medium and the various effects encountered are explored. We conclude at the receiver side with a detailed examination of Digital Signal Processing (DSP) tools in section 1.6, highlighting the techniques used to recover the transmitted data accurately, and then we introduce the knowledge of Forward Error Correction Codes (FEC) in section 1.7. Finally, we present the concept of Achievable Information Rate (AIR) in section 1.8, discussing its significance as a performance metric and methods for estimating AIR in the context of optical communications.

1.2 Development of Fiber Communications

Optical fiber communication technology, crucial in the current Internet era, is one of the foundational technologies of the information age. Its basic elements include light sources (predominantly lasers), optical fibers with exceptional energy transmission efficiency, and photodetectors (PDs) at the receiving end. In today's world, all types of information technologies depend on communication networks for data transmission. Optical fiber communication technology connects various networks, forming an arterial system in information transmission and playing a pivotal role. As illustrated in Fig. 1.1, the modern communication network architecture comprises wireless access networks, metro networks, and core networks, among others, interconnected by optical fiber technology. For instance, in wireless access networks, optical fibers link wireless base stations to metro networks, with wireless communication only employed for the final connection between base stations and users. The metro network connects the data transmission between different cities, countries, and even continents. In data center networks, optical interconnections are prevalent, utilizing optical fibers for data transfer within and between centers. Thus, optical fiber communication technology not only serves as the main thoroughfare in current communication networks but also as many vital byways.

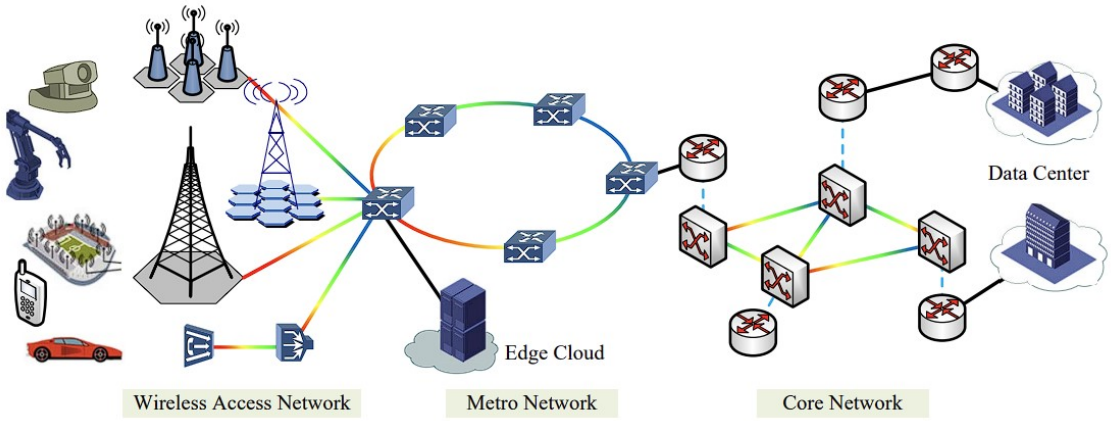


Figure 1.1: Architecture of multi-domain heterogeneous converged optical networks. Source: [4].

The optical fiber transmission network, comprised of optical fibers, lasers, modulators and many other optical and optoelectronic devices, is the fundamental bearer network for other service networks. With the emergence of technologies like the Internet of Things, big data, virtual reality, artificial intelligence (AI), and fifth-generation mobile communications (5G), the demands for information exchange and transmission have escalated. According to Cisco's 2020 research, the global number of Internet users is projected to increase from 3.9 billion in 2018 to 5.3 billion by 2023, growing at a compound annual growth rate of 6 percent. This growth signifies an increase in internet penetration from 51 percent of the global population in 2018 to an anticipated 66 percent

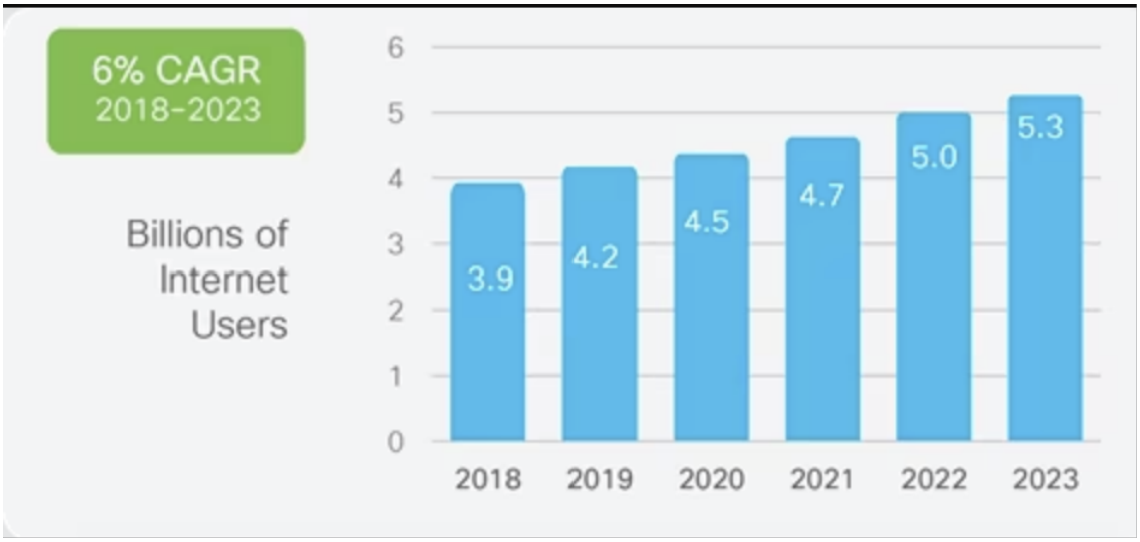


Figure 1.2: Global Internet user growth. Source: [5].

by 2023. These statistics and projections are visualized in Fig. 1.2. Facing this surge in users, optical fiber communication, as a backbone of the communication network, faces immense pressure to evolve. High-speed, large-capacity optical fiber communication systems, and networks are poised to be the primary direction of future development in this field.

The origins of fiber optics trace back to 1887 when Charles Vernon Boys crafted the first optical fiber in his laboratory. Yet, at that time, effective control of light remained a formidable challenge. A significant leap forward occurred in 1960 with Theodore Maiman's demonstration of the first laser, laying the groundwork for optical communications' future advancements. In 1966, Sir Charles Kao unveiled the potential of optical fiber communications [40], forming the foundation of contemporary communications systems.

Research into fiber optic communication systems began in earnest around 1975. Initially, GaAs lasers emitting at 800 nm were utilized. However, by 1970, it was discovered that single-mode fiber (SMF) in the 1300 nm range significantly reduced loss and dispersion. The 1550 nm region was later identified for its minimal losses, although initially limited to 60-kilometer repeater distances. This limitation was overcome in 1992 with the advent of optical amplifiers and Wavelength Division Multiplexing (WDM), which substantially enhanced throughput and extended repeater distances. The current focus is on broadening the bandwidth to include the S+C+L bands and improving the spectral efficiency of WDM systems.

Today, the telecommunications sector boasts a variety of optical fibers, each tailored for specific applications, with fiber losses reduced to below 0.2 dB/km and with different chromatic dispersion properties as shown in Fig. 1.3, such as "Standard" SMF (ITU-T G.652) also known as SMF-28, Low Water Peak Fiber (ITU-T G.652.C/D), Dispersion Shifted Fiber (ITU-T G.653), Low-Loss Fiber (ITU-T G.654), Non-Zero Dispersion Shifted Fiber (ITU-T G.655), and Bend

Insensitive Fiber (ITU-T G.657). Thanks to advanced transmitters, amplifiers, and receivers, communications can now span over several thousand kilometers. Optical fibers theoretically support billions of channels, and their cost-effectiveness further cements their role in modern telecommunications. Optical fiber communication technology, a pivotal application of laser information technology, has revolutionized both science and society, constructing the framework of modern communication networks and becoming a core component of information transmission.

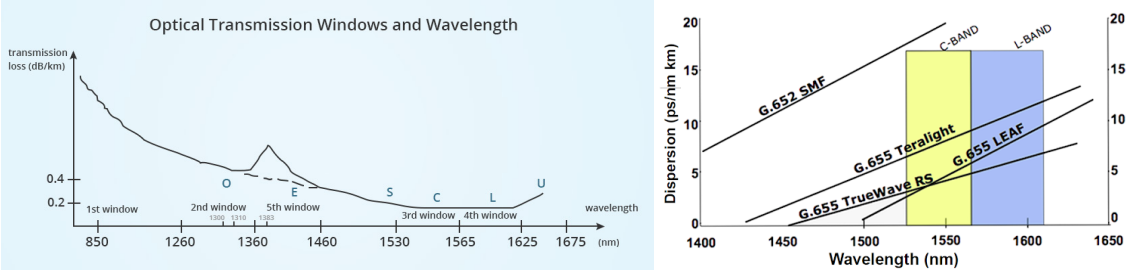


Figure 1.3: Left: Single-mode fiber attenuation distribution. Source: [6]. Right: Single-mode fiber dispersion distribution. Source: [7].

1.2.1 From IM·DD to Early Coherent Optical Communications

Early optical fiber communication systems, as illustrated in Fig. 1.4, predominantly utilized direct detection reception methods. An Intensity Modulation with Direct Detection (IM·DD) link comprises all necessary hardware for modulating Radio Frequency (RF) signals onto optical carriers, including directly modulated laser or Distributed Feedback Laser with an external modulator, and the optical transmission medium, typically a single mode fiber. Additionally, it involves the equipment needed to recover the RF from the optical carrier, a photodetector (PD), and any passive impedance-matching circuits to adapt the impedances of the modulation device and photodetector with the RF source and load impedances.

This method represents a simple approach to optical fiber communication. The PD operates as a square-law detector, capable of detecting only the intensity of the optical signal. Consequently, this communication method is limited to loading information onto the light intensity for transmission. The receiving sensitivity depends on the data transmission rate, while the transmission distance is determined by both the data transmission rate and the thermal noise of the receiver's trans-impedance amplifier.

This direct detection reception method was the standard from the inception of optical fiber communication technology in the 1970s until the early 2000s. The specific technical specifications during this period were predominantly based on GaAs and InPGasAs semiconductor lasers. These lasers, operating at 0.8 μm, were capable of emitting 45 Mbit/s signals over a 10 km relay-free transmission distance. Meanwhile, semiconductor lasers operating at 1.5 μm could emit 2.5

Gbit/s signals with a relay-free transmission distance of 100 km. Until recent years, for baud rates lower than 50 Gbaud over short-distance transmissions, IM·DD still had a certain degree of competitiveness due to its low cost, simple architecture, and receiver sensitivity that does not depend on the carrier phase or the state of polarization (SOP) of the input signal [41–43].

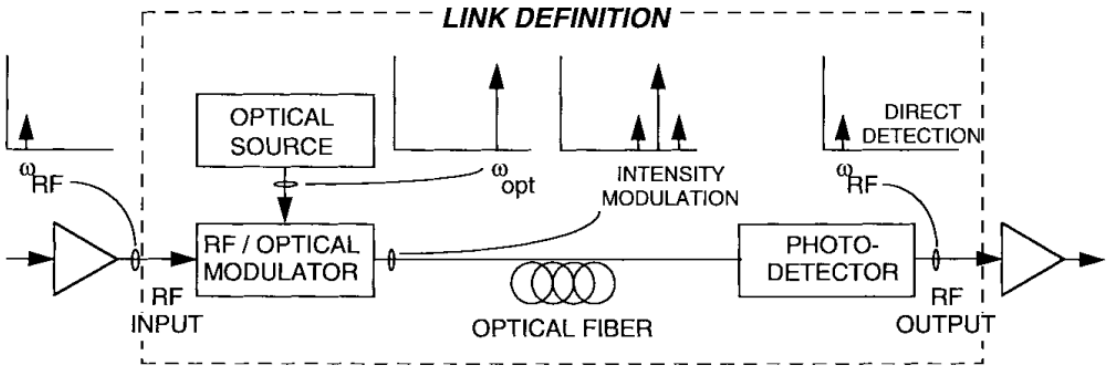


Figure 1.4: Definition of IM·DD analog optical link. Source: [8].

The concept of using heterodyne detection in coherent optical communication, originally proposed by DeLange [44], did not gain much traction in the 1970s when the IM·DD scheme was prevalent in optical fiber communication systems. However, the landscape shifted dramatically in 1980. This change was sparked by Okoshi and Kikuchi [45] and Fabre and LeGuen [46] who demonstrated precise frequency stabilization of semiconductor lasers for heterodyne detection in fiber optic communications. This advancement catalyzed a wave of coherent optical communication system demonstrations. Since the 1990s, coherent detection technology has increasingly become a focal point in optical fiber communication research [9].

The early schematic of coherent detection, shown in Fig. 1.5, represents the first generation of such systems. Coherent detection achieves optimum detection sensitivity, subject to the shot noise limit, especially when a high-power local oscillator is used. In direct detection systems, only the power P_s of the light signal is detectable. In contrast, coherent detection systems can detect a signal with an amplitude proportional to $\sqrt{P_s P_l}$, where P_l is the power of the local oscillator (LO) light. With sufficiently high power in the local oscillator, these systems can reach the detection sensitivity limit. The balanced detector output is then given by:

$$I(t) = I_1(t) - I_2(t) = 2R\sqrt{P_s(t)P_l} \times \cos\{\omega_{IF}t + \theta_{sig}(t) - \theta(t)\}. \tag{1.1}$$

where R is the responsivity of the photodiode, ω_{IF} is the intermediate frequency (IF) given by $\omega_{IF} = |\omega_s - \omega_l|$, and $\theta_{sig}(t)$ and $\theta_l(t)$ are the phases of the transmitted signal and the LO respectively.

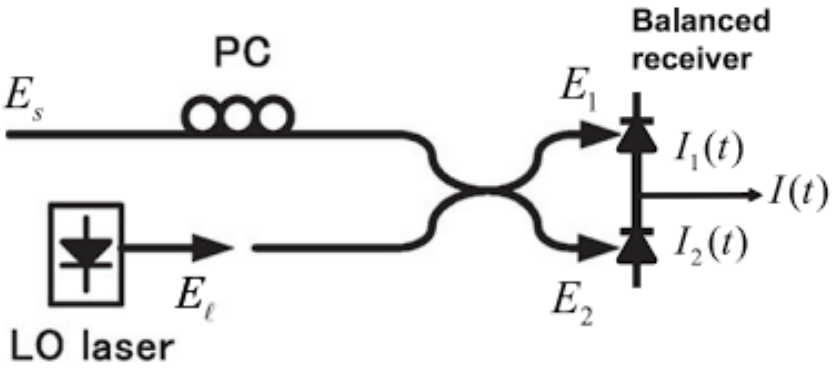


Figure 1.5: Configuration of the coherent receiver that measures the beat between the signal and the LO. Balanced detection eliminates the dc component and maximizes the beat signal. Source: [9].

Coherent detection in the 90s was considered to increase the sensitivity of receivers. The discovery of EDFA has made it possible to achieve similar levels of sensitivity but with direct detection. While single-polarization coherent receivers can recover the in-phase and quadrature (IQ) components, or amplitude and phase, on the optical carrier (scalar complex amplitude), these receivers are highly sensitive to random changes in the phase and SOP of the incoming signal, resulting in coherent systems being more complex in configuration compared to IM·DD systems.

Coherent receivers use the interference of a signal with an LO to extract phase information and include both heterodyne and homodyne architectures. Heterodyne detection refers to a scenario where the frequency difference between the signal carrier and the local oscillator is an intermediate frequency, while homodyne detection occurs when the frequencies of the signal carrier and local oscillator are identical, with a fixed phase difference.

Early coherent detection technology required complex optical phase-locked loops to ensure frequency locking and recover the carrier phase of the received signal. Additionally, a polarization controller (PC) was necessary to align the polarization state of the incident light signal with that of the local oscillator, ensuring maximum efficiency in reception.

1.2.2 Modern Coherent Optical Communications

Modern coherent optical communication systems maximize the information transmission potential by utilizing amplitude, phase, polarization states and wavelength. By adopting Polarization Division Multiplexing (PDM), which uses two orthogonal polarization states to transmit information, these systems effectively double the channel capacity. A critical challenge in these systems, as previously discussed, is aligning the polarization of the input signal with the State of Polarization (SOP) of the Local Oscillator (LO), especially due to the random birefringence changes in

the transmission fiber. To mitigate these challenges, coherent IQ receivers incorporate polarization diversity. This receiver architecture is coupled with real-time DSP techniques at frequencies greater than 50 GSamples/s. Such an approach effectively harnesses the full information-carrying capacity of amplitude, phase and polarization state modulation in a cost efficient way. The result is the achievement of real-time, application-specific integrated circuits (ASICs) for Polarization Division Multiplexing-Quadrature Phase Shift Keying (PDM-QPSK) receivers at 100 Gb/s, and extending up to 800 Gb/s/wavelength ASIC designs [47–50].

As illustrated in Fig. 1.6, a typical homodyne receiver employs both polarization and phase diversities by merging two phase diversity homodyne receivers into a single polarization diversity configuration [51]. An incoming signal with an arbitrary State of Polarization (SOP) is split into its x- and y- polarization components using a Polarizing Beam Splitter (PBS), represented mathematically as:

$$\begin{bmatrix} E_{sx} \\ E_{sy} \end{bmatrix} = \begin{bmatrix} \sqrt{B}E_s e^{j\delta} \\ \sqrt{1-B}E_s \end{bmatrix}. \tag{1.2}$$

Here, E_s denotes the complex amplitude of the incoming signal, B is the power splitting ratio between the two polarization components, and δ is the phase difference between them. Meanwhile, the LO is split into x- and y- paths by a quarter-wave plate and then being made circular by a quarter-wave plate (QWP), expressed as:

$$\begin{bmatrix} E_{LO,x} \\ E_{LO,y} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} E_{LO} \\ jE_{LO} \end{bmatrix}. \tag{1.3}$$

The electric fields generated by the 90° optical hybrids at the inputs of four balanced photodiodes PD1-PD4 are:

$$\begin{bmatrix} E_{1,4} & E_{2,3} \\ E_{5,8} & E_{6,7} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\left(E_{sx} \pm \frac{1}{\sqrt{2}}E_{LO}\right) & \frac{1}{2}\left(E_{sx} \pm \frac{j}{\sqrt{2}}E_{LO}\right) \\ \frac{1}{2}\left(E_{sy} \pm \frac{1}{\sqrt{2}}E_{LO}\right) & \frac{1}{2}\left(E_{sy} \pm \frac{j}{\sqrt{2}}E_{LO}\right) \end{bmatrix}. \tag{1.4}$$

The resultant photocurrents from PD1 to PD4 are proportional to the square root of the power splitting ratio B, the magnitude of E_s, its phase θ and the phase difference δ between the two polarization components, and are expressed as:

$$\begin{bmatrix} I_{PD1} & I_{PD2} \\ I_{PD3} & I_{PD4} \end{bmatrix} \propto \begin{bmatrix} \sqrt{B}\,|E_s|\cos(\theta+\delta) & \sqrt{B}\,|E_s|\sin(\theta+\delta) \\ \sqrt{1-B}\,|E_s|\cos\theta & \sqrt{1-B}\,|E_s|\sin\theta \end{bmatrix} \tag{1.5}$$

Ultimately, we can detect the signal at the reception side. The x and y polarization components of the signal's complex amplitude are accurately determined by the equation:

$$\begin{bmatrix} E_{sx} \\ E_{sy} \end{bmatrix} = \begin{bmatrix} I_{PD1} + jI_{PD2} \\ I_{PD3} + jI_{PD4} \end{bmatrix} \tag{1.6}$$

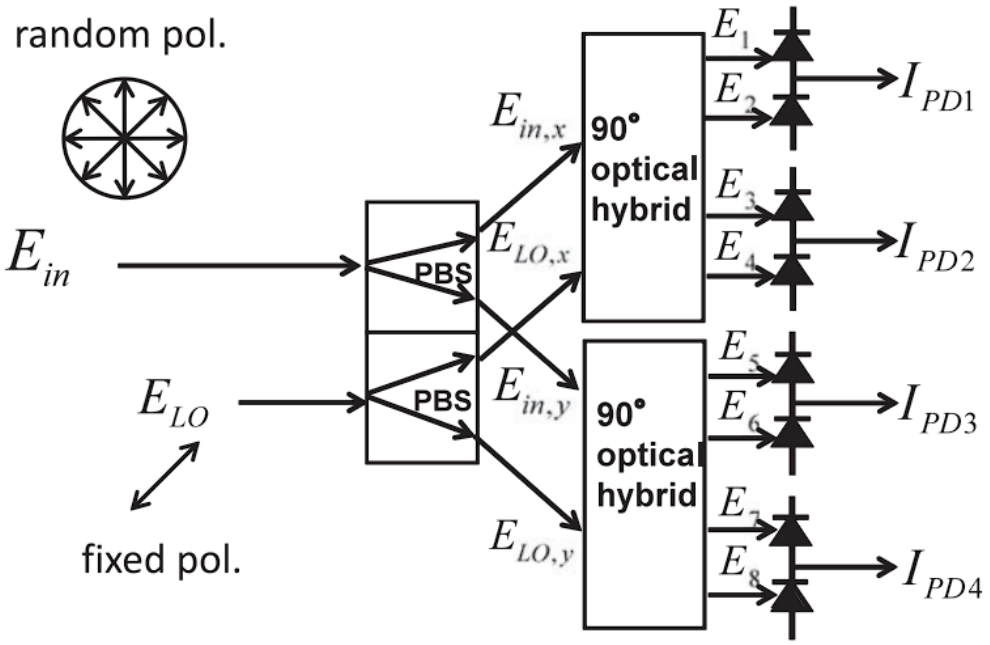


Figure 1.6: Configuration of the homodyne receiver employing phase and polarization diversities. Two phase-diversity homodyne receivers are placed in the polarization-diversity configuration, where a common LO is employed. Source: [9].

By utilizing the described method, we can fully recover and interpret the signal's features at the detection point. The recovery of the x and y polarization components of the signal's complex amplitude is essential for the receiver to reconstruct the transmitted information accurately. The configuration and capabilities of this receiver showcase the sophisticated techniques employed in coherent optical communication systems. These systems adeptly harness the information-carrying potential of phase and polarization states while addressing the challenge of random birefringence changes in transmission fibers.

Additionally, it is important to emphasize that at the receiving end, a suite of critical functions including frequency tracking, carrier phase recovery, equalization, synchronization, polarization tracking, and demultiplexing are all efficiently executed through Digital Signal Processing (DSP) technology [52, 53]. This technological progress not only streamlines the hardware design of the receiver but also considerably enhances the signal recovery capabilities. The complexities and intricacies of these DSP algorithms will be elaborated upon in the subsequent section.

1.2.3 Noise Sources

In coherent communications, several types of additive noise crucially affect system performance. Thermal and electronic noise arise from random electron motion in conductors and electronic devices, respectively, and are overshadowed by other noise sources in long-haul optical transmissions with coherent detection, which are typically shot noise from LO and accumulated ASE noise from inline amplifiers. Shot noise, a quantum phenomenon in photodetectors, results from the random fluctuation of photocurrent due to photon arrival, often assuming a Gaussian distribution in high-LO-power systems. In many practical coherent detection systems, shot noise is often considered negligible. Beating noise occurs when the LO beats with the received optical and ASE noise fields, causing power fluctuations. ASE noise, particularly prevalent in long-distance communications using optical amplifiers like EDFA and Raman, is a major additive noise due to the random multiplication of signal photons and spontaneous photon emission, typically exhibiting a circularly symmetric complex Gaussian distribution. In our subsequent simulations, to simplify the analysis, we consider an ideal coherent receiver, ignoring thermal and electronic noise, shot noise, and signal-ASE beating noise.

1.3 Simulation Framework for Optical Coherent Transmission

Our simulation system is constructed based on Optilux [54], an open-source suite of algorithms developed for the design, simulation, and analysis of optical communication systems. Implemented as a MATLAB/OCTAVE toolbox, Optilux provides efficient and advanced tools for system analysis of optical fiber transmissions, especially in modeling wave propagation within the fiber. This comprehensive toolbox was designed by Prof. Paolo Serena from the Department of Engineering and Architecture at the University of Parma, Italy. Leveraging Optilux, we have constructed a point-to-point optical fiber transmission system encompassing components from the transmitter to the receiver side, facilitating a detailed and accurate simulation environment for our studies.

1.4 Transmitter Side

The complete process flow of the transmitter configuration is depicted in the top row of Fig. 1.7. Initially, we generate random bits either from a Bernoulli (1/2) distribution or by using a pseudo-random binary sequence of the De Bruijn type [55]. These bits are subsequently modulated using specified N-dimensional (N-D) modulation techniques, which may include two-dimensional M-PSK or M-QAM, four-dimensional high-order modulation, or M-QAM based probability constellation shaping or sequence selection. Following modulation, a pilot sequence comprising a synchronization sequence [56] and a training sequence for channel estimation is prefixed to

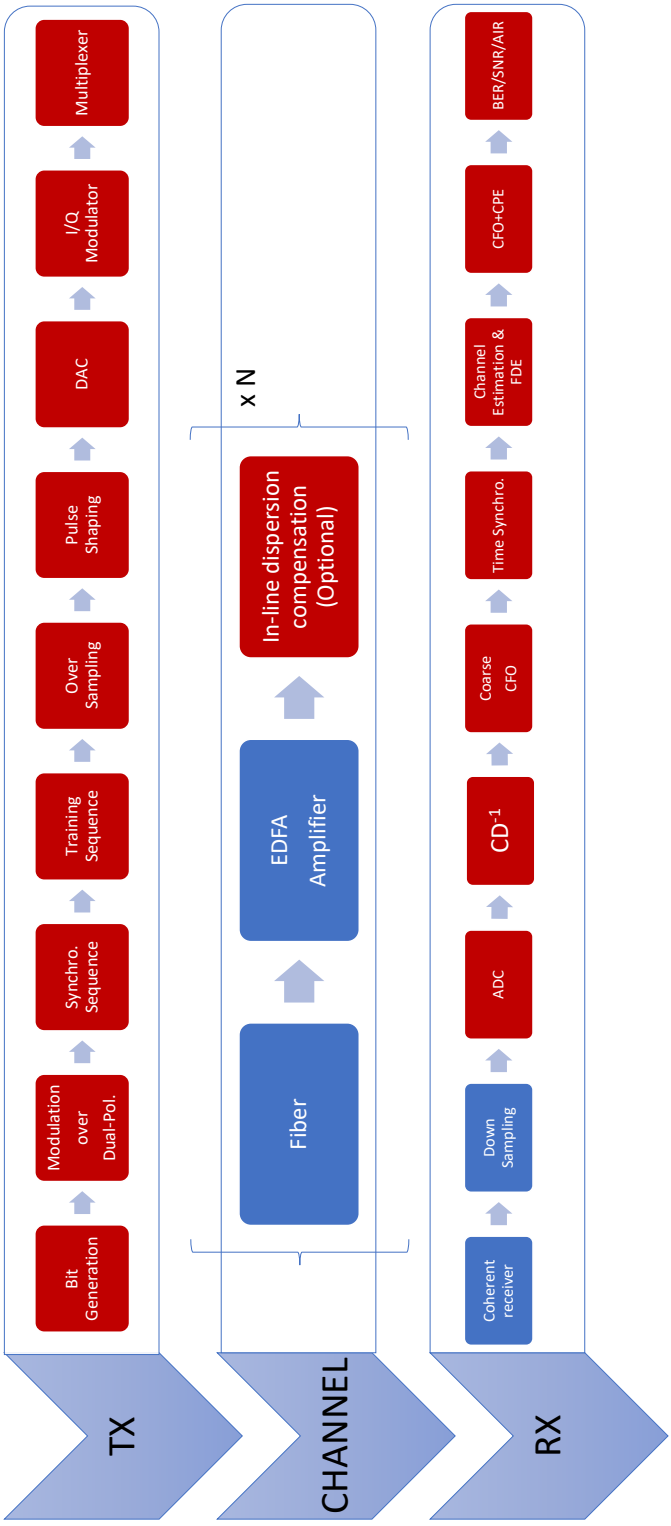


Figure 1.7: Block diagrams of the simulation system. In our simulation system, components from Optilux are utilized as-is and represented as the blue module, while the red module indicates the custom code that we have developed and integrated into the system.

the data stream. For channel estimation, we employ polarization-multiplexed constant amplitude zero auto-correlation (CAZAC) sequences [57, 58] for data-assisted estimation or use genie-aided equalization when assuming that we perfectly known the linear channel. After oversampling, pulse shaping is conducted using a root-raised cosine filter with a typical roll-off factor of 0.1 to limit the transmitted signal's bandwidth. The digital signal is then converted into an analog signal through a Digital-to-Analog Converter (DAC), which then drives the Mach-Zehnder Modulators (MZM) of the I/Q modulator to modulate the light wave produced from the laser. The structure of a single-polarization I/Q MZM is shown in Fig. 1.8 and it can also be used as a building block of the dual-polarization MZM. The modeled DACs implement a uniform quantization using a fixed number of bits, called the vertical resolution of the DAC. They may be switched on or off depending on the desired simulation parameters. Finally, for multi-channel configurations, a multiplexer is simulated to implement Wavelength Division Multiplexing (WDM) transmission.

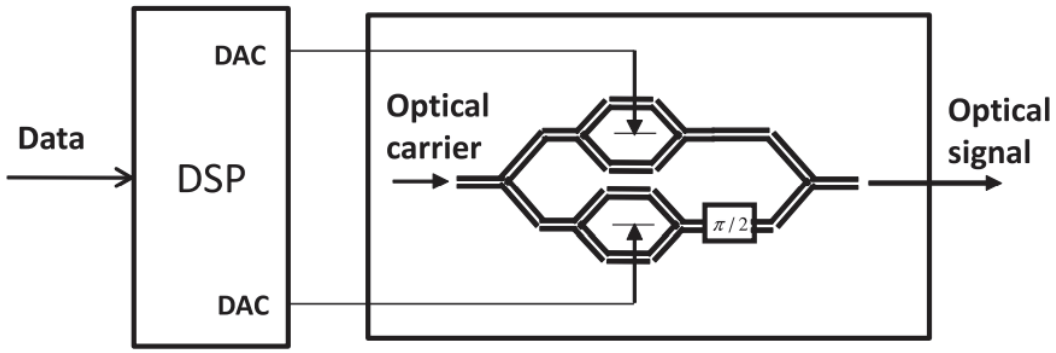


Figure 1.8: Configurations of the DSP-based optical transmitter. Analog signals after DSP and DAC drive the optical IQ modulator. Source: [9].

Coherent optical communication systems are complex and vulnerable to various impairments that can significantly affect their performance. Power losses, laser distortions and noise, for instance, can restrict the system's bandwidth and deteriorate the signal's integrity. Given that these effects can be notably frequency dependent, pulse shaping limiting the signal bandwidth such as the root-raised cosine (RRC) filter is particularly useful. On the transmitter side, a prevalent issue is the quantization noise introduced by Digital-to-Analog Converters (DACs). The limited quantization levels of DACs can compromise the accuracy of signal generation, leading to a less precise representation of the intended signalling scheme. Mach-Zehnder Modulators (MZMs) are also susceptible to imbalances such as gain imbalance, quadrature imbalance, and skew between the in-phase (I) and quadrature (Q) components of each polarization branch. These imbalances and skews can distort the resulting signal, negatively impacting the transmission quality and the overall system performance. Additionally, laser sources, which are fundamental to coherent systems,

can introduce impairments such as laser intensity noise and phase noise. The latter is particularly critical in coherent transmissions, as it can significantly affect the system's performance by inducing errors in the detection of phase-sensitive signals such as Quadrature Amplitude Modulation (QAM) formats. The laser's linewidth, a measure of its spectral purity, primarly dictates the amount of added phase noise to the system.

To address these challenges, a comprehensive strategy combining meticulous hardware design and sophisticated DSP techniques is essential. This approach ensures efficient and accurate data transmission over optical networks while maintaining the integrity and reliability of communications. The specific impairments related to DACs and MZMs will be further elaborated and discussed in Sec.3.3.2. This section primarily focuses on introducing the application of pulse shaping in the form of an RRC filter and the addition of laser phase noise.

1.4.1 The RRC Filter

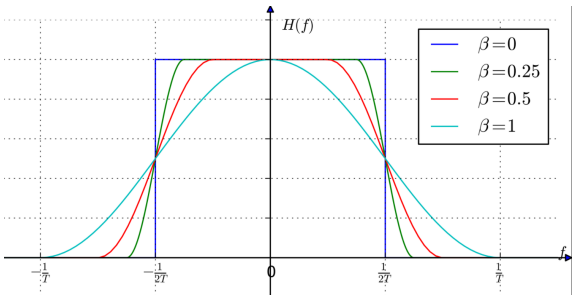


Figure 1.9: Frequency responses of RRC filters with various roll-off factors β. Source:[10]

The raised-cosine filter is commonly used for pulse shaping in digital modulations due to its ability to achieve enhanced out-of-band rejection and a spectrum free of ringing effects. It derives its name from the fact that the non-zero part of its simplest frequency spectrum shape (when the roll-off factor $\beta = 1$) is a cosine function, "raised" to sit above the frequency axis f. The RRC filter, sometimes known as the square-root-raised-cosine (SRRC) filter, gets its name because its frequency response, $H_{rrc}(f)$, is the square root of the frequency response of the raised-cosine filter, $H_{rc}(f)$. Its frequency domain representation is a piece-wise function defined as:

$$H_{rc}(f) = \begin{cases} 1, & |f| \leq \frac{1-\beta}{2T} \\ \frac{1}{2}[1+\cos(\frac{\pi T}{\beta}[|f|]-\frac{1-\beta}{2T})], & \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \\ 0, & \text{otherwise} \end{cases} \tag{1.7}$$

$$0 \leq \beta \leq 1. \tag{1.8}$$

$$|H_{rrc}(f)| = \sqrt{|H_{rc}(f)|} \tag{1.9}$$

where T is the symbol period of the communication system, and β is the roll-off factor. The occupied bandwidth can be controlled by adjusting β.

Fig. 1.9 illustrates the frequency response of the RRC filter with various β values. When $\beta = 1$, the RRC filter is a cosine function, and its single-side bandwidth is the same as the symbol rate. As β decreases, the filter's shape increasingly resembles a rectangular window, and when $\beta = 0$, the filter's bandwidth is the half of the symbol rate. Although β cannot be exactly 0 due to the fraction in the filter expression, it can be very small. In simulations, choosing $\beta = 0.1$, the filter's bandwidth is 0.55 times the symbol bandwidth. For a 100 Gbaud transmission, the RRC filter's single-side bandwidth is 55 GHz.

1.4.2 Laser Phase Noise

Coherent laser sources with highly pure spectra are pivotal in coherent communications [59] where ultra-narrow linewidth lasers with minimized phase noise are used to enhance performance towards fundamental limits. Recent advancements in frequency stabilization techniques have enabled coherent laser radiation with extremely narrow linewidth values, defined as the full width at half maximum of a laser, reaching down to millihertz levels. However, accurately characterizing such narrow linewidths remains a significant challenge.

The linewidth of a laser is fundamentally influenced by its radiation coupled with spontaneous emission, a quantum process producing pure white noise, leading to the 'Schawlow-Townes linewidth' [60] which theoretically exhibits a Lorentzian lineshape [61]. This intrinsic quantum-limited linewidth reflects the laser's short-term stability and temporal coherence.

Intrinsic linewidth modeling of lasers is crucial for understanding the random phase fluctuations introduced by spontaneous emission, a fundamental quantum process during laser generation. These fluctuations cause the laser's output electrical field to deviate from an ideal sinusoidal wave. A quasi-monochromatic laser field can be expressed as a sinusoidal wave with random phase fluctuations:

$$E(t) = E_0 e^{j[2\pi f_0 t + \varphi(t)]}. \tag{1.10}$$

Here, E_0 represents the stationary amplitude value, and f_0 is the nominal central optical frequency. Under ideal conditions, the frequency and phase of the light waves output by a single-frequency single-longitudinal mode laser are constant. The output is a single spectral line of frequency f_0. The spontaneous emission causes random fluctuations in phase and amplitude, resulting in laser spectral line broadening and output power fluctuations. The random intensity perturbation of E_0 is intensity noise, the random frequency perturbation of f_0 is called frequency noise, and the random phase perturbation of φ is called phase noise. Therefore, the randomness of spontaneous emission is the source of laser noise. The phase fluctuation, $\varphi(t)$, induced by

white frequency noise, is akin to a continuous random walk, typically modeled by the Wiener process [62]. The phase jitter after a positive time delay τ, denoted as $\Delta\varphi(t) = \varphi(t+\tau) - \varphi(t)$, is a random variable with a Gaussian distribution with a probability density function:

$$p(\Delta\varphi) = \frac{1}{\sqrt{2\pi\sigma_\varphi^2}} e^{-\frac{\Delta\varphi^2}{2\sigma_\varphi^2}}. \tag{1.11}$$

which is a zero mean Gaussian distribution of variance σ_φ^2 directly proportional to the time delay τ which is the sampling time in the simulation. The relationship between the intrinsic linewidth ($\Delta\nu_c$) and the phase jitter variance [63, 64] is given by:

$$\sigma_\varphi^2 = 2\pi\Delta\nu_c\tau. \tag{1.12}$$

The laser phase noise of the local oscillator also impacts the received signal through the beating terms between the incident signal and the LO in the coherent receiver. Furthermore, over long-haul multi-span dispersion uncompensated fiber transmissions, laser phase noise due to the no-zerolaser linewidth of the LO may be enhanced at the receiver side by the digital dispersion compensation block, resulting in equalization-enhanced phase noise (EEPN) that is enhanced as the cumulative dispersion increases, especially for high data rates [65, 66]. In these conditions, EEPN can impose a relatively large penalty, which needs to be compensated by DSP.

1.5 Propagation Channel Model

When simulating the optical channel for fiber transmission using the Optilux framework, two primary modes are employed: Unique Field Mode and Separate Fields Mode. In the Unique Field Mode, all WDM channels are considered as a single propagating signal and are multiplexed in the fiber using an optical multiplexer, with all the propagation effects being added directly to this unified signal. This mode simplifies implementation and reduces the complexity of the model. Conversely, in the Separate Fields Mode, each channel is propagated separately, allowing for individual consideration of effects and the addition of one-by-one interactions between channels. This mode offers a more granular investigation of each effect by facilitating the independent addition or removal of specific influences. For instance, in the Unique Field Mode, maintaining an intra-channel nonlinear effect, a pure self-phase modulation (SPM) effect, while disregarding inter-channel nonlinear effects such as cross-phase modulation (XPM) effects is unfeasible since all channels are combined into a single signal. In our simulations, we will mostly use the Unique Field Mode unless stated otherwise.

The channel model typically comprises homogeneous fiber spans consisting of a fiber cable followed by an Erbium-Doped Fiber Amplifier (EDFA). Various linear and nonlinear effects in the

fiber are accounted for and simulated using the Split-Step Fourier Method (SSFM) [67–76]. These effects include Chromatic Dispersion (CD), Polarization Mode Dispersion (PMD), random polarization rotations, fiber attenuation, and nonlinear Kerr effect, which encompasses Self-Phase Modulation (SPM), Cross Phase Modulation (XPM), Cross Polarization Phase Modulation (XpolM), and Four-Wave Mixing (FWM).

The presentation of the propagation channel model is divided into four distinct parts. First, we recall the basic Nonlinear Schrödinger Equation (NLSE) to set the ground for understanding signal propagation in the optical fiber, including the basic effects such as attenuation, CD, and Kerr effects. Next, we focus on linear optical transmission model to visualize all linear effects in the fiber, including CD, PMD, PDL, and dynamic state of polarization (SOP) rotation. Then, a comprehensive description of nonlinear effects is provided. This structured approach allows for a simpler comprehension of both the linear and nonlinear dynamics impacting the signal as it traverses the optical channel. Lastly, we address the impact of Amplified Spontaneous Emission (ASE) noise, a significant consideration in long-haul optical communication systems. This noise is emitted by EDFA and is an inherent factor in signal degradation and system performance evaluation.

1.5.1 Nonlinear Schrödinger Equation

Analyzing fiber effects on the signal is crucial for the successful recovery of information using the DSP stage of an optical communication system. Modeling the propagation of dual polarization modulated signals in optical fibers typically relies on the simplified coupled nonlinear Schrödinger equation (CNLSE) formalism [77]. This approach outlines the evolution of the slow-varying envelopes $S_x(z,t)$ and $S_y(z,t)$ along the propagation distance z, considering factors like loss, second-order chromatic dispersion, and Kerr nonlinear effects. The CNLSE is represented as:

$$\begin{cases} \frac{\partial S_x}{\partial z} + \beta_{1x}\frac{\partial S_x}{\partial t} + i\frac{\beta_{2x}}{2}\frac{\partial^2 S_x}{\partial t^2} + \frac{\alpha_x}{2}S_x = i\gamma\left(|S_x|^2 + \frac{2}{3}|S_y|^2\right)S_x \\ \frac{\partial S_y}{\partial z} + \beta_{1y}\frac{\partial S_y}{\partial t} + i\frac{\beta_{2y}}{2}\frac{\partial^2 S_y}{\partial t^2} + \frac{\alpha_y}{2}S_y = i\gamma\left(|S_y|^2 + \frac{2}{3}|S_x|^2\right)S_y \end{cases} \tag{1.13}$$

In Eq. (1.13), β_n denotes the n^{th} derivatives of the propagation constant β. The term α represents the fiber attenuation coefficient, while γ is the nonlinear coefficient of the fiber. It is assumed that both $\beta_{2,x/y}$ and $\alpha_{x/y}$ are equal for both polarization components, indicating that Polarization Dependent Loss (PDL) is neglected. Note that $S_x(z,t)$ and $S_y(z,t)$ are not merely the envelopes of a given channel but represent the total WDM field envelope of each polarization. Therefore, Eq. (1.13) inherently includes Four Wave Mixing (FWM) effects for each polarization state.

In scenarios involving dual-polarization transmission, additional complexities such as Polarization Mode Dispersion (PMD) and polarization crosstalk need to be considered. The polarization

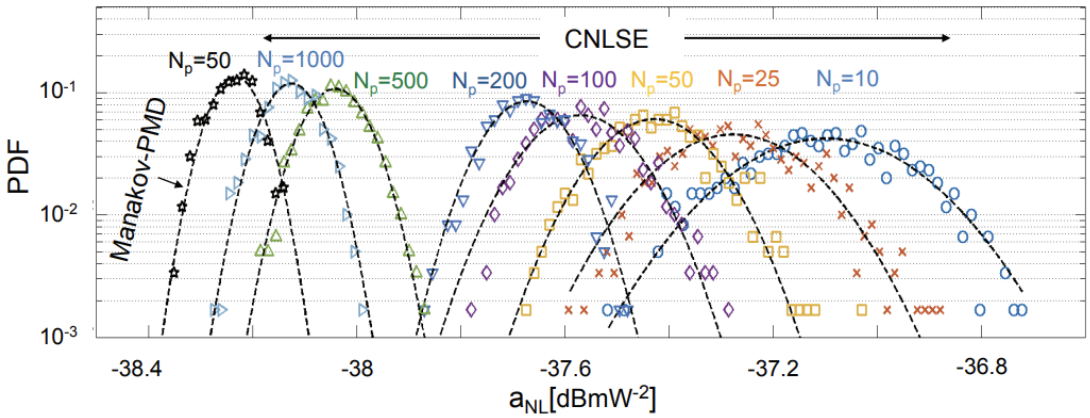


Figure 1.10: CNLSE vs Manakov-PMD in different wave-plates [11]. N_p is the number of wave-plates.

changes are simulated by cutting each fiber span into several wave plates (WP), and the XpolM can be simulated by CNLSE or Manakov-PMD equation [78] which neglects the nonlinear-PMD term and offers a faster simulation time than the CNLSE while being enough accurate for fibers with short birefringence correlation length (around less than 10 m). When the correlation length L_{corr} [79] is significantly lower than the nonlinear Kerr length L_{NL} [73], the nonlinear effects on signals can be effectively modeled by averaging over polarization components. This simplification neglects the nonlinear-PMD term and significantly reduces simulation computation time [73, 80–82]. The Manakov-PMD equation is given by:

$$\begin{aligned}
&\frac{\partial S_x}{\partial z}+\beta_{1x}\frac{\partial S_x}{\partial t}+i\frac{\beta_{2x}}{2}\frac{\partial^2 S_x}{\partial t^2}+\frac{\alpha_x}{2}S_x=i\gamma\frac{8}{9}\left(|S_x|^2+|S_y|^2\right)S_x\\
&\frac{\partial S_y}{\partial z}+\beta_{1y}\frac{\partial S_y}{\partial t}+i\frac{\beta_{2y}}{2}\frac{\partial^2 S_y}{\partial t^2}+\frac{\alpha_y}{2}S_y=i\gamma\frac{8}{9}\left(|S_y|^2+|S_x|^2\right)S_y
\end{aligned} \tag{1.14}$$

In the CNLSE (Eq. (1.13)), the terms $|S_x|^2S_x$ and $|S_y|^2S_y$ correspond to Kerr nonlinearities within each polarization tributary, while the terms $\frac{2}{3}|S_y|^2S_x$ and $\frac{2}{3}|S_x|^2S_y$ pertain to cross-polarization nonlinearities. The Manakov-PMD equation (Eq. (1.14)) introduces a coefficient of $\frac{8}{9}$, which adjusts the weights of intra-tributary nonlinearities about the weights of cross-polarization nonlinear effects. This coefficient reflects a more averaged approach to modeling the combined effects of phase and polarization state variations in the medium.

In [11], Manakov-PMD shows good accuracy with lower running time. Indeed, in Fig.1.10, Manakov-PMD with 50 WP gives a lower fluctuation than CNLSE with 1000 WP. Given the better accuracy achieved by Manakov-PMD, we choose this method with 50 WP per span to implement the SSFM throughout our study.

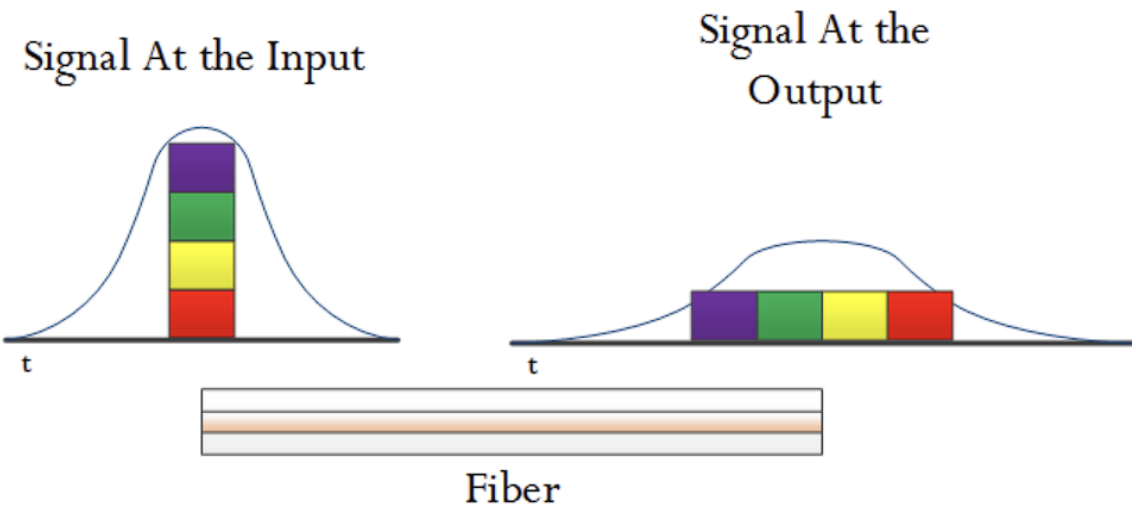


Figure 1.11: Illustration of Chromatic Dispersion on an optical silica fiber. A single optical pulse is composed of a range of wavelengths. Because each 'color' travels in the fiber at a slightly different speed, the wavelengths separate as the pulse travels down the fiber. Source: [12].

1.5.2 Linear Transmission System Model

Next, we will ignore the nonlinear Kerr effect for the time being and focus on the linear optical transmission system model to conduct a concentrated analysis and understanding of several impairments in the channel. These include chromatic dispersion, polarization mode dispersion, polarization dependent loss, and random polarization rotation:

- Chromatic Dispersion (CD) refers to the broadening of light pulses as they travel through the optical fiber due to the different speeds of wavelengths. As illustrated in Fig. 1.11, the signal undergoes broadening during its propagation through the fiber.

- Polarization Mode Dispersion (PMD): This results from the differential speed of light in various polarization states within the fiber, leading to pulse broadening. The Differential Group Delay (DGD) is a manifestation of PMD, indicating the time difference between the fastest and slowest polarization modes.

- Polarization Dependent Loss (PDL): This is the variation in attenuation of light through the optical medium depending on its state of polarization. It can cause pulse narrowing even with finite differential group delay [83] and mainly occurs in optical components such as amplifiers and wavelength selective switches rather than in the fiber itself.

- Random Polarization Rotation: Due to random fluctuations in the fiber's birefringence, the state of polarization of light changes randomly over time and length, complicating the

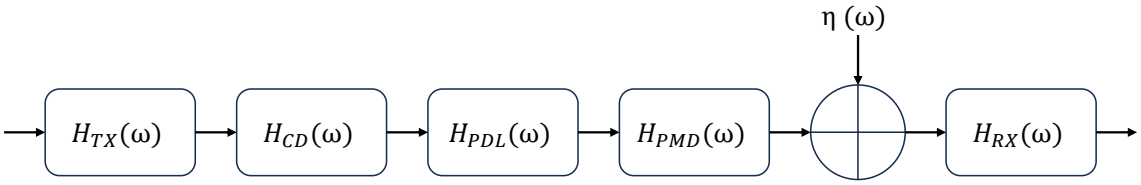


Figure 1.12: Optical transmission system model.

control of polarization effects. Higher PMD values lead to a larger range of polarization-dependent chromatic dispersion and principal state of polarization rotation speed [84].

The intertwined and complex nature of these effects in optical fiber communication necessitates a range of compensation and management techniques to maintain signal integrity and system performance. Comprehending and mitigating these effects is crucial to enhancing the reliability and speed of fiber optic communications. Consequently, we introduce the optical transmission system model, which encapsulates all deterministic channel impairments under the assumption of linear or weakly nonlinear conditions [85], the 2×2 frequency-domain transfer matrix of the channel $\mathbf{H}(\omega)$ at the angular frequency ω is described as:

$$\mathbf{H}(\omega) = \mathbf{H}_{Rx}(\omega)\mathbf{H}_{PMD}(\omega)\mathbf{H}_{PDL}(\omega)\mathbf{H}_{CD}(\omega)\mathbf{H}_{Tx}(\omega). \tag{1.15}$$

Here, $\mathbf{H}_{Tx}(\omega)$ is the transfer function accounting for any pulse shaping or low-pass filtering at the transmitter, and the transfer function $\mathbf{H}_{Rx}(\omega)$ contains the contribution of the optical and electrical low-pass filtering at the receiver [86]. PMD and PDL are approximately fixed values during the coherence time of the channel. Typically, this coherence time ranges between 1 μs to 1 ms in practical scenarios. Consequently, the transfer function $\mathbf{H}(\omega)$ can be effectively determined using learning sequences if the frame duration is shorter than the coherence time, specifically less than $1-10$ μs. The Chromatic Dispersion (CD) transfer function $\mathbf{H}_{CD}(\omega)$ is expressed as:

$$\mathbf{H}_{CD}(\omega) = \exp\left(j\frac{D\lambda^2\omega^2 L}{4\pi c}\right). \tag{1.16}$$

In the context of optical fibers, the coefficient $D = -\frac{2\pi c\beta_2}{\lambda^2}$ in ps/nm/km represents the accumulated chromatic dispersion of a signal after it has propagated through an optical fiber of length L km, with the group velocity dispersion β_2 characterized in ps^2/km. Here, λ in nm is the central wavelength of the signal and c is the speed of light in a vacuum. Additionally, j represents the imaginary unit, defined as $\sqrt{-1}$.

The Hermitian Polarization Dependent Loss (PDL) matrix $\mathbf{H}_{PDL}(\omega)$ is defined as [87]:

$$\mathbf{H}_{PDL}(\omega) = \begin{bmatrix} \cos(\theta') & -\sin(\theta') \\ \sin(\theta') & \cos(\theta') \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \kappa \end{bmatrix} \times \begin{bmatrix} \cos(\theta'') & -\sin(\theta'') \\ \sin(\theta'') & \cos(\theta'') \end{bmatrix} \tag{1.17}$$

where θ' and θ'' are the polarization rotation angles at the input and output of the equivalent accumulated-PDL element corresponding to the whole link, respectively, and κ ($0 < \kappa < 1$) is the PDL coefficient defined by the attenuation between the orthogonal lossy and lossless axes, expressed in dB as: $\kappa_{dB} = -20\log(\kappa)$. When many optical elements with non-zero PDL are present at different nodes in the link, κ_{dB} follows a Maxwellian distribution.

Polarization Mode Dispersion (PMD) can be modeled by concatenating N_p independently oriented fiber sections (or wave-plates) with independent uniformly distributed birefringence:

$$\begin{aligned}\mathbf{H}_{PMD}(\omega) &= \prod_{s=1}^{N_p} \mathbf{H}_{DGD_s}(\omega) \\ &= \prod_{s=1}^{N_p} \begin{bmatrix} \cos(\theta'_s) & \sin(\theta'_s) \\ -\sin(\theta'_s) & \cos(\theta'_s) \end{bmatrix} \times \begin{bmatrix} e^{j(\phi_s+\omega\tau_s)/2} & 0 \\ 0 & e^{-j(\phi_s+\omega\tau_s)/2} \end{bmatrix} \times \begin{bmatrix} \cos(\theta'_s) & -\sin(\theta'_s) \\ \sin(\theta'_s) & \cos(\theta'_s) \end{bmatrix}.\end{aligned} \tag{1.18}$$

In the PMD model, the input and output fields are rotated according to the principal State of Polarization (SOP) of each fiber section s. For each Differential Group Delay (DGD) element, the angle mismatch between the incident SOP of the signal and the principal slow and fast axis of the PMD element is θ'_s. Each Hermitian birefringence matrix is defined by ϕ_s which refers to the polarization phase and τ_s characterizing the DGD between the two orthogonal fast and slow axes. The all-order PMD approximation is obtained for $N_p = 50$, while the first-order PMD (also referred to as DGD) is achieved for $N_p = 1$.

1.5.3 ASE Noise

In optical fiber communication, the introduction of noise by optical amplifiers is a critical consideration. While amplifying an optical signal (of a specific frequency band) is necessary to increase the fiber's transmission distance, the amplification process is not entirely controllable. An EDFA is typically employed in the receiver part to compensate for the losses incurred during transmission. An essential consideration in this context is the ASE noise generated by the EDFA [88].

This noise arises from the spontaneous emission of the amplifier, occurring when activated particles in the EDFA return to their ground state from an excited state, simultaneously amplifying the optical signal and producing random, incoherent spontaneous emissions. These emissions are the primary source of ASE noise, which has several characteristics. Firstly, the spontaneous incoherent emission of stimulated particles can occur in any direction, potentially causing additional stimulated radiation and amplification, essentially equating to amplified noise. Secondly, the frequency band of ASE noise is very wide, often encompassing the entire gain bandwidth of the amplifier. Thirdly, the excited particle count is determined under certain conditions: the more

particles involved in generating ASE, the fewer are available to provide signal gain. Consequently, as ASE noise increases, the amplified signal power decreases, reducing the signal gain.

A simplified model for understanding the performance of EDFA involves the ASE power spectral density N_0, which is measured after the amplifier and is related to the amplifier's gain and its spontaneous emission factor, or population inversion factor n_{sp} [89]:

$$N_0 = 2hf_0(G-1)n_{sp} \quad . \tag{1.19}$$

In this equation, h is Planck's constant, f_0 is the central frequency of the signal, G is the amplifier gain, and n_{sp} is the population inversion factor ($1 < n_{sp} \leq 2$). Typically, the EDFA noise figure F is used instead of its spontaneous emission factor, defined as the ratio between the optical signal-to-noise ratio (OSNR) at the input and the output of the amplifier. By assuming a large gain value ($G \gg 1$), the noise figure F can be approximated as:

$$F = \frac{OSNR_{in}}{OSNR_{out}} = \frac{1}{G} + \frac{2n_{sp}(G-1)}{G} \approx 2n_{sp} \quad . \tag{1.20}$$

For long-haul mutli-span transmission links, ASE noise accumulates incoherently with each span of the optical transmission system, affecting the Optical signal-to-noise ratio (OSNR) evaluated at the end of the link. The OSNR over a given reference bandwidth B_{ref} , which is conventionally chosen to be 0.1 nm (12.5 GHz) can be calculated as:

$$OSNR_{B_{ref}} = \frac{P_{ch}}{N_{span}N_0B_{ref}} \approx P_{ch(dBm)} - (-58\ \text{dBm} + NF + 10\log_{10}(N_{span}) + \alpha L_{span}). \tag{1.21}$$

Here, P_{ch} represents the power of the channel, N_{span} is the number of spans, N_0 is the ASE power spectral density, L_{span} is the length of each span in the optical link, and NF is the amplifier noise figure in dB. This relationship underscores the need for managing ASE noise to maintain a high OSNR, which is crucial for the integrity and quality of the received optical signal.

1.5.4 Nonlinear Fiber Channel

Under the assumption of low to negligible accumulated PDL, linear impairments in optical communication systems can be effectively compensated for on the receiver side, allowing us to approach the capacity of linear Additive White Gaussian Noise (AWGN) channels by even increasing the Signal-to-Noise Ratio (SNR). However, in nonlinear fiber channels, nonlinear effects limit the capacity as illustrated in Fig. 1.13. At overly high SNR levels, significant nonlinear interference is generated, leading to degradation in transmission performance. This nonlinear interference acts as a fundamental barrier, preventing further improvements in capacity even when the SNR increases, marking a distinct difference from the behavior observed in linear AWGN channels.

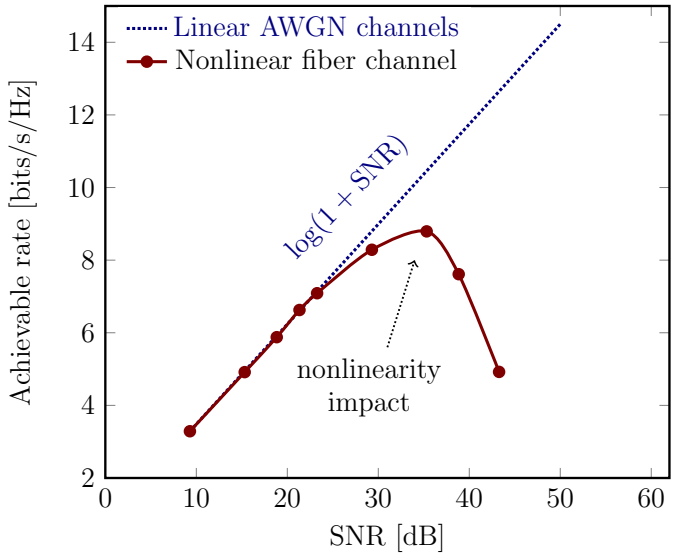


Figure 1.13: Capacity problem in fiber [13].

To effectively combat Nonlinear Interference (NLI) in Wavelength Division Multiplexing (WDM) systems, it is crucial to understand its specific composition. As depicted in Fig.1.14, NLI in WDM transmissions can be divided into two categories: deterministic and stochastic. Deterministic interactions are primarily between signal components, while stochastic interactions involve both signals and noise. It is worth mentioning that inter-channel interactions can also be considered stochastic when the data on adjacent channels is unknown to the channel under consideration. Moreover, we mainly focus on interactions between signal components themselves and signal and noise components, as noise-noise interactions are typically overwhelmed by the signal during propagation and therefore have a negligible impact [90].

The main contribution to NLI primarily comes from signal-to-signal interactions, which are generally divided into intra-channel and inter-channel types. Intra-channel NLI originates only from frequency components within the channel of interest, whereas inter-channel NLI involves other interfering channels. Notably, as the number of WDM channels increases, inter-channel NLI tends to become more significant than intra-channel NLI, which is particularly true for practical WDM systems.

In the frequency domain, signal-to-signal NLI is fundamentally divided based on the interacting frequency components. SPM is generated from a single frequency component of the channel of interest. Inter-channel Cross-Phase Modulation (XPM) and Four-Wave Mixing (FWM) occur when two or three signal frequency components from different WDM channels interact. These nonlinear effects result from the intensity variations of one channel impacting the phase of adjacent channels (XPM) or mixing to produce new frequencies (FWM). Conversely, XPM and FWM can also manifest within a single channel, termed intra-channel effects, when all interacting fre-

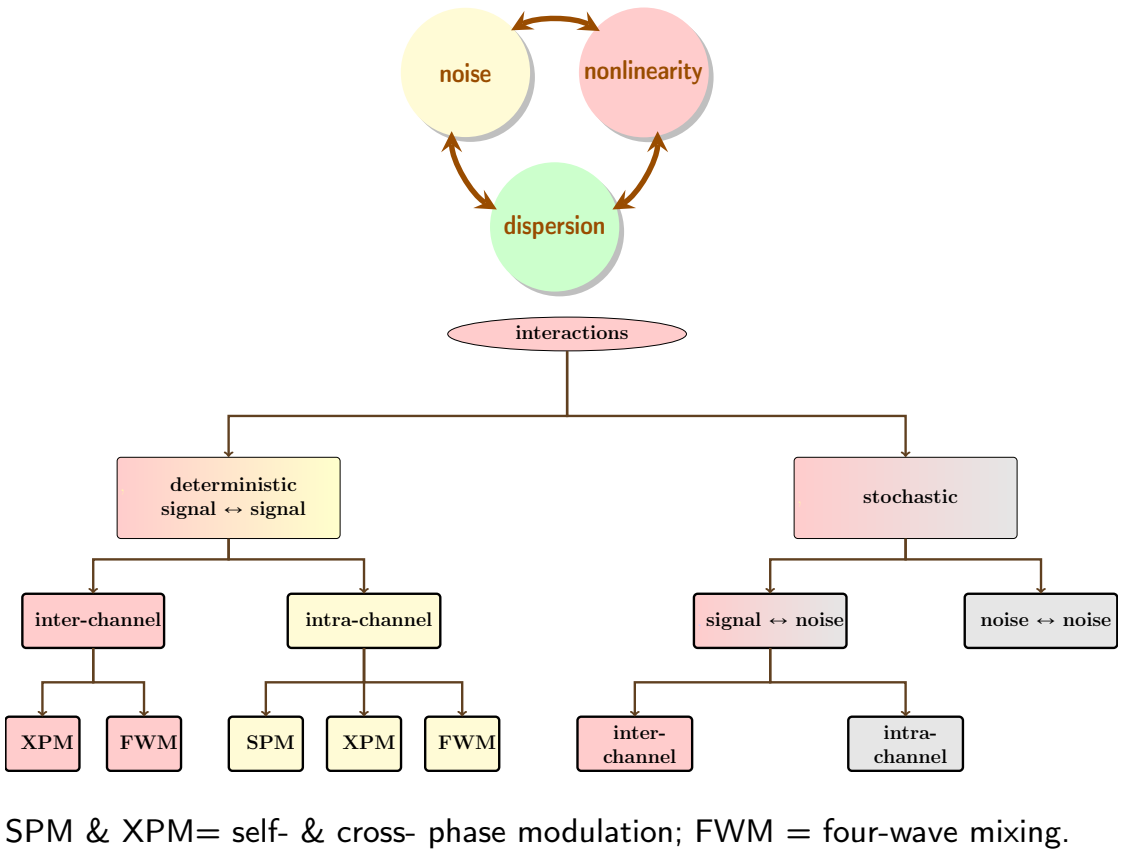


Figure 1.14: Decomposition of the NLI components [13].

quency components originate from the same channel. In this case, the complex interplay of the channel's own frequency components leads to phase modulation and frequency generation within the channel itself. Due to the non-zero nonlinear refractive index, the polarization state of a given WDM channel can be modified by the power in other channels of the WDM system, called cross-polarization modulation (XPolM). The process for signal-noise interaction is similar to signal-signal interaction except that one frequency component originates from the noise spectrum. The interaction between amplified spontaneous emission (ASE) noise and the Kerr effect leads to random rotation of the received signal constellation. This phenomenon is known as the Gordon-Mollenauer effect [91]. Particularly in long-haul fiber optical communication links that utilize cascaded EDFA, the system experiences the addition of Nonlinear Phase Noise (NLPN) [92].

Collectively known as nonlinear phase noise, these NLI contributions can lead to modulation instability, particularly in the presence of intra-channel interactions. Modulation instability results from the interplay between nonlinear effects and Chromatic Dispersion (CD), producing spectral sidebands that contribute to nonlinear phase noise [93]. Nonetheless, in typical WDM system configurations, XPM and FWM are the most dominant nonlinear impairments due to the large number of interfering channels, with FWM being particularly challenging to mitigate due to its

complex generation mechanism [94].

The CNLSE presented in Eq. (1.13) presents significant challenges due to its non-integrability, meaning that it lacks an analytical solution for precise Nonlinear Interference (NLI) modeling. As a result, the SSFM method is typically employed for numerical solutions. However, SSFM requires substantial computational resources and can be quite complex for extensive simulations. To overcome these limitations, noise models have been proposed in recent years, enabling us to approximate the NLSE using models like the Gaussian Noise (GN) model [95–97] or the Enhanced Gaussian Noise (EGN) model [98–100] which introduces a dependence on the modulation format (kurtosis $\widehat{\mu}_4$ and sixth-order moment $\widehat{\mu}_6$ of the amplitude level) and further hints at kurtosis as a performance metric. These models provide a pragmatic balance between computational complexity and accuracy, offering a viable approach to estimating the impact of NLI in optical communication systems. GN model and EGN model were designed for independent identically distributed (iid) signaling over single polarization transmission. Recently, an advanced 4D-nonlinear model for Polarization Division Multiplexing (PDM) systems has been proposed in [101], which applies to PDM-4D modulation and non-iid signaling (such as practical probabilistic constellation shaping schemes). This model represents a significant step forward in accurately representing the complex interactions in dual-polarization systems and offers the potential for improved performance analysis in optical communication systems employing advanced modulation schemes.

The fundamental concept of the GN model is to treat the NLI as a Gaussian process. This model specifically focuses on the Power Spectral Density (PSD) of NLI generated by the nonlinear term in the NLSE. While the GN and EGN models will not be introduced further here, it is important to understand the classification of NLI terms within these models. By comparing the model-specific classification with the general classification previously introduced, we gain a clearer understanding of NLI and its various manifestations in optical communication systems. In [96], an alternative taxonomy for Nonlinear Interference (NLI) in optical communication systems was proposed, dividing NLI into three distinct categories based on their interaction dynamics: Self-Channel Interference (SCI), Cross-Channel Interference (XCI), and Multi-Channel Interference (MCI), which tend to manifest as additive Gaussian noise and nonlinear phase rotation (NLPR), differing primarily in their spectral origin. Specifically,

- SCI: This type of NLI perturbs a given channel and is produced by that channel itself. It represents the nonlinear effects that occur within the same channel, such as SPM.

- XCI: This form of NLI perturbs a given channel due to nonlinear interactions between that channel and one other channel. It includes XPM, where the presence of another channel affects the phase of the signal of interest.

- MCI: This category of NLI perturbs a given channel resulting from nonlinear interactions involving that channel and two other channels, or by three entirely distinct channels. It is typically associated with complex interaction phenomena like FWM.

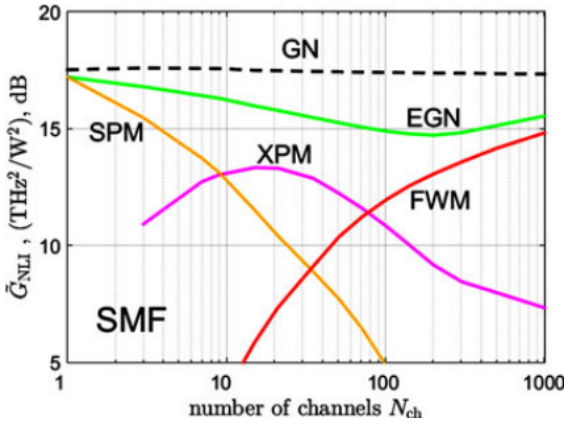


Figure 1.15: The normalized average NLI noise Power Spectral Density ($\widetilde{G}_{\mathrm{NLI}}$) over the center channel versus the number of channels (N_{ch}) for a WDM system with a fixed total bandwidth of 504 GHz. PDM-QPSK with quasi-Nyquist shaping, characterized by a roll-off factor of 0.05 and a channel spacing set at 1.05 times the symbol rate. The NLI is measured across 50 spans of SSMF with each span being 100 km in length. The contributions of SPM, XPM, and FWM components making up the EGN model curve are shown. Source: [14].

To better understand the variations of the three effects in fibers, we illustrate in Fig. 1.15 from [14], the changes in SCI, XCI, and MCI for an SSMF link as the number of channels increases within a fixed total bandwidth. The tested configuration is described in the caption of the figure. This fixed-total-bandwidth configuration leads to a decrease in the rate of each individual channel with an increasing number of channels on the x-axis. Notably, for ease of expression, we can consider SCI, XCI, and MCI to be roughly equivalent to SPM, XPM, and FWM respectively. At high symbol rates, SPM is predominant. As the symbol rate decreases, SPM gradually diminishes, XPM first rises and then falls, and FWM continuously increases. Intermediate values of symbol rates present a complex interplay of these three effects. According to the predictions of the EGN model, the optimal symbol rate corresponds to the lowest point of NLI influenced by these three effects. The GN model, due to its assumption of signal Gaussianity, cannot accurately predict NLI variations and tends to overestimate NLI. Conversely, the EGN model effectively incorporates FWM and provides an accurate representation of system behavior across the entire range of symbol rates.

This result also reveals to us the importance of symbol rate optimization (SRO) in multicarrier systems. As further discussed in App. A, SRO can be effectively achieved using Digital Multi-Band (DMB) technology to mitigate NLI [102]. However, the specific SRO strategies for different

fiber types, baud rates, and system configurations vary. Therefore, it is imperative to further investigate and understand the SRO for different modulation formats across various system conditions.

Beyond SRO, alternative methods were explored to mitigate or embrace nonlinear effects such as the use of new formats like the nonlinear Fourier transform, as outlined in [103, 104], or the implementation of equalizers for nonlinear effects, such as digital backpropagation [105], or employing perturbation analysis [106] to mitigate NLI using DSP. However, the practical application of low-complexity versions of these algorithms in next-generation optical systems offers limited performance enhancement.

In addition to NLI mitigation techniques, there has been a quest to identify new estimators of the power of the NLI generated by a given symbol sequence. The evolution of these estimators ranges from the initial focus on signal energy variance and Peak-to-Average Power Ratio (PAPR) to more sophisticated measures such as kurtosis and the 6th-order moment utilized in the EGN model. However, as signals become more optimized and baud rates increase, the accuracy of these simple indicators tends to diminish. Recently, the Energy Dispersion Index (EDI) [107] and windowed kurtosis [108] have been proposed as windowed indicators. These take into account the influence of surrounding signals on the central signal, providing a more accurate estimate of the NLI intensity. Yet, they still fall short of the precision offered by NLI calculations using numerical methods such as SSFM. In Chap. 5, we will detail these new metrics and introduce a novel windowed metric named D-EDI that incorporates dispersion effects. This new metric promises to deliver a more precise estimation of the NLI-generation capability of a given sequence, enhancing our ability to predict and mitigate NLI.

1.6 Digital Signal Processing Tools at The Receiver Side

At the receiver end which is shown at the bottom of Fig. 1.7, the optical channel under test is filtered and detected using a dual-polarization coherent receiver. Initially, according to the Shannon-Nyquist criterion, the received signal is sampled at a rate of at least twice its bandwidth which is two samples per symbol (SPS), and then its total power is normalized. Chromatic Dispersion (CD) is then compensated using a Finite Impulse Response (FIR) filter or in frequency domain designed to counteract the chromatic dispersion accumulated during transmission. Subsequently, the Carrier Frequency Offset (CFO) is compensated in two stages: initially through a coarse CFO estimator before synchronization and then refined with a fine CFO estimator after synchronization, as detailed in [109, 110]. Following coarse CFO compensation, the signal could be interpolated to recover the transmitter clock. Common algorithms employed include Godard's tone clock [111], Gardner's scheme [112], and Mueller-Müller [113].

Then, the 2×2 Multiple-Input Multiple-Output (MIMO) channel is equalized to separate the polarization tributaries. This step typically involves data-aided channel estimation in frequency or

time domains using constant-amplitude zero-autocorrelation CAZAC sequences [57, 58], or fully blind equalization such as the time-domain Constant Modulus Algorithm (CMA) [114] for M-PSK modulations, or even genie-aided channel knowledge, followed by Minimum-Mean-Square-Error (MMSE) equalization of the received signal. While CMA-based equalization is the most widely used technique in current systems using QPSK and squared-QAM modulation formats, the use of data-aided channel estimation and equalization are gaining ground with the introduction of new geometrically shaped or probabilistically shaped multi-dimensional formats.

Finally, Carrier Phase Estimation (CPE) is performed to mitigate the laser phase noise that impacts the signal's integrity. This is accomplished using the Viterbi–Viterbi Phase Estimation (VVPE) algorithm for M-PSK modulations [115, 116] or a Blind Phase Search (BPS) algorithm for M-QAM modlations [117], or also a pilot-aided phase-estimation algorithm applicable for all modulation formats, as referenced in [118, 119]. In our approach, we primarily utilize the following CPE receiver starting with a known sequence and then moving to a decision-directed equalization mode [120]. For each time index k, and each polarization tributary, the angle between the decided symbol $\hat{x}_k$ and the received symbol y_k is estimated. This angle includes both the correlated laser phase noise ϕ_k as well as additive interference and noise components n_k^{NLI} and n_k^{ASE}. It is crucial to note, however, that we can partially mitigate correlations within n_k^{NLI} through this CPE method.

The estimated angle $\widehat{\phi}_k$ is computed as a moving-window average over w preceding symbols:

$$\widehat{\phi}_k = \measuredangle \sum_{l=1}^{w} \hat{x}_{k-l}^{*} y_{k-l}. \tag{1.22}$$

Here, x^* represents the complex conjugate of x. We are allowed to use all y_k values, but only past x_k values. This calculation method aligns with the chain rule of mutual information between input and output sequences. The laser phase noise and XPM-induced phase noise affecting the received symbol y_k is then corrected through a phase rotation:

$$y'_k = y_k \cdot \exp(-j\widehat{\phi}_k). \tag{1.23}$$

Finally, the achievable rate is calculated from the compensated symbols y'_k, using circular Gaussian statistics on a symbol-by-symbol basis. This method can be changed to genie-aided or pilot-aided CPE by assuming that we use an ideal sequence or known pilots as x_k to estimate the phase.

This suite of sophisticated techniques at the receiver side ensures that the signals are accurately detected and processed, maintaining the transmission quality required for reliable communications. Finally, the received symbols are detected and decoded by hard-decision decoding or

soft-decision decoding of the FEC code. Several performance metrics, listed below, can be numerically computed to evaluate the transmission performance:

- Electrical SNR ($\mathrm{SNR}_{\mathrm{elec}}$) defined as the signal-to-noise ratio computed over 2D symbols at the output of the DSP chain;
- Bit error rate (BER) by counting bits in error and dividing over the total number of transmitted bits;
- Mutual information (MI) computed in a symbol-wise fashion and not accouting for the bit-mapping;
- Generalized mutual information (GMI) or the bit-wise mutual information that takes into account the bit mapping.

1.7 Forward Error Correction Codes

In the realm of DSP, forward error correction (FEC) plays a pivotal role in enhancing system performance [121]. FEC achieves this by adding redundancy in the form of parity bits to the binary information sequence, thus maximizing transmission reliability and safeguarding against errors. This addition allows the system to maintain the requisite rate for reliable transmission while reducing the SNR requirements. The net coding gain (NCG) is a crucial FEC performance metric, representing the amount of SNR reduction provided by a specific FEC scheme. According to Shannon's noisy channel coding theorem, there is a theoretical performance limit for FEC known as channel capacity. Under ideal conditions, FEC can facilitate data transmission with arbitrarily high reliability as long as the information rate remains below this channel capacity.

The essential purpose of FEC is to incorporate redundancy into the binary information sequence emanating from the source encoder. Post-transmission over a noisy channel, the FEC decoder leverages this redundancy to accurately reconstruct the source information. The code rate, R_c, is a fundamental parameter defined as the ratio of information bits (K_c) to the total number of bits (N_c) post redundancy addition:

$$R_c = \frac{K_c}{N_c}. \tag{1.24}$$

Within this framework, the FEC encoder processes blocks of K_c information bits, appending $N_c - K_c$ redundant bits. The FEC overhead is thus defined as:

$$OH = \frac{N_c - K_c}{K_c} = \frac{1 - R_c}{R_c}. \tag{1.25}$$

It is important to note that while Shannon capacity and GMI derivations assume ideal FEC codes with infinite block lengths, real-world FEC codes have finite lengths, impacting the net coding gain.

Historically, a variety of coding schemes have been implemented, including Hamming codes, Reed-Solomon (RS) codes [122], and Bose–Chaudhuri–Hocquenghem (BCH) codes [123]. In contemporary high-capacity systems exceeding 100 Gb/s, Low-Density Parity-Check (LDPC) codes, typically with about 20% overhead, are prevalent. The soft decoding LDPC (SD-LDPC) code ihas gained traction for its ability to handle nearly arbitrary long block lengths with manageable complexity, achieving NCG close to theoretical limits and operating within less than 1 dB SNR gap to the Shannon capacity [124, 125]. In our work, we use SD-LDPC from the DVB-S2 standard [126].

1.8 Achievable Information Rates

The typical requirements for the Bit Error Rate (BER) after FEC decoding are stringent, often less than 10^{-15}. This level is significantly lower than ones directly observed through simulations or actual testing, making direct measurement challenging. Consequently, the feasibility of achieving post-FEC BER $< 10^{-15}$ must be inferred from pre-FEC data. For Hard Decision (HD) FEC systems, the FEC threshold serves as a metric for measurement. This threshold is defined as the maximum allowable pre-FEC BER at the FEC input, assuming sufficient interleaving to disrupt potential error bursts.

However, the advent of Soft Decision (SD) FEC decoding in fiber-optic communications has made traditional FEC thresholds less accurate [127], necessitating more precise performance metrics. Information-theoretic quantities such as Mutual Information (MI), Generalized MI (GMI), or Normalized GMI (NGMI) are now being used as more accurate predictors of post-FEC BER [128]. Among these, GMI and NGMI are particularly noted for their accuracy in predicting post-FEC BER across various constellation sizes, especially when the system employs uniform square Quadrature Amplitude Modulation (QAM) [129]. NGMI, as a normalized form of GMI, is intuitive when comparing modulations with different spectral efficiencies and has become a key performance indicator.

Recent research has broadened the applicability of GMI, showing that it can serve as a reliable FEC threshold not only for uniform QAMs but also for Probabilistic Constellation Shaping (PCS) QAMs, which are being utilized in cutting-edge research works [130–132]. This demonstrates the importance of GMI as a tool for evaluating the performance of modern optical communication systems. By providing more reliable estimates of post-FEC BER, GMI, also referred to as **the achievable information rate (AIR)** aids in designing and optimizing these systems, ensuring that

they meet the high standards required for effective and accurate data transmission. We introduce in the following the basic knowledge for using these information-theoretic metrics for a scalar memoryless channel. The analytical calculation of GMI for an optical link is practical when the nonlinear interference noise adheres to a well-defined statistical distribution, typically Gaussian. In our system, the GMI calculation is conducted by comparing the transmitted signal and the received signal at the end of the DSP chain. Given that the noise at this stage closely resembles AWGN (quasi-AWGN), we base our GMI calculations on the assumption of an AWGN channel [121]. However, in broader applications where these specific conditions may not be present, GMI is often estimated empirically. This empirical approach relies on analyzing the received symbol measurements, which can be sourced either from real-world system operations or from detailed propagation simulations.

1.8.1 Channel Capacity

Channel capacity denoted as C, represents the maximum rate at which information can be reliably transmitted over a communication channel. This pivotal concept was first articulated by Claude Shannon in [133], laying the groundwork for modern information theory. According to Shannon's theorem, for communication to be deemed reliable, the transmission rate R of information must not exceed the channel capacity — symbolically, $R \leq C$. In essence, channel capacity serves as the upper limit for the rate of error-free information transfer through a given channel under specific conditions, dictating the feasibility and efficiency of communication systems.

The channel capacity is inherently determined by the statistical properties of the channel. For the simplest form of the channel, the additive white Gaussian noise (AWGN) channel is given by:

$$Y = X + N. \tag{1.26}$$

where X and Y are complex random variables representing the channel input and output respectively, and N is a complex circular Gaussian random variable of zero mean, accounting for noise and variance $N_0/2$ per real dimension. The AWGN channel transforms the input following the channel conditional probability:

$$p_{Y|X}(y|x) = \frac{1}{\pi N_0} e^{-\frac{|y-x|^2}{N_0}}. \tag{1.27}$$

Here, x and y are realizations of X and Y respectively. The mutual information, I, measures the maximum amount of information that can be transported over the channel [90]:

$$\begin{aligned} I(X;Y) &= \int\int p_{XY}(x,y)\cdot\log_2\left(\frac{p_{Y|X}(y|x)}{p_Y(y)}\right)dxdy \\ &= \int\int p_{XY}(x,y)\cdot\log_2\left(\frac{p_{XY}(x,y)}{p_X(x)\cdot p_Y(y)}\right)dxdy. \end{aligned} \tag{1.28}$$

where $p_{XY}(x,y)$ is the joint probability, and $p_X(x)$ and $p_Y(y)$ are the marginal probability density functions of X and Y. The channel capacity is the maximum mutual information, maximized over all possible input distributions $p_X(x)$, i.e., $C = \max_{p_X(\cdot)} I(X;Y)$. For the AWGN channel, Shannon demonstrated that the optimal source distribution maximizing C is itself complex and Gaussian:

$$p_X(x) = \frac{1}{\pi P} e^{-\frac{|x|^2}{P}}. \tag{1.29}$$

where P is the signal power. Therefore, the channel capacity is given by the celebrated Shannon capacity formula:

$$C = \log_2(1+\mathrm{SNR}). \tag{1.30}$$

measured in bits per channel use, where $\mathrm{SNR} = \frac{P}{N_0 B}$ is the signal-to-noise ratio, and B is the bandwidth. This formula represents the maximum achievable information rate for a given SNR over the AWGN channel.

1.8.2 Mutual Information with Discrete Modulation Constraint

Shannon's formulation of channel capacity is based on the premise of infinite input and output symbols, X and Y, and assumes that the input and output distributions, $p_X(x)$ and $p_Y(y)$, are continuous Gaussian functions. This theoretical model allows for the determination of the maximum rate at which information can be reliably transmitted over a communication channel, known as the Shannon capacity.

However, in real-world optical communication systems, the symbols X and Y are finite due to the discrete nature of practical modulation formats, and their distributions often deviate from the ideal Gaussian model. These practical constraints lead to a more realistic concept known as modulation-constrained capacity. This capacity is typically lower than the Shannon capacity, reflecting the reduction in performance due to the finite size of signal constellations and the non-Gaussian nature of signal distributions. The modulation-constrained capacity provides a more accurate measure of what is achievable in actual systems, accounting for the practical limitations inherent in real-world implementations.

Considering a finite symbol-constellation $X = \{a_1, a_2, \ldots, a_M\}$ of size M, the constrained capacity can be estimated using Monte-Carlo time averaging over a long-enough sequence of N

symbols:

$$I'(X;Y) = H(X) - H(X/Y) = H(X) - \frac{1}{N}\sum_{k=1}^{N} \log_2 \frac{p_X(x_k)p(y_k|x_k)}{\sum_{x\in A} p(y_k|x)p_X(x)} \tag{1.31}$$

Here, x_k and y_k represent the k-th symbol of the transmitted and received symbol sequences respectively and $H(X)$ is the source entropy defined as:

$$H(X) = -\sum_{x\in A} p_X(x)\log_2 p_X(x) \tag{1.32}$$

Finally, $H(X/Y)$ is the conditional entropy of X knowing Y. Eq. (1.31) corresponds to the maximum information rate achievable for a given discrete constellation. This limit is attainable in systems where the channel coding and constellation mapping are jointly designed and where the demapping and decoding are jointly processed, a scheme referred to as coded modulation.

1.8.3 Generalized Mutual Information

To mitigate the complexity associated with coded modulation in optical communication systems, Bit-Interleaved Coded Modulation (BICM) is often employed as a preferred technique. BICM distinguishes itself by separating the processes of coding and mapping, thereby allowing for a more flexible and practical system design. Within BICM systems, the bits that have been coded through FEC are interleaved. This interleaving process is crucial as it helps to disperse potential burst errors that might occur during transmission, significantly improving the robustness of the communication link.

Once interleaved, these bits are then grouped into blocks and mapped onto modulation symbols that will be transmitted over the optical channel. On the receiver side, the system needs to recover the original information from these symbols. To facilitate this, soft information about the coded bits is extracted from the demodulated symbols. This information, typically in the form of bit-wise a posteriori probabilities or log-likelihood ratios, is then relayed to the decoder. This process is known as Bit-Metric Decoding (BMD).

BMD is particularly effective in environments where the transmission channel can introduce errors in a non-uniform manner, as it provides a nuanced measure of reliability for each bit based on the received symbol. The flexible structure of BICM, coupled with the efficiency of BMD, makes it an attractive coding scheme for modern optical communication systems, balancing performance with complexity.

BICM can be viewed as m parallel, memoryless, and independent binary channels where $m = \log_2(M)$ for a constellation with M symbols. The maximum information rate of BICM is referred to as the Generalized Mutual Information (GMI), defined as [134, 135]:

$$GMI_{BMD} \approx H(X) - \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{m}\log_2 \frac{\sum_{x\in A} q_{Y|X}(y_k|x)p_X(x)}{\sum_{x\in A_{b_{k,i}}} q_{Y|X}(y_k|x)p_X(x)}. \tag{1.33}$$

Here, $q_{Y|X}$ is the auxiliary channel conditional probability, and $A_{b_{k,i}}$ is the set of N-dimensional constellation symbols whose i-th bit is $b_{k,i} \in \{0,1\}$. Employing a BICM scheme results in a capacity loss compared to the coded modulation-constrained capacity. However, adopting a Gray mapping when feasible, this loss is relatively small and acceptable for many practical applications.

As a final note, for the implementation of probabilistic constellation shaping (PCS), a Distribution matcher (DM) is employed to shape the distribution of transmitted symbols to enhance the system's efficiency and performance. The consideration of finite-length DM is vital as it closely represents practical implementations where the length of DM is constrained. The GMI is conventionally estimated under the assumption of infinite-length DM, where the rate loss due to the finite nature of practical DM is negligible. However, in real-world scenarios, the length of DM is finite, leading to a potential rate loss that needs to be accounted for.

The computation of GMI for a finite-length one-dimensional DM involves evaluating the information content per Amplitude Shift Keying (ASK) symbol for both infinite and finite lengths of DM. This process should consider the rate loss due to the finite nature of DM and adjust the GMI accordingly. By incorporating the DM rate loss into the conventionally estimated GMI, the actual performance of the system using finite-length DM can be assessed more accurately, reflecting the realistic capabilities and limitations of the communication system. Hence, The GMI for the finite-length DM is defined as:

$$GMI_{DM} = GMI_{BMD} - R_{loss}. \tag{1.34}$$

where R_{loss} will be introduced in Eq. (2.10) and the detailed derivation is shown in the appendix of [136]. We use Eq. (1.34) in the following chapters to compare the GMI of DMs that have different rate losses.

Summary

In this chapter, we embarked on a comprehensive journey through the evolution of optical fiber transmission systems, beginning with Intensity Modulation-Direct Detection (IM·DD) and culminating in the advanced realm of coherent communications. We recalled the building blocks of coherent communication systems, delving into the key aspects of each part. This exploration provides a thorough understanding of the technological advancements and methodologies that have shaped modern optical communications. The chapter unfolded the main layers of development, illustrating how innovations in technology and theory have driven the transition from basic

IM·DD to sophisticated coherent communication systems. Each section was dedicated to critical sub-systems, concepts and metrics, that will be used in the subsequent chapters. As we conclude this chapter, the next chapter shifts the focus toward the evolution of signal modulation. We will examine the progression of modulation techniques, from the earliest methods to the complex schemes that enable today's high-capacity, efficient optical communication systems. This continuation ensures a comprehensive understanding of both the infrastructure and the signal shaping and processing techniques that are fundamental to optical fiber communication.

Chapter 2

Advanced Signaling Technologies for Optical Communications

2.1 Introduction

This chapter delves into the rich history and development of signaling technologies in optical communications. It charts the journey from the earliest optical signals to the complex, high-speed methods employed in modern networks.

In the domain of optical communications, utilizing Single-Mode Fiber (SMF) and fully exploiting the C+L+S bands as discussed in [137, 138], could theoretically propel transmission capacities to extraordinary levels, potentially reaching or surpassing 100 Tb/s per fiber. Achieving such impressive capabilities necessitates the collaborative advancement of a range of sophisticated technologies. This includes the development and integration of Multi-Band Optical Amplifiers, Multi-Band Reconfigurable Optical Add-Drop Multiplexers (ROADMs), effective algorithms for fiber impairment compensation, advanced coding and modulation schemes, advanced signal processing techniques, meticulous end-to-end system optimization and control, and the implementation of high-bandwidth modulators and receiver devices. The convergence of these advanced technologies is key to unlocking the full potential of optical communication systems. In this chapter, we are particularly interested in the part on advanced modulation schemes in coherent transmission systems.

Using the Mach-Zehnder Modulator (MZM), it is possible to modulate symbols with distinct amplitudes and phases. By solely altering the phase, a 2-bit/4D-symbol Polarization Division Multiplexing Binary Phase-Shift Keying (PDM-BPSK) [139] can be achieved, featuring two constellation points with different phases per polarization tributary. Increasing the number of phases further allows for more complex modulation formats like 4-bit/4D-symbol PDM-Quadrature Phase-Shift Keying (QPSK) or even higher-order Phase-Shift Keying (PSK) modulations. However, PSK do

not fully exploit the potential of modulation techniques. By modifying both phase and amplitude, more advanced formats such as 6-bit/4D-symbol PDM-8QAM and 8-bit/4D-symbol PDM-16QAM, or even higher Quadrature Amplitude Modulation (QAM) levels, can achieve higher spectral efficiencies. By employing higher-order modulation formats, which improve spectral efficiency, the actual bit data transmission rate can be significantly increased within the same bandwidth. For instance, at a 100G baud rate, the simplest BPSK can achieve a data rate of 100 Gbit/s. In contrast, a 256QAM format can boost the data rate to an impressive 800 Gbit/s. This example underscores the transformative potential of advanced modulation techniques in enhancing the throughputs of optical communication systems. Yet, with the increase in modulation order, the Minimum Squared Euclidean Distance (MSED) of high-order modulations tends to decrease under constant average power conditions, leading to reduced noise tolerance. Additionally, the disparity in amplitude levels between constellation points grows, making the transmitted signal more prone to the generation of Kerr-induced nonlinear interference during fiber transmission.

Multidimensional modulation formats, such as the 2-bit/4D-symbol X-constellation proposed in [18], offer larger MSED compared to PDM-BPSK, achieving linear gains. Furthermore, the X-constellation and a following 3-bit/4D-symbol scheme, called PB-6B-8D [140] also strategically design the polarization states of 4D-symbols to eliminate cross-polarization modulation (XpolM), yielding considerable nonlinear gains in the Dispersion-Managed (DM) Large Effective Area Fiber (LEAF) systems compared to PDM-BPSK and Polarization-Switched QPSK (PS-QPSK) [141], respectively. However, in the widely deployed Dispersion-Unmanaged (DUM) Single-Mode Fiber (SMF) systems, these nonlinear gains tend to diminish due to the accumulated dispersion. We will design multidimensional modulation techniques that offer higher spectral efficiency over DUM systems in Chap. 3.

Over an AWGN channel, conventional PSK or QAM with uniformly distributed, equally spaced constellations are not optimal from a capacity point of view. Instead, modifying the probability and spatial distribution of constellation points through signal shaping can significantly enhance both the transmission distance and rate in high-speed optical communication systems. Signal shaping can generally be categorized into two types: geometric shaping (GS) and probabilistic constellation shaping (PCS). GS involves the use of constellation points that are uniformly probable but non-equidistant [142–145]. In contrast, PCS alters the probability distribution of constellation points, typically conforming to a Maxwell-Boltzmann distribution [1, 130, 146–149]. This adjusts the likelihood of different energy constellation points being used. Both methods are highly effective in approaching the Shannon limit of channel capacity. However, in scenarios where the number of constellation points is limited, PCS has shown to be more effective in closely approximating the Shannon limit compared to geometric shaping [150, 151]. Recent research efforts have also focused on optimizing the combination of both PCS and GS to achieve even better

performance [152, 153].

In section 2.2, we present the state-of-the-art of conventional constellation shaping. Subsequently, in section 2.3, we delve into a few of the first prominent multi-dimensional modulations tailored for NLI reduction and present first results of our preliminary tests. In section 2.4, we introduce the geometric shaping (GS). It is worth noting that we will summarize a comparison of various performance metrics for different multi-dimensional GS constellations in Tab. 3.1 of the next chapter. Finally, we present the basics of distribution matching for PCS in section 2.5.

2.2 Conventional Constellation Shaping

Light, as an electromagnetic wave, possesses multiple degrees of freedom that can be exploited for encoding information in optical communications. The mathematical description of electromagnetic waves (electric fields) is shown in Eq. (2.1). The electric field of these electromagnetic waves is mathematically described and comprises two orthogonal polarization components, E_X and E_Y. These components are crucial in Polarization Division Multiplexing (PDM), where they are used as two separate channels to transmit independent signals simultaneously. This approach effectively doubles the data transmission capacity of a single optical fiber.

$$\vec{\mathbf{E}} = \begin{bmatrix} E_X \\ E_Y \end{bmatrix} = \begin{bmatrix} A_X e^{j\Phi_X} \\ A_Y e^{j\Phi_Y} \end{bmatrix} e^{j(\omega \cdot t - k \cdot z))} = \begin{bmatrix} I_X + jQ_X \\ I_Y + jQ_Y \end{bmatrix} e^{j(\omega \cdot t - k \cdot z))} \tag{2.1}$$

In Wavelength Division Multiplexing (WDM), different frequencies (denoted as ω) of light are utilized as distinct channels. Each of these channels carries its own independent data stream at a specific frequency or wavelength. WDM significantly increases the total data transmission capacity of optical fibers. Modulating the amplitude A and phase Φ of each wavelength enables the definition of various modulation symbols used in schemes such as QPSK (Quadrature Phase Shift Keying) or its variant DQPSK (Differential Quadrature Phase Shift Keying) initially proposed along with a non-coherent receiver, QAM modulations and others. By altering the phase and amplitude, more data can be encoded into each light pulse, increasing the data transmission rate beyond what is possible with binary modulation techniques like OOK (On-Off Keying) or BPSK (Binary Phase Shift Keying).

These advancements in modulation technology allow for more efficient use of the available bandwidth in optical fibers, leading to higher data rates and improved performance in optical communication systems. This is especially important in applications requiring high data throughput, such as in backbone internet infrastructure or in long-distance communication links like undersea cables.

The evolution of optical data transmission has mirrored that of electronic data transmission in many ways. Initially, it adopted the simplest and lowest-cost digital encoding schemes, such

as Return-to-Zero (RZ) or Non-Return-to-Zero (NRZ) On-Off Keying (OOK), which used ideal rectangular sequences to represent 1 (on) and 0 (off), compatible with Directed detection and easier to implement (no DACs needed, cost-effective transmitters). However, this approach encountered limitations as transmission rates reached 40 Gbps. With the advent of 40 and 100 Gbps clock rates, the bandwidth used by OOK signals exceeded that of the 50 GHz channels defined by the International Telecommunication Union (ITU). This led to the signal's spectrum expanding and overlapping with adjacent channels. The signal degradation occurred due to crosstalk and bandwidth narrowing when passing through wavelength filters and due to important inter-symbol interference (ISI) from chromatic dispersion.

To overcome these challenges, the transition from OOK to PSK or QAM schemes long employed in the wireless industry, became necessary for high-speed optical fiber transmission. These advanced modulation techniques, depending on the symbol clock rate, can reduce the required bandwidth, enabling higher data rates within the 50 GHz ITU channel grid. Additionally, these new approaches facilitate compensation for Chromatic Dispersion (CD) and Polarization Mode Dispersion (PMD) through digital signal processing combined with coherent detection. Dispersion, caused by different frequencies and polarization states of optical waves propagating at varying speeds, can lead to pulse broadening and signal degradation, a significant issue in long optical fibers.

With coherent detection and digital equalization techniques, there is no need for optical PMD compensation, polarization control or dispersion compensating fibers, eliminating their associated losses and delays. Furthermore, complex-valued modulation can be paired with techniques like Polarization Division Multiplexing (PDM). In PDM, a second orthogonally polarized optical signal carries independent information over the same fiber, effectively doubling the transmission speed without needing an additional fiber. Compared to OOK, QPSK maintains the same symbol rate but doubles the transmission rate as each symbol encodes 2 bits. Adding PDM can double the transmission rate again, allowing QPSK with PDM to transmit four times the number of bits at the same symbol clock rate. Additionally, the use of pulse shaping filters can reduce the spectrum further, achieving more than 100 Gbps transmission rates in a 50 GHz channel.

2.2.1 Quadrature Phase Shift Keying

In optical communication, QPSK offers significant advantages over simpler schemes like OOK and BPSK, achieving 40 Gb/s in [154], and further improving to 100 Gb/s in [48]. QPSK represents 2 bits per symbol by having four constellation points, each with a phase that is $\frac{\pi}{2}$ apart from the others while maintaining a constant amplitude as shown in Fig. 2.1. This configuration allows QPSK to effectively double the data rate compared to OOK and BPSK without increasing the bandwidth. In other words, it can deliver the same data rate as OOK or BPSK but uses only

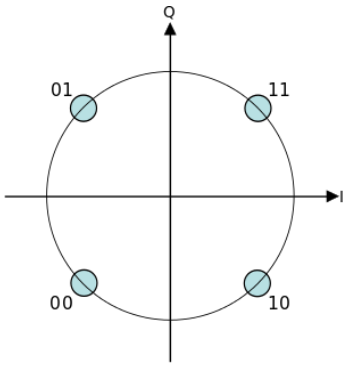


Figure 2.1: Construction of QPSK. Source: [15].

half the bandwidth. However, phase detection in QPSK requires coherent detection techniques, which are more complex than the direct detection used in OOK and BPSK.

PSK constellations can theoretically be designed with a varying number of phases. While it is possible to use any number of phases in PSK constellations, the need to accommodate binary data typically dictates the use of a number of symbols that is a power of two. This approach allows each symbol in the constellation to represent an integer number of bits, aligning with the binary nature of the data being transmitted. As a result, PSK constellations in practical communication systems are often designed with 2, 4, 8, or possibly 16 phases, with each phase shift corresponding to a distinct symbol representing a unique combination of bits. However, in practical applications, 8-PSK often represents the upper limit for high-order PSK constellations. When the number of phases in a PSK constellation exceeds eight, the error rate tends to increase significantly, making the system less efficient. In such cases, it may be more advantageous to switch to more advanced modulation schemes, such as Quadrature Amplitude Modulation (QAM), which offer better performance.

2.2.2 M-Quadrature Amplitude Modulation

In advanced optical communication systems, modulation schemes like Amplitude-Shift Keying (ASK) and Phase-Shift Keying (PSK) can encode information simultaneously in both amplitude and phase. These techniques can be combined with other multiplexing methods to significantly increase the data transmission rate.

Quadrature Amplitude Modulation (QAM) is a prime example of such a scheme. It combines amplitude and phase modulation to achieve higher data rates, with capabilities reaching 400 Gbps or more. QAM signals have a more intricate modulation and demodulation process compared to simpler schemes, making them more expensive to implement. However, the trade-off is that higher-order QAM constellations, which are more spread out compared to pure PSK constellations (like BPSK or 8-PSK), are less susceptible to noise and distortion. This spread results in a lower Bit Error Rate (BER).

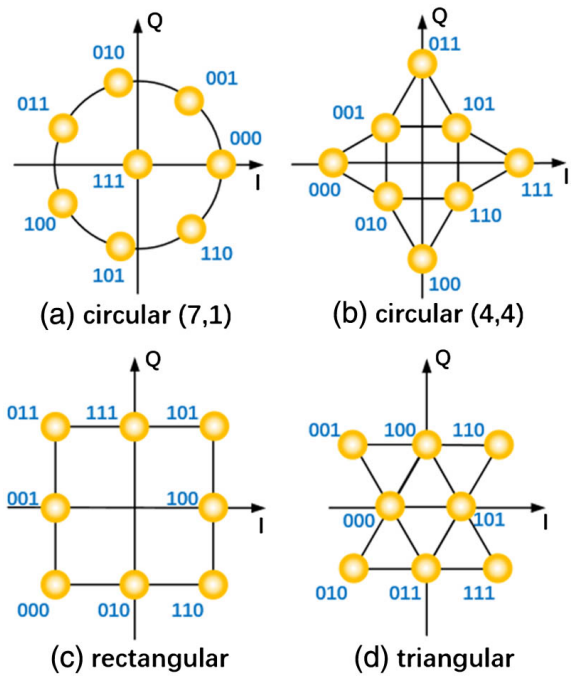


Figure 2.2: Construction of 8QAM. Source: [16, 17].

In 2^m-QAM schemes, $M = 2^m$ constellation points represent a group of m bits. These points are typically arranged in a square grid. For example, in the simplest form, 2-QAM (equivalent to BPSK), each symbol represents one bit. The amplitude between the constellation points for 1 and 0 remains constant, but there is a phase difference of π, making 2-QAM essentially identical to BPSK. In 4-QAM (similar to QPSK), there is one amplitude value and the phase difference between the four constellation points is $\frac{\pi}{2}$. In the realm of 8-QAM constellation shaping, Thomas [17] carried out a theoretical analysis of the performance of four uniquely shaped 8-QAM constellations, as depicted in Fig. 2.2. The central design principle was to distribute the constellation points across circles, regular triangles, squares, or similar regular polygons. Thomas's study primarily focused on the AWGN channel and concluded that the circular (7,1) constellation yielded the best performance. Building upon this foundation, Nolle and colleagues [155] expanded the research to encompass more complex channel scenarios, specifically tailored to the nuances of flexible optical fiber communications. They revisited the first three constellations illustrated in Fig. 2.2. Through experimental analysis, they corroborated that the circular (7,1) configuration remained the most effective in these contexts.

16-QAM is more popular due to its balance between spectral efficiency and BER. In this scheme, each symbol represents 4 bits, with the 16 constellation points arranged in a square grid. These points are typically encoded using Gray code, where adjacent points differ by only one bit as shown in Fig. 2.3. This encoding minimizes the impact of errors due to noise, as incorrect assignments of measured points to nearby points in the constellation will typically result in only

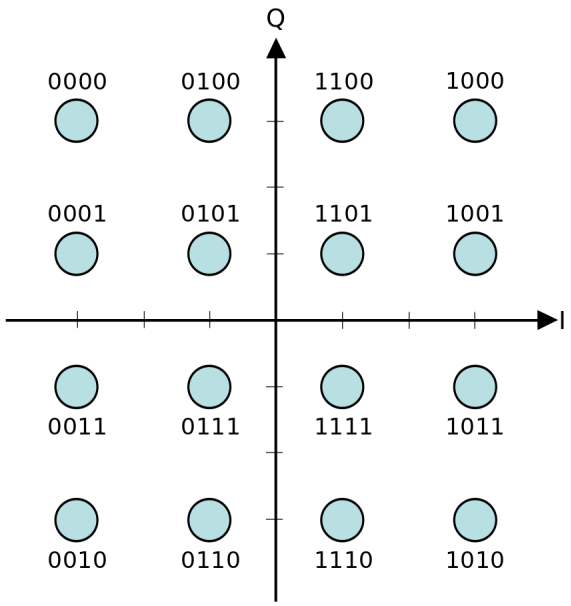


Figure 2.3: Construction of 16QAM. Source: [15].

a single-bit error. For a clearer comparison, we give the calculation formulas of the bit error probabilities for different modulations in Tab. 2.1 and show the comparison of BER versus the bit-level SNR equal to $\frac{E_b}{N_0}$ where E_b is the bit energy, for different modulations in Fig. 2.4. We clearly see that QAM with the same spectral efficiency has better performance than PSK.

Table 2.1: Bit error probability (exact or upper bounds) versus $\frac{E_b}{N_0}$ and $\frac{E_s}{N_0}$, where $m = \log_2(M)$

Modulation	bit error probability	bit error probability
BPSK	$0.5\text{erfc}(\sqrt{\frac{E_b}{N_0}})$	$0.5\text{erfc}(\sqrt{\frac{E_s}{N_0}})$
QPSK	$0.5\text{erfc}(\sqrt{\frac{E_b}{N_0}})$	$0.5\text{erfc}(\sqrt{\frac{E_s}{2N_0}})$
M-PSK	$\frac{1}{m}\text{erfc}(\sqrt{\frac{mE_b}{N_0}}\sin(\frac{\pi}{M}))$	$\frac{1}{m}\text{erfc}(\sqrt{\frac{E_s}{N_0}}\sin(\frac{\pi}{M}))$
M-QAM (m = even)	$\frac{2}{m}(1-\frac{1}{\sqrt{M}})\text{erfc}(\sqrt{\frac{3mE_b}{2(M-1)N_0}})$	$\frac{2}{m}(1-\frac{1}{\sqrt{M}})\text{erfc}(\sqrt{\frac{3E_s}{2(M-1)N_0}})$

As transmission demands continue to escalate, the quest for greater spectral efficiency becomes increasingly critical. In this context, higher-order modulation formats like 32-QAM and 64-QAM [156, 157] are gaining prominence in medium to long-distance transmission applications. 2^{32} QAM has even appeared in encrypted communications [158]. These advanced QAM formats offer significantly enhanced spectral efficiency. However, they also lead to a continual decrease in the Minimum Squared Euclidean Distance (MSED) and nonlinear tolerance. This trend indicates a trade-off where the benefits of higher spectral efficiency with higher-order QAMs come at the cost of greater susceptibility to nonlinear effects and a reduced margin for errors, factors that are crucial in the design and optimization of novel schemes for optical fiber communications. The

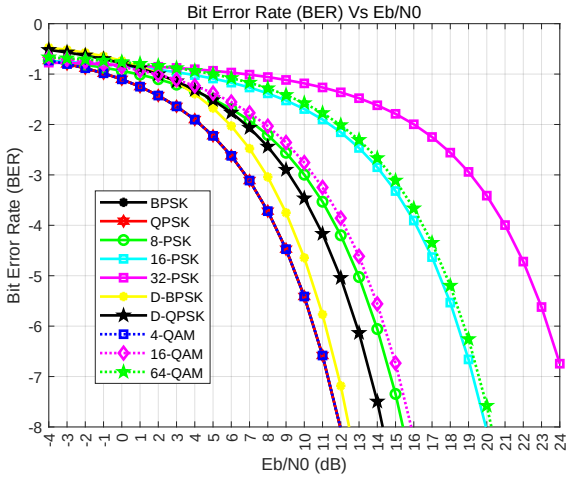


Figure 2.4: Comparison of the bit error rates of different modulation formats.

multi-dimensional shaping presented through selected schemes for optical fiber transmission in the next section can lead to further advancements.

2.3 Examples of Multi-Dimensional Shaping

2.3.1 The X-constellation

After successfully building our simulation system and validating its performance over an AWGN channel, we implemented and tested firstly the X-constellation proposed in [18] for dispersion-managed (DM) systems. X-constellation is a polarization-balanced format, built using 2D symbols from the QPSK alphabet arranged such that the polarization states are orthogonal in two consecutive time slots, in order to minimize the effect of a specific NLI manifestation which is cross-polarization modulation (XpolM)[1]. By imposing orthogonal polarization states in consecutive time slots for all WDM channels, XPolM can be minimized over two consecutive time slots. X-constellation symbols also have a constant modulus, which gives it more resilience to nonlinear effects over dispersion-managed systems. Fig.2.5 shows the structure of the X-constellation over two time slots A and B, along with the bit mapping and the Stokes representations S_A and S_B for each of the 16 8D constellation points. The basic idea behind X-constellation is to choose QPSK symbols on two polarization states and over two consecutive time slots (making it an 8D modulation). Hence, 256 combinations are possible. Among those, 64 combinations with opposite states of polarization (SOP) (diametrically opposed in Stokes space) in two consecutive time slots exist. The authors in [18] choose only 16 of these combinations to increase the MSED. Thanks to a higher MSED value, X-constellation gets a superior performance in the AWGN channel compared

[1] As a consequence of the non-zero nonlinear refractive index, the polarization state of a given wavelength-division multiplexed (WDM) channel can be modified by the power in other channels of the WDM system, a phenomenon called XpolM

Binary Value		0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
Slot A	x-pol	−1-i	−1-i	−1-i	−1-i	−1+i	−1+i	−1+i	−1+i	1-i	1-i	1-i	1-i	1+i	1+i	1+i	1+i
	y-pol	1+i	−1+i	1-i	−1-i	−1-i	1-i	−1+i	1+i	−1-i	1-i	−1+i	1+i	1+i	−1+i	1-i	−1-i
	$S_A = (S_1,S_2,S_3)$	(0,-4,0)	(0,0,-4)	(0,0,4)	(0,4,0)	(0,0,4)	(0,-4,0)	(0,4,0)	(0,0,-4)	(0,0,-4)	(0,4,0)	(0,-4,0)	(0,0,4)	(0,4,0)	(0,0,4)	(0,0,-4)	(0,-4,0)
Slot B	x-pol	1+i	−1+i	1-i	−1-i	1+i	−1+i	1-i	−1-i	1+i	−1+i	1-i	−1-i	1+i	−1+i	1-i	−1-i
	y-pol	1+i	−1-i	−1-i	1+i	1-i	−1+i	−1+i	1-i	−1+i	1-i	1-i	−1+i	−1-i	1+i	1+i	−1-i
	$S_B = (S_1,S_2,S_3)$	(0,4,0)	(0,0,4)	(0,0,-4)	(0,-4,0)	(0,0,-4)	(0,4,0)	(0,-4,0)	(0,0,4)	(0,0,4)	(0,-4,0)	(0,4,0)	(0,0,-4)	(0,-4,0)	(0,0,-4)	(0,0,4)	(0,4,0)
DOP	$\lvert S_A + S_B \rvert$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*The Stokes vectors for the two time slots are reported as S_A and S_B, respectively. Note that the cumulative Stokes vector $S_A + S_B$ is zero for all symbols.

Figure 2.5: The optical field Jones vectors for the two consecutive time slots (Slot-A and Slot-B) that define the 8-dimensional X-constellation symbols, and their corresponding binary symbol labels. Source: [18].

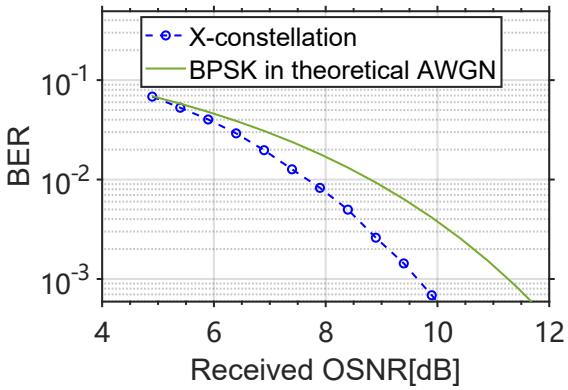


Figure 2.6: AWGN performance of X-constellation vs PDM-BPSK.

to PDM-BPSK which has the same spectral efficiency. In Fig. 2.6, our AWGN simulations show a gain of 1.5 dB for a BER of 10^{-3}, which is in line with the original paper.

Now, let us consider a nonlinear propagation. Since not only intra-channel but also inter-channel XpolM are reduced when the propagating signals are de-polarized through the use of X-constellation, the constellation should be more robust to the change of SOP of adjacent channels during WDM transmission. We test the performance gain in a nonlinear regime through numerical simulations. We compare the signal-to-noise ratios of the X-constellation to the ones of PDM-BPSK at the output of the channel equalization at the receiver side, denoted $\mathrm{SNR}_{\mathrm{elec}}$. The transmission signal consists of 5 WDM channels with 50 GHz spacing around 1550 nm, each modulated at 28, 35, or 40 GBaud respectively. All channels are pulse-shaped with a root-raised-cosine filter (RRC) with a roll-off factor equal to 0.1 which limits the spectrum to 1.1 times the baud rate. The tested fibers are LEAF or SSMF fibers with a chromatic dispersion coefficient of 4 or 17 ps/nm/km respectively.

We also study two dispersion management options, with or without inline dispersion compensation, named DM or DUM systems respectibely. For DM systems, we apply 5% pre-dispersion compensation, 90% dispersion compensation distributed at the end of each fiber span, and then

5% of the remaining dispersion is compensated at the receiver side. For DUM, all chromatic dispersion is compensated at the receiver side. Apart from CD, polarization mode dispersion is added with a PMD parameter equal to 0.04 ps/$\sqrt{\text{km}}$.

The link configuration is 20 spans of 80 km of fiber. In each span, an EDFA with a noise figure of 5dB compensates for the accumulated loss in the span. The SSFM implemented through the Manakov equation is used with 50 wave plates per span to simulate the polarization mixing [11]. Following propagation and coherent detection, we apply an RRC filter with the same roll-off factor to optimally match the transmitter filter. Then, after dispersion compensation, a CAZAC-aided channel estimation filter is used to compensate for the polarization crosstalk and PMD prior to de-mapping and bit error count.

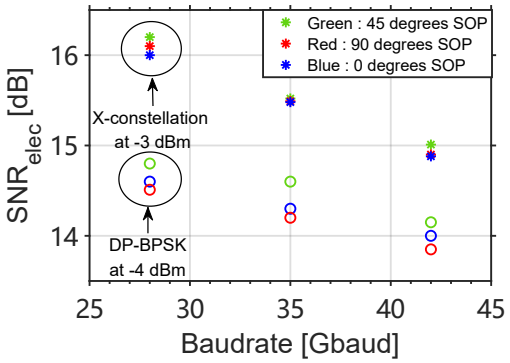


Figure 2.7: SNR_{elec} at the optimal power for different baud rates and SOPs of adjacent channels over a dispersion-managed system with LEAF fibers (20×80km spans).

First, we look into the results for the LEAF fiber with in-line dispersion management. In Fig. 2.7, we compare the fluctuations of SNR_{elec} for the optimal optical power value for 28, 35, and 42 GBaud, while changing the SOP of adjacent channels relative to the central channel. 0 degrees means aligned SOPs, and 90 degrees means that the signal on Pol. X is exchanged with the signal on Pol. Y. The performance of the central channel is shown in the figure. We see that PDM-BPSK has more fluctuations for different SOP values for all baud rates which means that it is more sensitive to the SOP change. On the other hand, X-constellation offers a larger SNR_{elec} and fewer fluctuations mainly from the elimination of XpolM. As the baud rate increases, the signal becomes more susceptible to dispersion effects, resulting in a rapid degradation of its nonlinear tolerance. This impact manifests in two significant ways. Firstly, it is more difficult to maintain constant power, leading to an elevation in SPM, which diminishes SNR_{elec}. Secondly, the XPolM elimination characteristic of X-constellation becomes more difficult to maintain effectively, leading to a diminishing gain relative to BPSK.

Then, we test the performance over SSMF fibers that have higher dispersion values than LEAF fibers. We see from the right part in Fig. 2.8 that the fluctuations at the optimal point with respect to

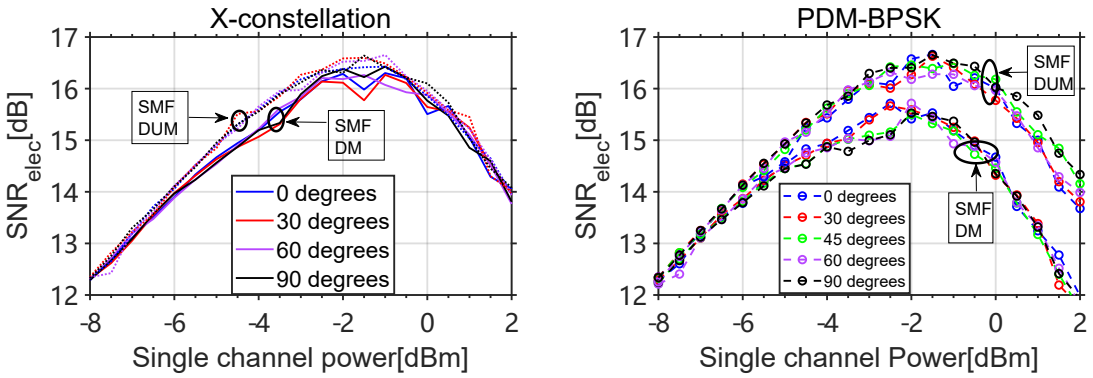


Figure 2.8: PDM-BPSK vs X-constellation in 35 Gbaud over SSMF fiber DM / DUM with different SOP of adjacent channel over 1600 km (20×80km spans) transmission

the SOP orientation of the adjacent channels vanish both in DM and DUM systems for 35GBaud PDM-BPSK. The SNR values are also higher than the ones obtained for the transmission over LEAF fibers. That is related to the mitigation of XpolM with the increased accumulated chromatic dispersion. However, from the left part of the figure, we see that the X-constellation still offers a 0.7 dB gain in SNR_{elec} at the optimal point and has a 1 dB gain in optimal power for DM systems. But in DUM systems, X-constellation loses the nonlinear tolerance advantage and shows a similar behavior as PDM-BPSK. Since the constant-modulus property of the X-constellation and the phase-matching favoring XPolM are completely removed by accumulated chromatic dispersion over 1600km transmission, the X-constellation only keeps the linear gain coming from a larger MSED over PDM-BPSK.

2.3.2 Extensions of Polarization-balanced schemes

In the 8D-QPSK (over 2 pol. and 2 consecutive time slots) alphabet containing 256 combinations, there are $128 = \log_2(7)$ combinations with different SOP over two consecutive time slots. This subset was called polarization-balanced 7 bits 8 dimensions (PB-7B8D) [140]. If we now apply a second screening and pick only the combinations of SOP that are orthogonal (equivalently opposite in Stoke space) in two consecutive time slots, we are left with $64 = \log_2(6)$ out of 128 and we call this set PB-6B8D [19]. Lastly, $32 = \log_2(5)$ out of the 64 combinations can be selected to increase the MSED and the obtained constellation is called PB-5B8D [140]. In the previous section, X-constellation selects only 16 out of the 64 combinations with opposite SOP. We focus, in this section, on the comparison of the nonlinear gains between the PB-6B8D and the polarization-switched QPSK (PS-QPSK), which both have a spectral efficiency (SE) of 3 bits/4D.

The bit mapping for PB-6B8D can be generated using Boolean equations. Each 8D symbol is labeled by an eight-bit vector $b_1, \cdots, b_8$. The first six bits are taken from the sequence of information, the last two are parity bits and are computed as in the equations in Fig. 2.9, where $\oplus$

$$b_7 = \overline{b_2} \oplus b_3 \oplus b_5 \oplus (b_1 \oplus b_2) \cdot (b_3 \oplus b_4 \oplus b_5 \oplus b_6) \oplus (b_3 \oplus b_4) \cdot (b_5 \oplus b_6)$$
$$b_8 = \overline{b_1} \oplus b_4 \oplus b_6 \oplus (b_1 \oplus b_2) \cdot (b_3 \oplus b_4 \oplus b_5 \oplus b_6) \oplus (b_3 \oplus b_4) \cdot (b_5 \oplus b_6)$$

Figure 2.9: Boolean equations to generate PB-6B8D, taken from [19]

and $\cdot$ denote binary additions and multiplications, respectively, and $\bar{b}$ denotes the logical negation of b.

PS-QPSK is a 4D constellation that uses one polarization to transmit a QPSK symbol (2 bits), and the choice of the polarization tributary is made with a third bit as shown in Tab. 2.2 [141]. Hence, only one polarization tributary carry a signal in each time slot.

Table 2.2: Symbol code book of PS-QPSK

	000	**001**	**010**	**011**	**100**	**101**	**110**	**111**
Pol. X	**Null**	**Null**	**Null**	**Null**	**1+i**	**1-i**	**-1+i**	**-1-i**
Pol. Y	**1+i**	**1-i**	**-1+i**	**-1-i**	**Null**	**Null**	**Null**	**Null**

PS-QPSK has the same performance as PB-6B8D over an AWGN channel as shown on the BER vs E_b/N_0 in Fig. 2.10 which matches the results in [20]. E_b is the energy per bit and N_0 is the variance of the Gaussian noise per complex dimension. Indeed, PB-6B8D has the same MSED and the same total number of pairs of neighboring symbols at MSED, denoted n_d, as PS-QPSK. On the right figure, the PB-6B8D called PB-PS-QPSK in ref [20], shows the same achieved mutual information as PS-QPSK.

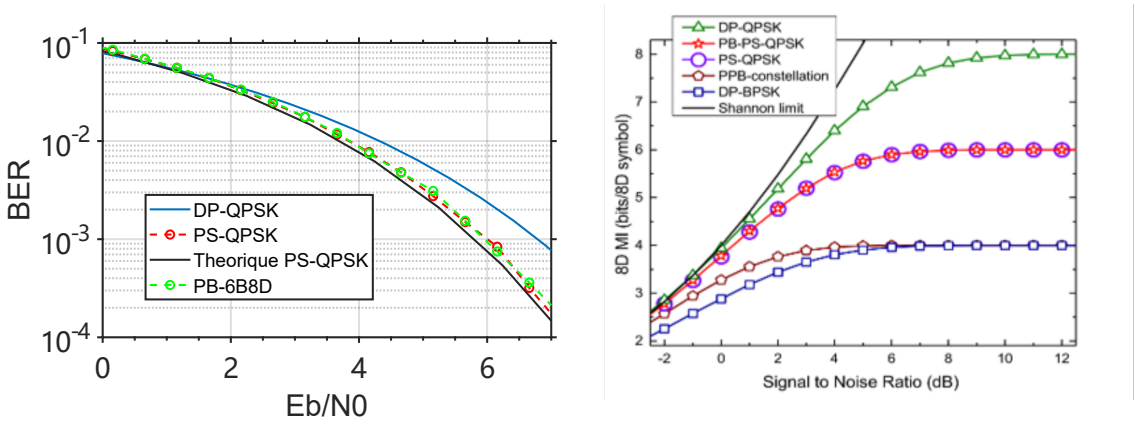


Figure 2.10: Validation of PS-QPSK and PB-6B8D over AWGN channel. Left: Our simulation. Right : Results in [20]

For the transmission over a nonlinear fiber link, we compare the electric signal-to-noise ratios $\mathrm{SNR}_{\mathrm{elec}}$ of the PB-6B8D to the ones of PS-QPSK. The transmission is made in the same conditions as the ones for the X-constellation in the previous section. In Fig. 2.11, we show the results

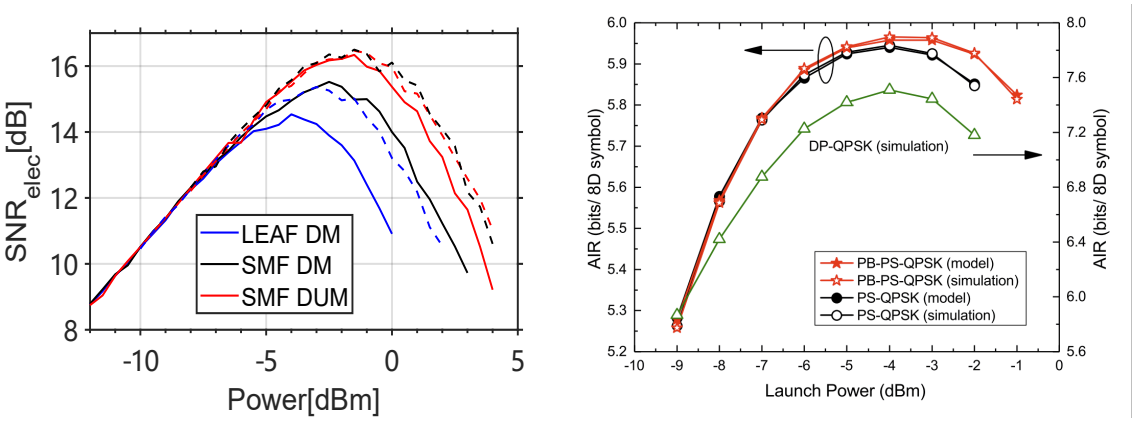


Figure 2.11: PB-6B8D vs PS-QPSK at 35 GBaud over SSMF or LEAF fiber for DM and DU. Link configuration: 20×80 km transmission with 5 WDM channels. On the left: our simulation results (Dashed line: PB-6B8D. Solid line: PS-QPSK). Right: Achievable information rate results from [20]

for a baud rate of 35 GBaud. Our simulation results show a 0.8 dB gain in SNR_{elec} and 1 dB gain in optimal power of PB-6B8D over DM systems with LEAF fibers which is in line with the demonstrated advantages in [20]. Over DM systems with SSMF fibers, PB-6B8D still gets 1 dB gain in SNR_{elec} and 1 dB gain in optimal power. Finally, for DUM systems with SSMF, unlike the X-constellation, PB-6B8D still holds a small gain at optimal power and a 1dB gain in SNR_{elec} in highly nonlinear regime.

To interpret these results, we look at the expected linear and nonlinear gains of each modulation format. From the Tab.3.1 in the next chapter that summarizes performance metrics of several multi-dimensional schemes, if we look at the rows corresponding to the studied modulations in this section, we see that even though both PB-6B8D and PS-QPSK have the same MSED and the same total number of 8D symbol pairs at MSED, PB-6B8D has no variations of the degree of polarization (DOP) within the 8D symbol (just like the X-constellation). However, it also has a lower kurtosis compared to PS-QPSK, which might explain the higher SNR in the nonlinear regime, a behavior that was not observed for the X-constellation over DUM systems.

2.4 Geometric shaping

In this section, we present an overview of Geometric Shaping (GS) in optical communications, a detailed discussion of its integration with multidimensional coding will be presented in Chap. 3. GS, a critical technique in enhancing the performance of long-distance optical communication systems, encompasses a spectrum of implementation methods. These include Geometrically Shaped-QAM (GS-QAM) [21], which boosts system capacity and mutual information performance. However, the complexity of GS-QAM implementation can be greater than traditional

modulation schemes (more stringent requirements on the linearity of devices and high-resolution DACs). Deep Learning for Geometric Constellation Shaping [159] has shown superior performance in reducing symbol error rates, but also demands significant computational resources and extensive volume of data for training.

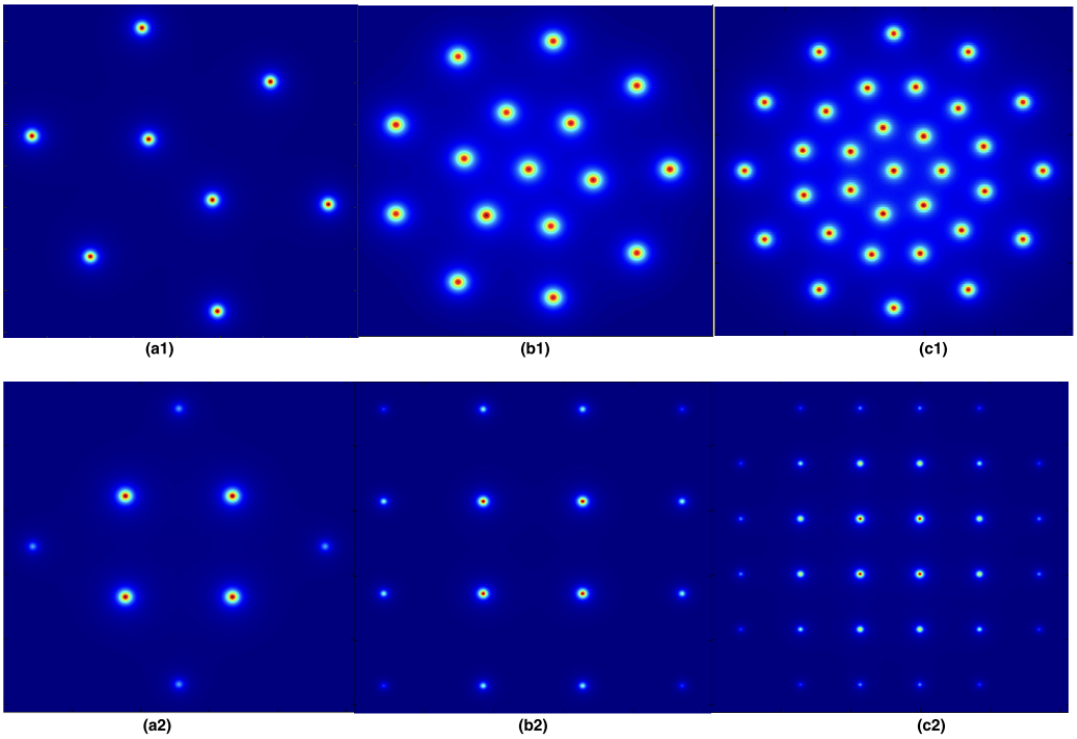


Figure 2.12: The constellation diagrams for (a1) GS-8QAM, (b1) GS-16QAM, (c1) GS-32QAM, and (a2) PCS-8QAM, (b2) PCS-16QAM, (c2) PCS-32QAM. Source: [21]

The optimization of GS constellations in optical communications is intricately linked to the used optimization criterion and the available SNR. Various criteria are available for optimizing these constellations, including but not limited to minimizing average symbol error probability, maximizing mutual information, and minimizing mean-square error in the representation of Gaussian sources. A typical characteristic of these constellations is that points with lower amplitude are spaced more closely than those at the outer regions as shown in Fig. 2.12, reflecting a strategic approach to spatial distribution. For the GS-8QAM/16QAM/32QAM, the symbols are uniformly distributed but the position of each symbol is optimized. Compared to GS, conventional PCS keeps the symbols at their initial positions, but the distribution of all the symbols is transformed into a Maxwell-Boltzmann (MB) distribution.

In the field of GS-QAM formats, several methods have been proposed to achieve shaping gains. Notably, earlier works introduced formats such as the generalized cross constellation [160], which have been instrumental in advancing GS techniques. More recent developments in this area include, but are not limited to, iterative polar modulation [161], generalized mutual information

(GMI)-optimized QAM formats [162], and multidimensional constellations that employ a multi-sphere distribution [163].

2.5 Probabilistic Constellation Shaping

Unlike GS which consists in finding the target constellation point position distribution and then maps it to bit streams, in PCS we are committed to making the output symbol sequence reach a target probability distribution. As proved in [164], the desired symbol probability distribution in the additive white Gaussian noise (AWGN) channel under a fixed average energy constraint is a Maxwell-Boltzmann (MB) distribution. The distribution matcher (DM) is a major part of achieving the desired distribution. DM involves transforming independent sequences of uniformly distributed symbols into sequences that conform to a specific target distribution. These sequences represent discrete memoryless sources (DMS) and are designed to be invertible, allowing the original input symbols to be accurately recovered. The dematcher plays a key role in this process, reversing the DM process to restore the original input symbols.

2.5.1 Different Types of Distribution Matching

Distribution matching can be categorized into three distinct types, based on the variability of the lengths of the input and output sequences:

- Variable-to-Fixed (v2f): In this type, the length of the input sequence is variable, while the output sequence length remains fixed. This approach addresses challenges associated with variable transmission rates but introduces complexities in implementation [165].
- Fixed-to-Variable (f2v): Here, the input sequence has a fixed length, but the output sequence length varies. This method can be particularly useful in systems where the transmission rate is not constant, but it may pose synchronization challenges due to the variable length of the output [166, 167].
- Fixed-to-Fixed (f2f): In fixed-to-fixed length DM, both the input and output sequences have fixed lengths. This type often utilizes block codes, which can simplify synchronization issues that arise in v2f and f2v types. However, f2f DM may not always guarantee error-free recovery of the input sequence. To improve the reliability of the process, implementing a one-to-one mapping in f2f DM is recommended [22].

Each DM type has its own set of challenges and complexities. For instance, variable-to-fixed and fixed-to-variable DMs must address issues related to transmission rates and synchronization, whereas fixed-to-fixed length DMs, though avoiding synchronization problems, might struggle

with ensuring accurate recovery of the input sequence. In the context of PCS in optical communications, the use of f2f DM is quite prevalent due to its effectiveness in matching specific target distributions. This approach is instrumental in optimizing the performance of communication systems by tailoring the distribution of transmitted symbols to better suit the channel's characteristics. Various methods of f2f DM have been developed and are widely used in PCS, including Constant Composition Distribution Matching (CCDM), Enumerative Sphere Shaping (ESS), and Shell Shaping, etc.

- Constant Composition Distribution Matching (CCDM), as introduced by Schulte and Böcherer in [22], focuses on creating sequences with a constant composition, ensuring that each symbol occurs with a predetermined frequency.
- Enumerative Sphere Shaping (ESS), proposed by Amari et al. in [149] for optical communications and further developed by Gültekin et al. in [24], involves mapping input bits onto a sphere in a high-dimensional space, optimizing the energy efficiency of the transmitted signals.
- Shell Shaping, as discussed in works by Dar et al. and Geller et al. in [168–170], targets the optimization of signal shapes within a defined geometric structure or 'shell' to improve transmission efficiency.

2.5.2 Theory and Example of f2f DM

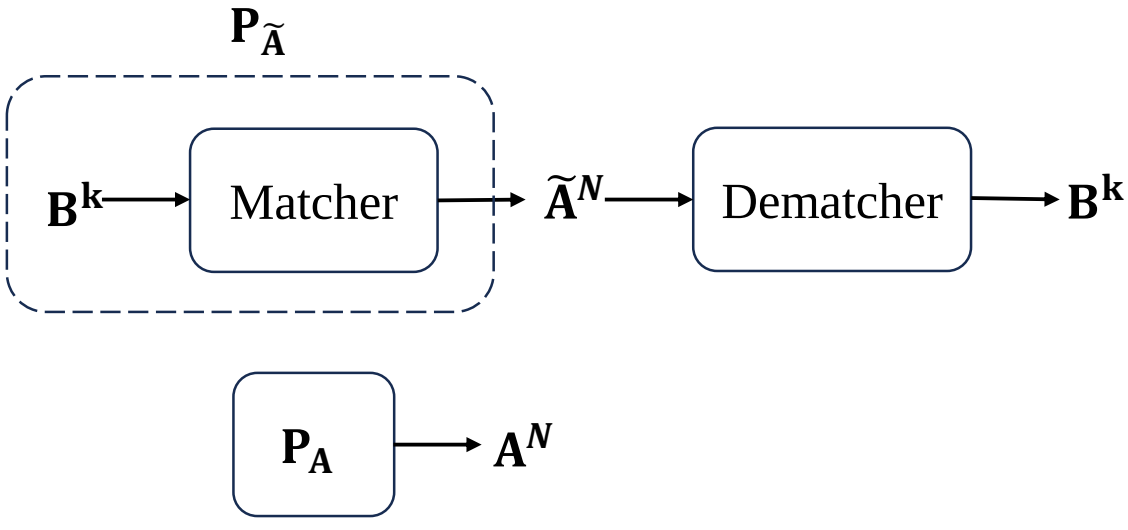


Figure 2.13: Matching a data block $\mathbf{B^k} = \{B_1 \cdots B_k\}$ to output symbols $\tilde{\mathbf{A}}^{\mathbf{N}} = \{\tilde{A}_1 \cdots \tilde{A}_N\}$ and reconstructing the original sequence at the dematcher. The rate of the matcher is $\frac{k}{N}$ expressed in bits/ output symbol. The matcher can be interpreted as emulating a discrete memoryless source $P_A(a)$, $P_{\tilde{\mathbf{A}}}$ is the output distribution of this mapping. Source: [22].

In this section, we will introduce the theory and give an example of f2f distribution matching. As shown in Fig. 2.13, in PCS, the bits in the input sequence, denoted as $\mathbf{B^k} = \{B_1 \cdots B_k\}$, are typically independent and follow a Bernoulli distribution with a probability of $\frac{1}{2}$. In the f2f DM process, a one-to-one invertible function f is used for mapping. This function f takes the k Bernoulli-distributed bits $\mathbf{B^k}$ and maps them onto strings of length N, $\tilde{\mathbf{A}}^{\mathbf{N}} = f(\mathbf{B^k})$, which are elements of the set A^N. The output distribution of this mapping, denoted as $P_{\tilde{\mathbf{A}}}$, closely imitates the desired distribution P_A.

This process is crucial in PCS as it allows for the efficient and tailored shaping of signal constellations to match specific channel conditions, thereby enhancing the overall system performance. The one-to-one nature of the function ensures that the mapping is reversible, allowing for accurate recovery of the original bit sequence through the inverse function f^{-1}. This reversibility is key to maintaining the integrity and reliability of the data transmission in optical communication systems.

In our analysis, we focus on 2^m-ary amplitude-shift keying (ASK) alphabets as the 1D dimension representation of square QAM symbols, represented as $X = \pm 1, \pm 3, \cdots, \pm(2^m - 1)$, This set can be decomposed into $X = S \cdot A$ where $S = \pm 1$ and $A = \{1, 3, \cdots, 2^m - 1\}$ denote the sign and amplitude alphabets, respectively, with $|A|$ being the cardinality of the amplitude alphabet.

Considering the symmetric nature of the capacity-achieving distribution for the AWGN channel around zero, our focus is primarily on the amplitude distribution $P_A(a)$ (which leads to a probabilistic amplitude shaping). We assume that the sign distribution $P_S(s)$ is uniform and independent of the amplitudes, leading to the channel input distribution $P_X(x) = P_S(s)P_A(a)$.

For ASK constellations under an average power constraint, the optimal distribution for maximizing the mutual information lacks a known analytical form for the AWGN channel. However, as indicated in [164, 171], Maxwell-Boltzmann (MB) distribution is commonly utilized for shaping amplitudes in this scenario:

$$P_A(a) = \begin{cases} K(\lambda)e^{-\lambda a^2} & , \quad \text{for } a \in A, \\ 0 & , \quad \text{otherwise.} \end{cases} \tag{2.2}$$

with λ dictating the probability distribution's variance and $K(\lambda)$ ensuring its normalization. Research studies [172] demonstrated that the MI gap between the MB distribution and the capacity-achieving distribution for ASK constellations is negligible.

To facilitate the generation of higher-order modulation formats like 2^m-ASK for $m \geq 2$, a binary labeling strategy is essential. Detailed discussions on binary labeling are available in [150], where it is posited that the binary label $B_1, B_2 \cdots B_k$ of a channel input X can be split into a sign bit B_1 and amplitude bits $B_2, B_3 \cdots B_k$. We use binary reflected Gray codes (BRGCs) for labeling, as defined in [150].

In the realm of distribution matching, the primary function is to map sequences of k bits to sequences of N-amplitudes in a reversible manner. The objectives of this process include (i) the creation of a shaping codebook $A^\star \subseteq A^N$ and (ii) the development of a shaping encoder for sequence indexing. The first objective pertains to defining the desired set $A^\star$, while the second involves the algorithmic aspects of the mapping process. The upcoming section will delve deeper into these concepts and the parameters associated with various shaping techniques.

The energy of an amplitude sequence $\mathbf{a^N} = (a_1, a_2, \ldots, a_N)$ is quantified as follows:

$$E(\mathbf{a^N}) = \sum_{i=1}^{N} a_i^2. \tag{2.3}$$

Representing N-sequences as points in an N-dimensional (n-D) space, the set $A^\bullet$ is defined as:

$$A^\bullet = \{\mathbf{a^N} : E(\mathbf{a^N}) \leq E^\bullet\} \tag{2.4}$$

comprising all sequences within or on the surface of an n-sphere with a squared radius of $E^\bullet$. The center of this sphere represents the zero-energy point. The composition of a sequence $\mathbf{a^N} \in A^N$ is expressed as $C = [n_1, n_2, \ldots, n_{|A|}]$, where n_j indicates the number of occurences of the j th element of the set A in $\mathbf{a^N}$. The number of n-sequences sharing the same composition C is given by the multinomial coefficient MC:

$$MC(C) = \frac{n!}{n_1! n_2! \cdots n_{|A|}!}. \tag{2.5}$$

For a set $A^\star$ with an average amplitude distribution $P_A(a)$ over A, the mean energy per symbol is given by:

$$E = \sum_{a \in A} P_A(a) a^2. \tag{2.6}$$

The shaping rate R_s of the set $A^\star$ is then defined as:

$$R_s = \frac{\lfloor \log_2 MC(C) \rfloor}{N}. \tag{2.7}$$

in bits per 1-Dimension. The input block length for a shaping algorithm indexing sequences from $A^\star$ is:

$$k = \lfloor \log_2 MC(C) \rfloor. \tag{2.8}$$

expressed in bits, where $\lfloor \cdot \rfloor$ is the floor operation. Typically, 2^k is less than $MC(C)$ for two main reasons: (i) $MC(C)$ is not usually an exact power of two, and (ii) k is often set as a variable in the range $1 \leq k \leq \lfloor \log_2 MC(C) \rfloor$ to facilitate rate adaptivity.

The parameters of a shaping code $A^\star$ satisfy the following inequality:

$$H(A) \geq \frac{\log_2 MC(C)}{N} \geq \frac{k}{N} \tag{2.9}$$

where the first inequality accounts for the finite block length N, and the second reflects the binary-input nature of the shaping algorithm. Here, $H(A)$ represents the entropy of a source with distribution P_A in bits. Both inequalities achieve equality when $N \to \infty$, indicative of asymptotically optimum amplitude shaping architectures. Finally, the rate loss R_{loss}, for a set $A^\star$ with average distribution $P_A(a)$, is defined as:

$$R_{\text{loss}} = H(A) - \frac{k}{N}. \tag{2.10}$$

expressed in bits per 1-Dimension.

2.5.3 Implementation of a PCS Transmitter

In [22, 23], an implementation of a PCS transmitter with M-QAM constellations is described including 1D distribution matchers and forward error correction (FEC) encoders. Figure 2.14 shows a block diagram of a 1D distribution matcher. The total net rate of the PCS CCDM transmitter with polarization division multiplexing is $R_T = 4(U_1 + U_2)/V_1 = 4R_{DM} + 8R_C - 4$ bits per 4D symbol where U_1 and U_2 are bit streams at the input of the transmitter (Here U_1 is equivalent to k, and v_1 is equivalent to N mentioned in Eq. (2.10). U_1 is used to better express the bit stream together with U_2). R_{DM} is the rate of the 1D distribution matcher. Each 1D DM encodes U_1 bits into $V_1 = U_1/R_{DM}$ symbols. $R_c = \frac{K_c}{N_c}$ is the rate of the embedded FEC encoder that encodes the bits coming from the U_2 bits and the Gray-mapped V_1 symbols. In the formula for the total net rate, for a practical implementation, the finite-length distribution matching will lead to a slight modification of U_1.

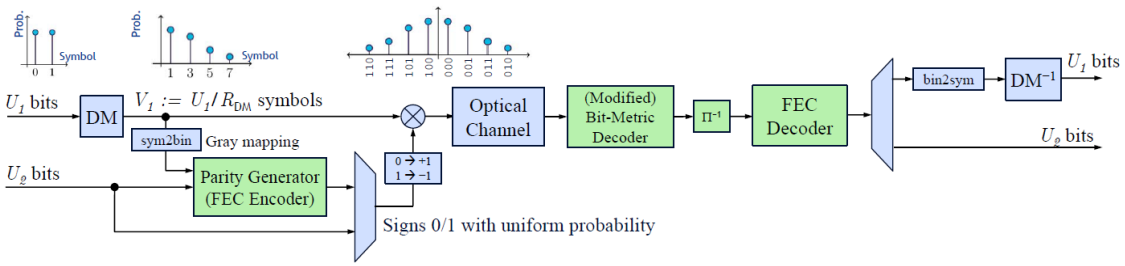


Figure 2.14: Block diagram of a PCS transmitter for a single dimension of the conventional PCS for polarization-division multiplexed systems. Source: [23].

In the next sections, we will present various techniques for implementing the 1D distribution matching.

2.5.4 Constant Composition Distribution Matching (CCDM)

In this section, we delve into the Constant Composition Distribution Matching (CCDM), a concept first introduced in [22] and later integrated into the amplitude shaping framework for PCS, as expounded in [171]. CCDM operates on the principle of constant composition, generating output

sequences that are permutations of a specific base sequence. This base sequence is defined by its composition C, representing the frequency of each amplitude's occurrence. The aggregate of these output sequences, or the cardinality of the shaping set $A^\circ \subseteq A^N$, is computed using the multinomial coefficient $MC(C)$, as outlined in Eq. (2.5). Each amplitude sequence within A° shares the same energy E°, positioning them on an n-shell characterized by a squared radius of E°, illustrated in Fig. 2.17. This consistent energy distribution is a hallmark of CCDM, pivotal for maintaining reliable performance in PAS applications. The performance of CCDM is optimal at infinite block length, where $N \to \infty$, allows for the realization of any desired distribution. However, CCDM faces challenges at short block lengths, primarily due to significant rate loss. It restricts the amplitude composition to a specific count of symbols for each possible amplitude, resulting in uniform per-block energy for all sequence permutations. This characteristic of CCDM, while advantageous in certain scenarios, may pose limitations in diverse communication contexts.

We provide two straightforward examples to elucidate the concept and coding mechanism of CCDM, focusing on scenarios involving sequences transformed from equal to unequal probabilities. These illustrative cases are designed to clarify the fundamental principles of CCDM. For an in-depth theoretical exploration of CCDM and its intricacies, readers are referred to the seminal works by Schulte and Böcherer, as detailed in [22, 171]. These references offer comprehensive insights into the underpinnings of CCDM, laying the foundation for understanding the examples presented here.

Example 1: Consider a 4-ASK constellation defined as $X = \pm 1, \pm 3$ with block length $N = 5$. The amplitude set is $\mathcal{A} = 1, 3$. Let us assume that the target distribution is:

$$P_A(1) = \frac{4}{5},\ P_A(3) = \frac{1}{5}. \tag{2.11}$$

Given that these probabilities are integer multiples of $\frac{1}{5}$, the corresponding class for a sequence of length $N = 5$ is:

$$\mathfrak{T}^5_{P_A} = (1,1,1,1,3), (1,1,1,3,1), (1,1,3,1,1), (1,3,1,1,1), (3,1,1,1,1). \tag{2.12}$$

where in each sequence, amplitudes 1 and 3 occur $n_1 = 4$ and $n_3 = 1$ times, respectively. A CCDM maps binary strings of length length k to sequences in $\mathfrak{T}^5_{P_A}$. with 5 sequences in $\mathfrak{T}^5_{P_A}$, we have:

$$k = \left\lfloor \log_2 \left| \mathfrak{T}^5_{P_A} \right| \right\rfloor = \lfloor \log_2 5 | \rfloor = 2. \tag{2.13}$$

As 5 is not a power of two, the sequence $(1,3,1,1,1)$ is eliminated. The following lookup table defines the considered CCDM mapping:

$$\begin{array}{ll} 00 \mapsto (1,1,1,1,3), & 01 \mapsto (1,1,1,3,1), \\ 10 \mapsto (1,1,3,1,1), & 11 \mapsto (3,1,1,1,1), \end{array} \tag{2.14}$$

This mapping effectively translates binary strings to amplitude sequences, thus providing a binary interface. Furthermore, the mapping is one-to-one, ensuring its invertibility.

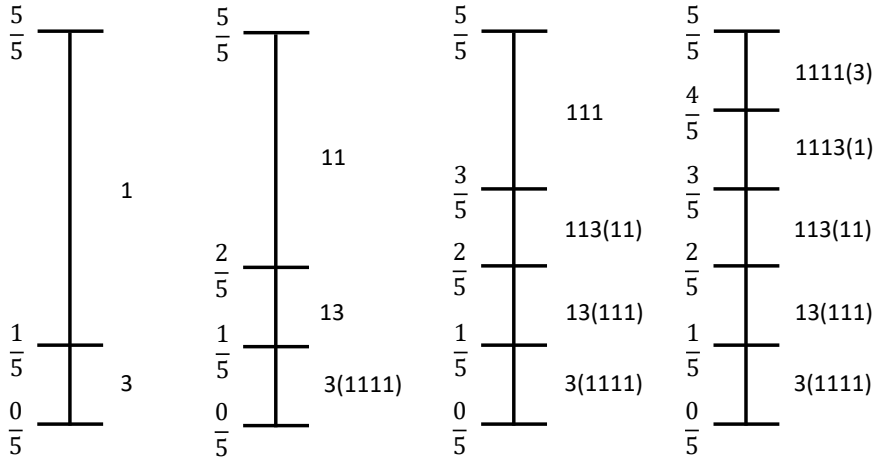


Figure 2.15: Refinement of the output intervals. Parentheses indicate symbols that must follow with probability one.

Example 2: In this example, we delve into a simple illustrative example of sequence indexing using arithmetic coding, particularly focusing on the encoding aspect as detailed in [22]. This example builds on the previous discussion and aims to shed light on how arithmetic coding refines output intervals. Fig. 2.15 serves as a visual aid in this explanation. Initially, our 'bag' for the encoding process contains four instances of the symbol '1' and one instance of the symbol '3'. The initial probabilities of drawing each symbol are $P_{\bar{A}_1}(1) = \frac{4}{5}$ and $P_{\bar{A}_1}(3) = \frac{1}{5}$, where $P_{\bar{A}_n}$ is the temporary probability of n^{th} step. Once a '1' is selected, the composition of the bag shifts to three '1's and one '3'. This alteration changes the drawing probabilities; specifically, the chance of drawing another '1' reduces to $P_{\bar{A}_2}(1) = \frac{3}{4}$, while the likelihood of selecting a '3' rises to $P_{\bar{A}_2}(3) = \frac{1}{4}$. This sequence of probability adjustments continues until all symbols are allocated. If a '3' is drawn at the outset, the remaining sequence is predetermined to consist of four '1's, in order to adhere to the desired output distribution. The symbols in parentheses within the figure represent those that are necessitated by the chosen distribution. This example underscores the dynamic nature of arithmetic coding, where the probability of selecting the next symbol is invariably influenced by the preceding selections, leading to varying probabilities at each step of the sequence.

Then, we illustrate the application of arithmetic coding for sequence indexing, as exemplified in Fig. 2.16. This example features four input sequences 00, 01, 10 and 11, each with equal probability, and five output sequences of equivalent likelihood. The input intervals are defined as [0, 0.25), [0.25, 0.5), [0.5, 0.75), and [0.75, 1). The output intervals, derived from the last column of Fig. 2.15, are set as $\left[0, \frac{1}{5}\right)$, $\left[\frac{1}{5}, \frac{2}{5}\right)$, $\left[\frac{2}{5}, \frac{3}{5}\right)$, $\left[\frac{3}{5}, \frac{4}{5}\right)$ and $\left[\frac{4}{5}, 1\right)$.

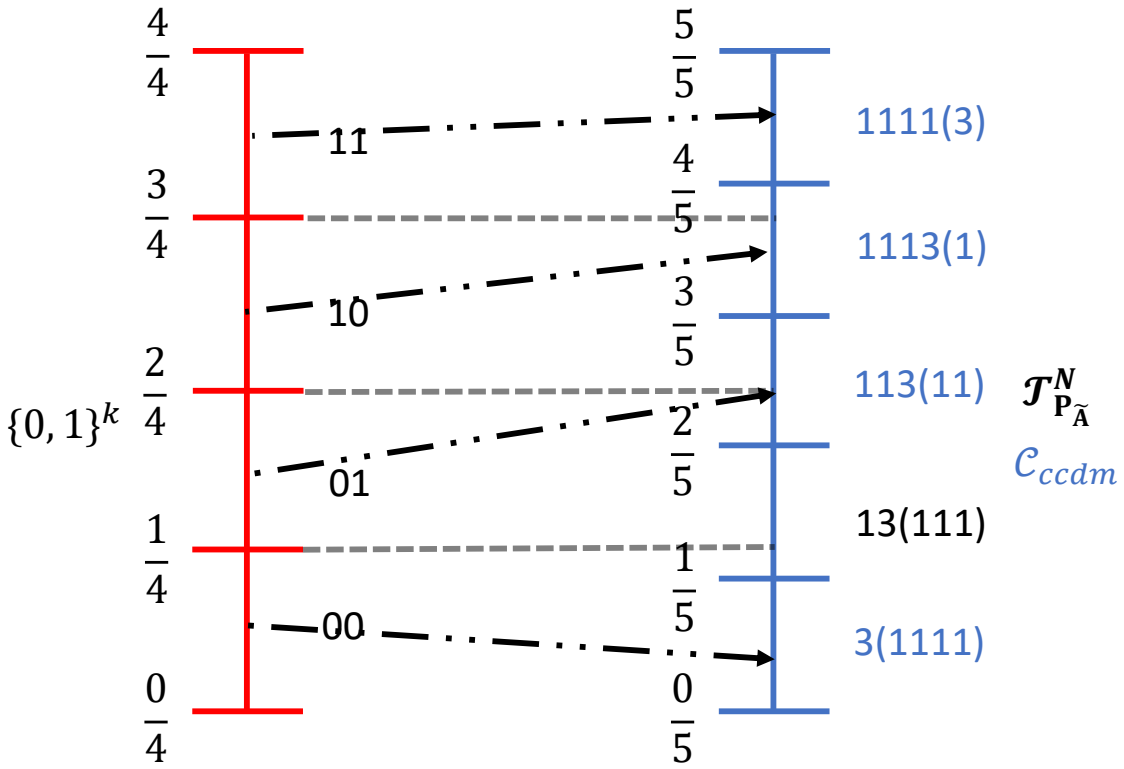


Figure 2.16: Diagram of a constant composition arithmetic encoder with $P_{\tilde{A}}(1) = \frac{4}{5}$, $P_{\tilde{A}}(3) = \frac{1}{5}$, $k = 2$ and $N = 5$.

The arithmetic encoder establishes a link between an output and an input sequence when the lower boundary of the output interval is within the range of the input interval. For instance, '00' may correspond to both '13111' and '31111', while '01' uniquely links to '11311'. The limitation of having at most two choices arises because the size of the input interval is less than twice the size of the output interval. Both choices remain valid, allowing for an inverse operation. In our implementation, the encoder selects the output sequence with the lowest interval boundary. Consequently, the codebook C_{ccdm} for this example includes {'11113', '11131', '11311', '31111'}.

Generally, the cardinality of C_{ccdm} is 2^k, with $2^k \leq |\mathfrak{T}^N_{P_{\tilde{A}}}| < 2^{k+1}$. The complete indexing of the set $\mathfrak{T}^N_{P_{\tilde{A}}}$ is feasible only when $2^k = |\mathfrak{T}^N_{P_{\tilde{A}}}|$. This analysis applies to all subsets $C_{\text{ccdm}} \subseteq \mathfrak{T}^N_{P_{\tilde{A}}}$. The specific subset included in the codebook is implicitly defined by the arithmetic encoder's operations.

2.5.5 Enumerative Sphere Shaping (ESS)

Sphere Shaping (SpSh) in the discrete domain [170, 173] can be adopted for amplitude shaping. In [174], it is shown that the average distribution over $A = \{1, 3, \cdots, 2^m - 1\}$ converges to an MB distribution as $N \to \infty$ when an N-spherical region of X^N is utilized as the signal space as

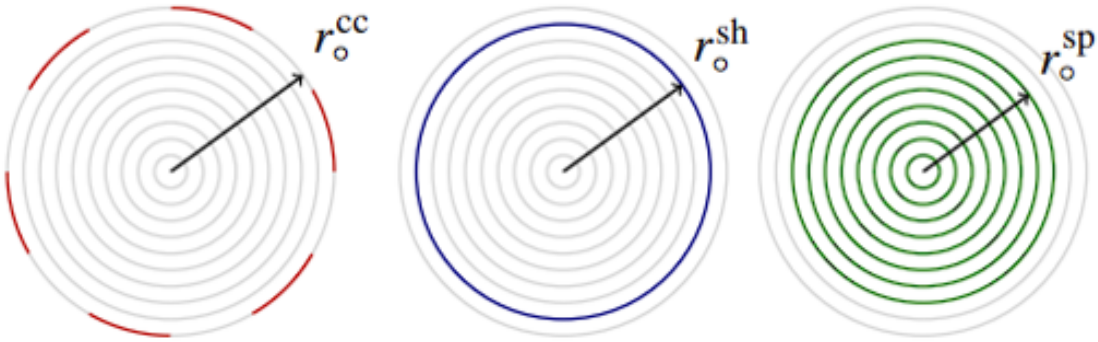


Figure 2.17: The illustration demonstrates three distinct shaping approaches applied to sequences in N-dimensional coding: constant composition, single-shell, and sphere shaping. Each approach is represented by a circle, symbolizing an N-dimensional shell. Within these shells, the darker areas signify the sequences that are actively utilized in each respective shaping method. Notably, when all three approaches encompass an identical number of sequences, there is a consistent relationship between their radii, denoted as $r_\circ^{cc}$, $r_\circ^{sh}$, and $r_\circ^{sp}$ for constant composition, single-shell, and sphere shaping, respectively. In every dimension, these radii satisfy the condition $r_\circ^{cc} \geq r_\circ^{sh} \geq r_\circ^{sp}$. This graphical representation provides a clear visual comparison of how each coding strategy occupies space within an N-dimensional framework, highlighting the differences in their sequence utilization. Source [24].

shown in Fig. 2.17. Furthermore, [175] indicates that at finite N, SpSh effectively minimizes the informational divergence between the average distribution and an MB distribution.

CCDM suffers from significant rate loss at short block lengths and requires long block lengths to achieve high shaping gain which makes its implementation more challenging. Enumerative sphere shaping (ESS) [24, 149] has been suggested as an alternative distribution matching technique with lower rate loss than CCDM at the same shaping rate. As shown in Fig. 2.18, if we look at the block length of 100, the rate loss of ESS is only one-fifth of the one of CCDM. ESS has a low rate loss in short block lengths because it uses all possible energy levels below a fixed maximum energy value as shown in Fig. 2.17. The performance of ESS and CCDM eventually converge as the block length grows.

Furthermore, in [149], the authors observed that finite block-length ESS and CCDM exhibit higher effective signal-to-noise ratios than their infinite block-length counterparts in the nonlinear transmission regime. It has been shown that for a 400 Gbps (Symbol rate 45 GBaud) dual-polarization 64-QAM WDM transmission system, ESS with short block-lengths outperforms both uniform signaling and CCDM. ESS with a block-length of $N = 200$ was shown to provide a reach extension of about 200 km in comparison with CCDM with the same block-length. As a last note, when assessing nonlinear gains, it is important to consider the interplay between the block length of the signal shaping technique and the capability of CPE algorithms in partially removing nonlinear distortions such as nonlinear phase rotations of the constellation points as observed

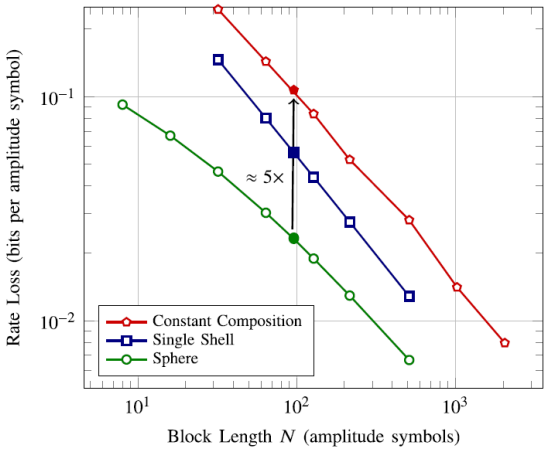


Figure 2.18: Comparison of constant composition and sphere codes in terms of rate loss. The target shaping rate is $R_s = 1.75$ bit/1D. The used 1D constellation is an 8-ASK. Source [24].

in [176–178].

Sphere coding is the process of constraining the coded sequences to be selected from within an N-dimensional sphere. In this case, the induced distribution in one real dimension converges to a Maxwell-Boltzmann distribution as $N \to \infty$. Hence, a set of bounded-energy, say E_{max}, amplitude sequences are used to create the shaped signaling. To index the selected sequences, ESS is used to lexicographically index the sequences as an ordered set. ESS can be graphically represented with a bounded-energy trellis in which each state represents an energy level and each branch designates an amplitude.

In the described trellis structure, each sequence $\mathbf{a^N} = (a_1, \cdots, a_N)$ is distinctly represented by a unique path, consisting of N branches. A branch connects nodes between adjacent columns, specifically from column $n-1$ to column n, symbolizing the n^{th} component a_n for $n = 1, 2, \ldots, N$. The nodes in the n^{th} column indicate the cumulative energy up to the first n dimensions, computed as:

$$E((a_1, \cdots, a_n)) = \sum_{i=1}^{N} a_i^2. \tag{2.15}$$

In Fig. 2.19, the energy values at each node are shown as black numerals. Every path originates from the zero-energy node in the bottom-left corner, culminating in a node in the final column, where $N = 4$ is the sequence's length. In our example, the cumulated energy will increase by 1 or by 9 for $a_n = 1$ or $a_n = 3$, respectively. The nodes in the terminal column ($n = N$) contain the possible sequence energies $E(\mathbf{a^N})$, with values ranging from $\{N, N+8, N+16, \ldots, E_{max}\}$, particularly with the amplitude set $\mathcal{A} = \{1, 3\}$. The energy increases with the minimum energy increment difference, so the energy at any node in column n is given by:

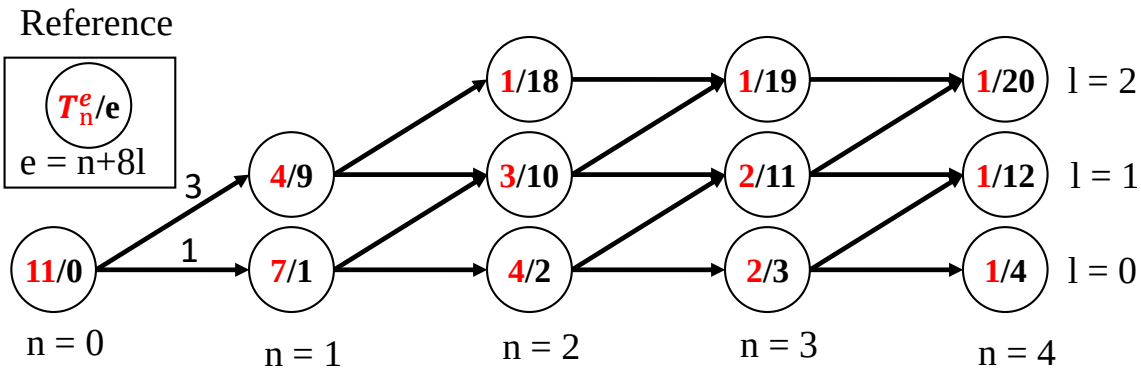


Figure 2.19: Bounded-energy enumerative amplitude trellis for $\mathcal{A} = \{1,3\}$, $N = 4$ and $E_{max} = 20$.

$$e = n + 8l. \tag{2.16}$$

Here, l ranges from 0 to $L-1$, with L defined as:

$$L = \left\lfloor \frac{E_{max} - N}{8} \right\rfloor + 1. \tag{2.17}$$

This represents the total number of energy levels in the final column, allowing the identification of a node by the pair (n,e). The red numbers in the figure reflect the count of paths leading to a final node at $n = N$, calculated recursively for $n = N-1, N-2, \ldots, 0$ and $0 \leq e \leq E_{max}$ as:

$$T_n^e = \sum_{a \in A} T_{n+1}^{e+a^2}, \tag{2.18}$$

starting with the last column initialized to ones (i.e., $T_e^N = 1$). Only states with energy levels e in the range between n to $E_{max} + n - N$ are considered in each column.

The trellis also serves for sequence enumeration within a sphere. For example, from node $(3,11)$, paths lead to nodes $(4,12)$ or $(4,20)$, using branches of amplitude 1 or 3 respectively. Hence, for node $(3,11)$, $T_3^{11} = 2$. From node $(1,1)$, paths lead to nodes $(2,2)$ or $(2,10)$, using branches of amplitude 1 or 3. Hence, for node $(1,1)$, $T_1^1 = 4+3 = 7$. At the zero-energy node, $T_0^0 = 11$ designates the total number of sequences represented in this trellis, which could encode $k = 3$ bits. Hence, for the toy example in Fig. 2.19, $R_s = 0.75$ bits/amplitude, $P_A(1) = \frac{7}{11}$, $P_A(3) = \frac{4}{11}$, $E_{av} = 4 \times (\frac{7}{11} \times 1^2 + \frac{4}{11} \times 3^2) = 15.6364$, and $R_{loss} = 0.95 - 0.75 = 0.2$ bits/amp, which are calculated using the following equations:

$$k = \left\lfloor \log_2 \left| T_0^0 \right| \right\rfloor \text{ (bits)}. \tag{2.19}$$

The rate of this ESS can then be expressed as

$$R_s = \frac{k}{N}(\text{bits/amp}) \tag{2.20}$$

The amplitude distribution $P_A(a)$ over $\mathcal{A}$ is calculated as [Lemma 5.1 in [179]]:

$$P_A(a) = \frac{T_1^{a^2}}{\sum_{b\in\mathcal{A}} T_1^{b^2}}. \tag{2.21}$$

The average energy of sequences, E_{av}, is:

$$E_{av} = N\sum_{a\in\mathcal{A}} P_A(a)a^2. \tag{2.22}$$

The calculation of ESS rate loss is different from that of CCDM, but the formula of ESS rate loss is also applicable to CCDM. We use A_{MB} to represent a random variable following an MB distribution, with its average symbol energy defined as:

$$N\sum_{a\in\mathcal{A}} a^2 P(A_{MB} = a) = E_{av}. \tag{2.23}$$

The rate loss is the rate difference between the MB distribution and the ESS for reaching the same average energy value:

$$R_{loss} = H(A_{MB}) - R_s. \tag{2.24}$$

When the block length is infinite, ESS achieves MB distribution, and the rate loss of ESS approaches 0.

Enumerative shaping and deshaping are key processes in this context. Shaping computes a sequence from a given index, while deshaping identifies the index of a sequence. The shaping process is given in Algorithm 1 and the deshaping process is summarized in Algorithm 2. These processes involve counting sequences and tracing paths in the trellis, as outlined by Cover's indexing formula [180].

We illustrate the encoding and decoding processes in ESS by Algorithm 1 and Algorithm 2 using the example in Fig. 2.19 as a reference. Consider an initial binary sequence $(1,1,0)$. The sequence is first converted into a decimal representation, $i_1 = 6$. We then navigate through the trellis tree from left to right, aiming to find the minimum energy amplitude at each node.

During encoding:

1. Compare i_1 with T_1^1. Since $i_1 = 6$ is less than $T_1^1 = 7$, the first output amplitude is determined as 1.

2. Moving to the second node, $i_1 = 6$ is greater than T_2^2, so the output amplitude is 3. We then ascend a node in the same column and compute the difference, $i_2 = i_1 - T_2^2 = 2$.

Algorithm 1 Enumerative Shaping

Ensure $0 \leq i < T_0^0$
Initialize $i_1 = i$
for $n = 1, 2, \ldots, N$ **do**
    for $b \in A$ **do**
        Choose $a_n = b$ such that

$$\sum_{b<a_n} T_n^{b^2+\sum_{j=1}^{n-1} a_j^2} \leq i_n < \sum_{b\leq a_n} T_n^{b^2+\sum_{j=1}^{n-1} a_j^2} \tag{2.25}$$

        if $n < N$ **then**
            Update

$$i_{n+1} = i_n - \sum_{b<a_n} T_n^{b+\sum_{j=1}^{n-1} a_j^2} \tag{2.26}$$

        end if
    end for
end for
Output the sequence $a_1, a_2, \ldots, a_N$

Algorithm 2 Enumerative Deshaping

Given $a_1, a_2, \ldots, a_N$:
Initialize the algorithm by setting $i_{N+1} = 0$.
for $n = N, N-1, \ldots, 1$ **do**
    for $b \in A$ **do**
        if $b < a_n$ **then**
            Update

$$i_n = \sum_{b<a_n} T_n^{b+\sum_{j=1}^{n-1} a_j^2} + i_{n+1} \tag{2.27}$$

        end if
    end for
end for
Finally output $i = i_1$.

3. For the third node, i_2 equals T_3^{11}, leading to an output amplitude of 3. Subsequently, $i_3 = i_2 - T_3^{11} = 0$.

4. Finally, i_3 is less than T_4^{20}, resulting in the last output amplitude being 1.

The resulting output amplitude sequence is $(1,3,3,1)$.

For decoding:

1. Starting with $i_5 = 0$, the fourth amplitude is identified as 1, setting $i_4 = 0+0 = 0$.

2. The third amplitude is 3, leading to $i_3 = i_4 + T_3^{11} = 2$.

3. The second amplitude is also 3, which updates $i_2 = i_3 + T_2^2 = 2+4 = 6$.

4. The first amplitude is 1, giving $i_1 = i_2 + 0 = 6$.

Finally, converting i_1 back to binary yields the original sequence $(1,1,0)$.

Summary

In this chapter, we provided a concise review of the progress in modulation and coding technologies within the realm of coherent optical communications. The chapter began with the basics of quadrature and amplitude modulation, exemplified by techniques like Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM). We then delved into enhancing their linear and nonlinear performance, leveraging the full potential of multi-dimensional flexibility. The narrative advanced to the transition from conventional constellation shaping to more sophisticated, tailored constellation shaping methods. We introduced the fundamental structure and techniques for Geometric Shaping (GS) and focused on basic structures and distribution matching techniques for Probabilistic Constellation Shaping (PCS). This exploration lays the groundwork for the forthcoming research findings in the next chapters, equipping the reader with the required knowledge and insights.

Chapter 3

Multi-Dimensional Shaping Designs

3.1 Introduction

This chapter provides an in-depth exploration of the evolution of multi-dimensional shaping techniques for NLI mitigation, a critical area in the field of communications that has seen significant advancements and innovations. It begins by tracing the history and development of existing multi-dimensional shaping methods. The narrative then transitions to the design of our novel multi-dimensional shaping.

In recent years, several studies have shown that multi-dimensional modulation formats can improve the performance of coherent transmission systems as they offer higher Euclidean distances, better spectral efficiency (SE) granularity, and a potential to mitigate nonlinear fiber effects compared to conventional formats such as polarization division multiplexed 2D formats like PDM quadrature phase shift keying (PDM-QPSK) or PDM M-QAM [141, 181]. For instance, in [18], the mitigation of cross-polarization modulation (XpolM) was realized by an 8D-QPSK using the polarization balanced concept with an SE of 2 bits over two polarization states and one-time slot (4D) as seen in Chapter 2. The experimental demonstration showed improvement in net system margin relative to PDM-BPSK over dispersion-managed (DM) systems. Later on, the same concept was used in [140] to design modulation formats with a higher SE of 2.5 and 3.5 bits per 4D symbol respectively. However, their non-linear gains in DM systems vanish over dispersion-unmanaged (DUM) systems as shown in Chap. 2, because a large accumulated dispersion destroys their 8D polarization balanced structure and their low-energy-variation property. In this chapter, our interest is in achieving performance gains over DUM systems using low-complexity multi-dimensional formats and aiming at higher spectral efficiency values.

Over DUM systems, for a higher SE of 6 bits per 4D symbol, several 4D modulation formats such as 4D-64 Set Partitioned (SP)-12QAM [25], Honey Comb (HC) grid [27] and others [182–184] were designed by increasing the Euclidean distances and enhancing the non-linearity tolerance through the minimization of the peak-to-average-ratio (PAPR) of the constellation which

leads to lower non-linear interference (NLI) generated by both intra-channel and inter-channel effects [185]. While PAPR does not capture all the intricacies of the NLI accumulation through the link, it has the merit of being a simple design metric. Then, the 4D-2A-8PSK which is proposed by Kojima et al. in [28, 186], and the 4D-64PRS which is proposed in [187] further increase the Euclidean distances and maintain the constant modulus characteristic, and achieve both linear and nonlinear gains. They are designed by two different constellation structures, combining the multi-dimensional coding techniques with geometric shaping. More recently, new geometric shaping schemes achieving higher spectral efficiency were also investigated in [188, 189].

Inspired by the findings in the existing literature, we propose a 4-dimensional 2-ary amplitude ring-switched modulation format with 64 symbols, which is denoted as 4D-2A-RS64 constellation encoded over two polarization tributaries to improve the transmission performance over long-haul optical fibers in the presence of the non-linear Kerr effect. At a spectral efficiency of 6 bits per 4D, our simulation results show that this format outperforms the polarization division multiplexed (PDM) 8QAM-star modulation as well as the 4D-2A-8PSK over links without inline dispersion management. We evaluate the performance for a WDM transmission of 11×90 Gbaud channels over a multi-span SSMF link. For an achievable information rate of 4.8bit/s/Hz, the maximum transmission distance is improved by 10.6% (400 km) and 4% (160 km) compared to PDM-8QAM-star and 4D-2A-8PSK respectively. The achieved gains are composed of a linear part and a non-linear part, respectively from the improved Euclidean-distance distribution and the constant power property of the 4D modulation. Furthermore, the geometric shaping of the proposed scheme is easy to implement and is robust to Mach-Zehnder modulator (MZM) imbalances and quantization errors stemming from the finite digital-to-analog converter (DAC) vertical resolution. This robustness is compared to the one of other geometric-shaped non-linearity tolerant 4D schemes such as the 4D-2A-8PSK and the 4D-64PRS that can be both outperformed by our scheme under severe transmitter-impairment conditions.

In section 3.2, we present the state-of-the-art in multi-dimensional shaping designs for achieving both linear and non-linear gains. Subsequently, in section 3.3, we introduce our design, the 4D-2A-RS64 constellation. Finally, we give the conclusion and the perspective on multi-dimensional shaping designs.

3.2 Existing 6 bit/4D-symbol Multi-dimensional Shaping Designs

3.2.1 4D-64-set-partitioned-12QAM and 4D-128-set-partitioned-16QAM

The 4D-64SP-12QAM proposed in [25], is a four-dimensional (4D) coded modulation format. The projection of 4D-64SP-12QAM over one I/Q plane is shown in Fig. 3.1. The 4D-64SP-12QAM demonstrates superior performance compared to existing PDM-8QAM, particularly in achieving

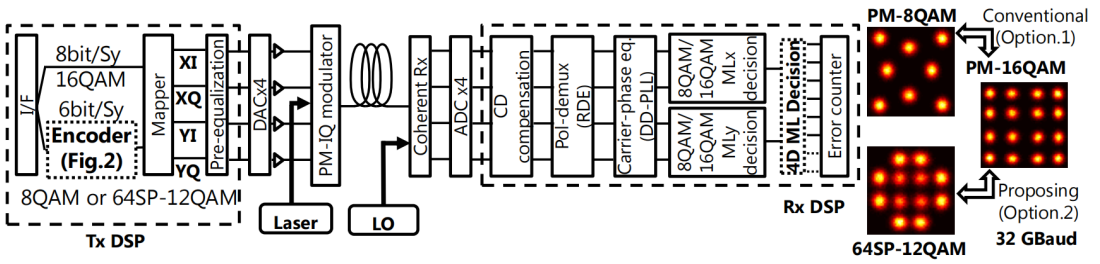


Figure 3.1: Schemes of switching between two transmission modes with SE of 6 and 8 bit/symbol. Source: [25]

longer transmission distances with the same spectral efficiency. Experimentally, it has shown over 0.4 dB better received Q factor than PM-8QAM over 6000 km at a spectral efficiency (SE) of 6bit/s/Hz. Set-partitioning schemes can be effectively integrated into conventional polarization-diverse QAM transmitters with minimal increases in complexity. The technical approach involves the elimination of specific 4D symbols from the PM-16QAM constellation to increase Euclidean distance between symbols by set partitioning techniques, thus reducing symbol error rates. This strategy also minimizes the bit error rate by decreasing the number of neighboring points for each constellation point. Notably, the 4D-64SP-12QAM format is compatible with existing PM-QAM systems, which facilitates cost-effective implementation by allowing shared components in the transmission system, such as the mapper and demodulation digital signal processing (DSP). The performance of the 4D-64SP-12QAM signal is validated through both back-to-back simulations and real-world experiments, demonstrating its superior sensitivity and extended reach. This is particularly evident in the long-haul WDM transmission experiments, where the new format achieved a transmission reach of 6000 km, surpassing PM-8QAM in terms of both sensitivity and distance.

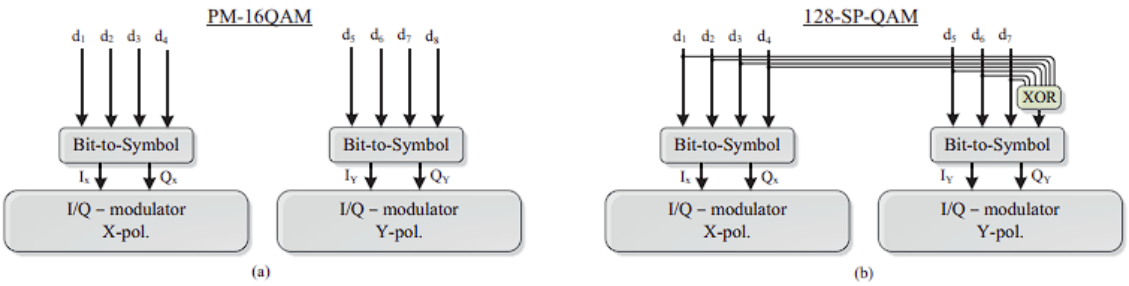


Figure 3.2: Concept for (a) generating PM-16QAM and (b) generating 128-SP-QAM. Source: [26]

Another set-partitioning-based design with higher spectral efficiency of 7 bits/4D-symbol is the 4D-128SP-16QAM in [26], where the authors experimentally compared the performance of 128-level set-partitioned quadrature amplitude modulation (128-SP-QAM) with PM-16QAM in optical fiber communications. Fig.3.2(a) illustrates a standard PM-16QAM transmitter where eight

binary sequences are merged to create 4-level signals, modulating 16QAM across two polarization states. For implementing 128-SP-QAM, maintaining constant parity among the eight bits is essential. This can be achieved by encoding one of the bit sequences through an XOR operation applied to the other seven bits, as depicted in Fig.3.2(b). Consequently, the existing architecture of a PM-16QAM transmitter can be adapted for 128-SP-QAM generation with only slight modifications. The principle is similarly applicable to 4D-32-SP-16QAM and 4D-64-SP-16QAM, with the primary variation being the number of XOR gates required, as detailed in [190]. Key findings in [26] demonstrate that 128-SP-QAM offers significantly better back-to-back sensitivity and increased maximum transmission distance compared to PM-16QAM, under both single-channel and WDM transmission scenarios.

3.2.2 Honey Comb (HC) Grid Constellations

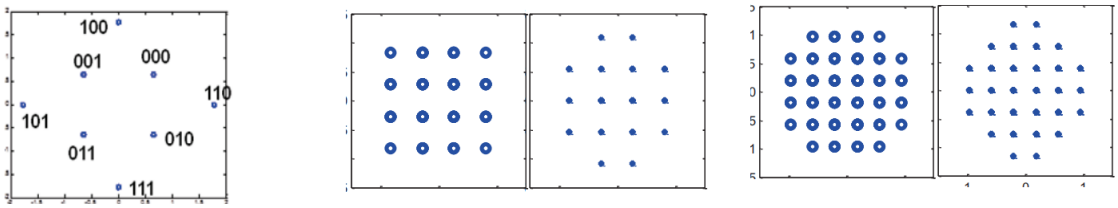


Figure 3.3: 2D-projection of three modulation formats. Left: PM-8QAM star with mapping (same projection for Pol X and Pol Y); Middle: HC64 constellation (Left: Pol.X, Right: Pol.Y); Right: HC256 constellation (Left: Pol.X, Right: Pol.Y). Source: [27]

Honey Comb (HC) grid is designed in [27], to explore 4D constellations surpassing the performance of PM-8QAM and PM-16QAM. The authors began with a compact lattice grid, initially filled with 6P-QPSK points as described in [139, 191], and then expanded these grid points. This expansion led to the identification of two point sets residing within two distinct 4D hyperspheres, encapsulating 64 and 256 points each. The orthogonal projections of these points on X and Y polarizations are depicted in Fig. 3.3. Drawing a parallel to the most efficient packing in a 2D Honey Comb (HC) grid, they named these new constellations HC64 and HC256. Compared to their PM M-QAM equivalents, having the same minimum Euclidean distance (d_{min}), HC constellations demonstrated an enhanced asymptotic power efficiency ($d^2_{\text{min}}/(4E_b)$) [139], where E_b is the average energy per bit, with improvements of 1.7 dB and 1.5 dB for HC64 and HC256, respectively. This design contributes to the field of optical communications by proposing new modulation formats that offer higher power efficiency and improved performance in terms of optical signal-to-noise ratio (OSNR) sensitivity and reach.

3.2.3 4D-2A-8PSK Constellation

Later on, Kojima et al. in [28, 186] proposed the 4D-2A-8PSK achieving higher gains, in both linear and nonlinear regimes, than PDM-8QAM-star, PDM-8PSK, and circular PDM-8QAM over DM systems. The 4D-2A-8PSK is a part of a constellation family detailed in [186, 192, 193] and depicted in Fig. 3.4.

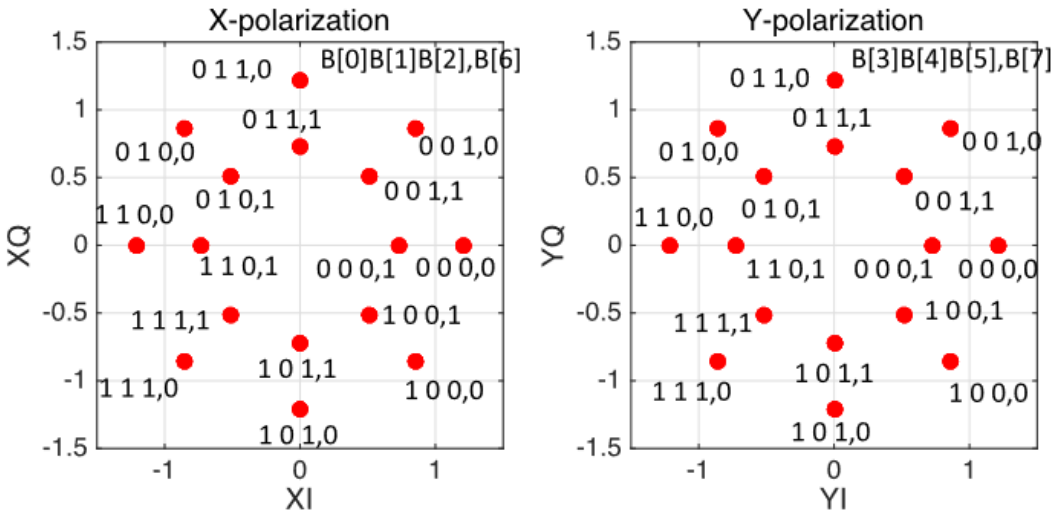


Figure 3.4: Constellation and bit-to-symbol mapping of 4D-2A-8PSK. To better unify the expressions in other articles, $B[i]$ will be expressed as b_i below. Source: [28]

The constellation is fundamentally two 8PSK with two distinct amplitudes, represented by radii R_1 and R_2. In the context of combined X- and Y-polarizations (i.e., in 4D space), it forms 256 unique combinations, equating to 8 bits per 4D symbol. By imposing a condition where the two polarizations exhibit complementary radii (i.e., choosing R_1 for one polarization imposes R_2 for the other), it is possible to implement set-partitioned (SP) 4D codes. These codes maintain a 4D constant modulus property, contributing to superior nonlinear transmission characteristics. The ratio R_1/R_2, which is less than or equal to 1 ($R_1 < R_2$), is defined as the ring ratio. When the ring ratio equals to 1, this modulation format simplifies to regular Polarization Division Multiplexing (PDM)-8PSK. Fig. 3.4 also illustrates the mapping rule for 4D-2A-8PSK [194]. Set-partitioning can be applied here. The eight bits for modulation, denoted as $b_0, \cdots, b_7$, are allocated such that b_0, b_1, b_2 and b_3, b_4, b_5 are used for Gray-mapped 8PSK at X- and Y-polarizations, respectively, while b_6 and b_7 determine the amplitude in each polarization. By applying Eq. (3.1) and Eq. (3.2), we select the 64-point constellation from the 256 combinations, corresponding to the 4D-2A-8PSK constellation with a spectral efficiency of 6 bits/4D symbol. The authors also proposed other constellations with 5 or 7 bits/4D-symbol by applying different Boolean equations.

$$b_6 = b_0 \bigoplus b_1 \bigoplus b_2 \bigoplus b_3 \bigoplus b_4 \bigoplus b_5 \tag{3.1}$$

$$b_7 = \overline{b_6}. \tag{3.2}$$

The observed gains stem from a larger Euclidean distance, a 4D constant modulus constraint, and the use of Gray labeling. The Euclidean distance gain comes from the nesting of two 8PSK constellations of different amplitudes followed by set partitioning. The complementary amplitudes of the two polarization tributaries ensure a constant energy property at each time slot, thus eliminating energy variations and reducing nonlinear effects. The Gray labeling also brings an improved achievable information rate (AIR) compared to PDM-8QAM-star which is not Gray labeled. The ring ratios of the 4D-2A-8PSK are computed based on an optimization of the AIR at different signal-to-noise ratios (SNR) [195]. However, the optimal ring ratio value varies between 0.6 and 1 for SNR values ranging between 5 and 12 dB, which shows that its performance is very sensitive to changes in the ring ratio. The performance of 4D-2A-8PSK was not explored over DUM systems.

3.2.4 4D-64PRS Constellation

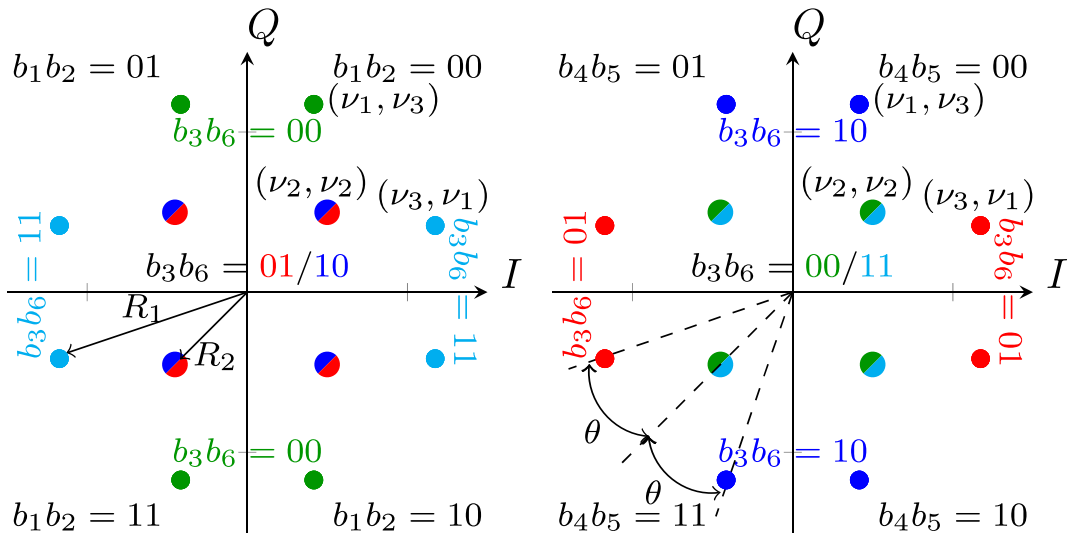


Figure 3.5: 2D-projections of the designed 4D-64PRS modulation and associated bits / polarization bits mapping. The squared values of the rings are $R_1^2 = \nu_1^2 + \nu_3^2$ and $R_2^2 = 2\nu_2^2$. Source: [29]

Later on, another four-dimensional 64-ary polarization ring switched modulation format, called 4D-64PRS with spectral efficiency of 6 bits/4D-symbol was introduced in [29] by Chen et al. The modulation uses a new mapping technique named 'Orthant-symmetric labeled constellation', a method that maps the possible combinations in quadrants separately to binary labels to facili-

tate the upcoming set partitioning. The construction of 4D-64PRS is done over two rectangular 16QAM constellations from which the 4 outermost points of maximum energy are removed. Hence, there are two rings of different radii and there are $2 \times 4 \times 8 = 64$ symbol combinations for which we will present the design rules in a few lines. The bit mapping techniques associated with these modulations are quite interesting and inspire our novel modulation design in the next section. We illustrate, in the following, the bit mapping for the 4D-64PRS modulation. 2D projections of the constellation over the two polarization tributaries are shown in Fig. 3.5. The inner ring contains 4 symbols and the outer ring contains 8 symbols. The combinations over two polarization tributaries can only occur between two different rings, leading to 64 combinations. Hence, 6 bits could be distributed to 64 combinations.

The quadrants in each polarization are defined by 4 out of the 6 bits, namely, $[b_1, b_2]$ and $[b_4, b_5]$ for the first and second polarization respectively. After two quadrants are selected, 2D projected symbols in the first and second polarization tributaries are valid 4D symbols only if they share the same color shown in Fig. 3.5. The color coding is chosen in order to maximize the MSED between neighboring signals. It also respects the constant 4D-modulus constraint. These rules yield a total of four 4D symbols in each pair of quadrants, which could be labeled to the two remaining bits $[b_3, b_6]$. The 4D-64PRS format can be seen as the transmission of alternating rings in two polarization states. In other words, when an inner ring (R_2) is transmitted in the first two dimensions (I and Q of first polarization), the outer ring (R_1) is used in the second two dimensions (second polarization). This case corresponds to dark blue or red points. The opposite alternation (outer ring, inner ring) is used for light blue or green points. Finally, the authors in [29] add a geometric shaping to optimize the ratio between R_1 and R_2 and also a tilt angle θ of the outer ring.

The 4D-64PRS constellation defined above was then compared to PDM-8QAM-star. Even though the former has a slightly lower MSED than the latter, it has 6 times fewer symbol pairs at MSED, also known as the kissing number. This offers up to 0.7 dB gain in SNR over an AWGN channel. For a long-haul non-linear optical fiber transmission over 8000 km, gains of up to 0.27 bit/4D in achievable information rate (5.5% data capacity increase) were observed. The 4D-64PRS has also a slightly lower MSED than 4D-2A-8PSK but also has a lower kissing number, which yields gains of up to 0.4 dB in SNR in the AWGN channel compared with 4D-2A-8PSK. It also has a better performance in the presence of non-linear effects because of its 4D constant modulus constraint. AIR-based optimization combined with geometric shaping is used to find the optimum ring ratio and angle for SNR levels between 3 and 13 dB. The optimized ring ratio varies in the range of $0.53 - 0.61$ and the relative rotation angle between the two rings is fine-tuned between 23.4 to 27.2 degrees.

3.2.5 Conclusions from the different strategies

At 6 bits/4D symbol, the two 4D constructions in Sec. 3.2.3 and Sec. 3.2.4 show better performance than other designs. They both combine geometric shaping [143, 144] and AIR-based optimization [145, 196] to find the best ring ratio and constellation point locations that maximize the AIR at a given SNR value. However, the fine geometric optimizations increase the constraints on the generation of the designed constellation (DAC resolution, stability and linearity of optical modulators, etc.). Moreover, their performance fluctuates considerably with slight changes in the design parameters.

In the field of optical communications, specifically in geometric shaping research, it is common to assume the use of perfect digital-to-analog converters (DACs) and analog-to-digital converters (ADCs) in transceivers. As highlighted in the study [117], advanced, geometrically-shaped modulation formats require DACs and ADCs with high resolution for effective generation and detection. However, the 8-bit quantizers commonly available commercially, which effectively provide around 5-6 bits of resolution due to additional distortions, limit the performance of these sophisticated modulation formats due to their limited resolution. For example, the work in [197] observed a significant 20% drop in power consumption when the resolution is lowered from 8 to 4 bits. Moreover, geometric shaping often leads to higher peak amplitudes and, thus higher PAPR, which further aggravates the quantization issues caused by low-resolution DACs. Therefore, it is essential to take into account the resolution limitations of DACs and ADCs when designing practical Multidimensional (MD) modulation formats.

The advent of high baud rates and geometric shaping in optical communications has led to an increased need for precision in the tolerance of optical components and the accuracy of control loops. A fundamental requirement is the ultra-precise alignment of path timing, especially between the in-phase (I) and quadrature (Q) paths, demanding a sub-picosecond level of accuracy. Additionally, to achieve optimal system performance, the gains of electric drivers must be precisely matched. The accuracy of modulator phase bias control loops is also crucial, as they need to keep the quadrature phase error within a narrow range of just a few degrees. Any deviations in the bias of the Mach-Zehnder (MZ) modulator or suboptimal extinction ratios can lead to direct current (DC) offsets in the transmitted signal constellations, thereby significantly degrading the signal quality. As noted in research [198–200], typical MZM error values might include a $\pm 10^{\circ}$ quadrature-phase imbalance, a ± 3 dB I/Q gain imbalance, and an I/Q offset ranging from $-40 \sim -10$ dB as defined in Eq. (1-3) in [199]. These issues are crucial for Multidimensional (MD) modulation formats and geometric shaping strategies, highlighting the essential need for high precision in the design of advanced optical transmitters.

In the realm of MD modulation formats post-GS, it is also important to address the increased complexity that these formats introduce. For MD modulation formats post-GS, the standard Gray

mapping typically becomes impractical, complicating the transformation of symbols into soft-decision bit metrics such as Log-Likelihood Ratios (LLRs). To handle this increased complexity, the max-log approximation is often used, which is particularly effective in AWGN channels, simplifying LLR calculations by avoiding exponential functions and reducing the need for Euclidean Distance (ED) calculations in Maximum Likelihood (ML) optimal demappers. For square QAM or PDM-QAM modulations, this approximation leads to a simpler piece-wise linear relationship between received symbols and LLRs, calculated for each real dimension. This simplification is a key reason for the widespread use of the max-log approximation in practical settings [201]. However, for non-square QAM or non-gray-labeled formats, the demapping process becomes more complex, often requiring extensive ED calculations or larger pre-computed look-up tables. To mitigate these challenges, MD modulation formats with regular structures, such as lattice, symmetric, or shell structures, are preferred to decrease computational demands. These structured formats offer a balance between performance and complexity, making them more manageable for practical applications in high-performance optical communication systems.

3.2.6 Metric Comparisons for Different 4D and 8D Constellations

To better compare different constellations introduced in Chap. 2 and Chap. 3, we summarize in the following table, Tab.3.1, different properties and metrics for major works and include in the last row our proposed scheme that will be described in the next section.

Table 3.1: Metrics and properties for different 4D and 8D constellations

	SE_{4D}	d^2_{min}	n_d	PAPR	Kurtosis	$\sigma^2_{P_s}$	α	β	CM	Sym.
PDM-BPSK	2	4	4	1	-1	0	1	0.5	Y	Y
PDM-QPSK	4	2	32	1	-1	0	1	0.6	Y	Y
PDM-8QAM-rec[155]	6	0.67	128	1.33	-0.89	0.23	1	0.66	N	Y
PDM-8QAM-star[155]	6	0.85	192	1.58	-0.67	0.68	1	0.69	N	N
PDM-8QAM-circular[155]	6	0.865	119	1.14	-0.86	0.29	1	0.68	N	N
PDM-16QAM	8	0.4	768	1.8	-0.68	0.64	1	0.7	N	N
X-Constellation[18]	2	8*	32*	1	-1	0	0	0	Y	Y
PS-QPSK[141]	3	4	24	1	0	0	1	0.5	Y	Y
PB-6B8D[140]	3	4*	384*	1	-1	0	0	0	Y	Y
4D-PRS64[29]	6	0.684	32	1	-0.7	0	1	0.66	Y	N
8D-2048PRS-T1[29]	5.5	1.15*		1		0	0.96	0.64	Y	N
8D-2048PRS-T2[29]	5.5	0.76*		1		0	0.87	0.55	Y	N
4D-2A-8PSK (6 bits)[28]	6	0.88	128	1	-0.83	0	1	0.67	Y	Y
4D-2A-8PSK (7 bits)[28]	7	0.27	64	1	-0.87	0	1	0.65	Y	Y
4D-OS128[202]	7	0.14	16	1.54		0.8			N	N
4D-128SP-16QAM[26]	7	0.8	864	1.8		0.645			N	N
4D-2A-RS64 (our proposal)	6	0.8	40	1	-0.64	0	1	0.66	Y	Y

The calculation of each parameter in the table is based on the 4D symbols. Each 4D symbol consists of two 2D symbols in two orthogonal polarization tributaries. Our designed modulation format is highlighted in bold. The asterisk $*$ designates 8D MSED values and number of neighbors at MSED n_d in 8D. The definition of each metric is given below:

- SE_{4D} is the spectral efficiency in two polarization tributaries and one time slot.
- d^2_{min} is the minimum squared Euclidean distance (MSED) of 4D-symbol. The computed values correspond to symbols that are normalized to a unit power per 2D. d^2_{min} is the minimum value of all possible $\mathrm{d}^2(\mathbf{x},\mathbf{y})$:

$$d^2(\mathbf{x},\mathbf{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2 \tag{3.3}$$

  where $\mathbf{x} = (x_1, ...x_n)$ and $\mathbf{y} = (y_1, ...y_n)$ are two n-Dimensional symbols.
- n_d is the total number of pairs of constellation points at MSED.
- PAPR is the peak-to-average power ratio, variance of the signal energy with respect to its mean-average energy:

$$\mathrm{PAPR} = \frac{||\mathbf{s}_{peak}||^2}{s^2_{rms}} \tag{3.4}$$

  where $\mathbf{s}_{peak}$ is the 4-D symbol with the peak amplitude and s_{rms} is the root mean square value of the 4D-symbol energy levels.
- Kurtosis is the fourth-order moment of the 2D projections of the constellations [203]:

$$K_x \triangleq E[|x^4|] - 2(E[|x^2|])^2 - \left|E[x^2]\right|^2 \tag{3.5}$$

  where $\mathrm{K_x}$ is the Kurtosis of a 2D symbol x.
- $\sigma^2_{P_s}$ is the variance of the 4D symbols' energies:

$$\sigma^2_{P_s} = E[\|\mathbf{s}\|^2 - E[\|\mathbf{s}\|]^2] \tag{3.6}$$

  where $\mathbf{s}$ is a 4D symbol.
- α and β are the maximum DOP (degree of polarization) change and the average DOP change between two consecutive 4D symbols respectively. For the 8D polarization balanced schemes, α and β are computed within the 8D symbol and not between 4D symbols from different 8D combinations.

The Stokes vector $\mathbf{S}$ of a 4D symbol $\mathbf{s}$, composed in Jones formalism of $\mathbf{s} = [s_{pol_1}, s_{pol_2}]^T$ is:

$$\mathbf{S} = [S_1, S_2, S_3] \text{ where } S_1 = \left|s_{pol_1}\right|^2 - \left|s_{pol_2}\right|^2,\ S_2 = 2\Re(s_{pol_1} s^*_{pol_2}),\ S_3 = 2\Im(s_{pol_1} s^*_{pol2}) \quad (3.7)$$

where $\Re(a)$ is the real part of a and $\Im(a)$ is the imaginary part of a . The DOP for the i-th transmitted 8D symbol is defined as:

$$p_i = \frac{\|\mathbf{S}_{t_1} + \mathbf{S}_{t_2}\|}{\left|s_{pol_1,t_1}\right|^2 + \left|s_{pol_1,t_2}\right|^2 + \left|s_{pol_2,t_1}\right|^2 + \left|s_{pol_2,t_2}\right|^2} \quad (3.8)$$

where $0 \leq \text{pi} \leq 1$, and t_1 and t_2 indicate time slot 1 and time slot 2. Then, α and β are defined as:

$$\alpha = \max_{i \in \{1,2,\ldots,M\}} (p_i) \quad (3.9)$$

$$\beta = \frac{1}{M} \sum_{i=1}^{M} p_i \quad (3.10)$$

- Constant modulus (CM) indicates whether the power in each time slot is the same.
- 'Sym' indicates whether the constellation structure is symmetric. A symmetric structure is one where all the neighboring symbols with one-bit difference, for any given symbol, have equal Euclidean distance from that symbol. This property reduces the mapping penalty with respect to Gray mapping.

3.3 Our Proposal: The 4D-2A-RS64 Constellation

3.3.1 Principle of 4D-2A-RS64

In multi-span transmission systems, non-linear interference (NLI) arising from the Kerr effect, and accumulated amplified spontaneous emission noise from EDFAs dominate the AIR performance of a given modulation format. Fewer temporal power variations help in reducing the former and a larger minimum squared Euclidean distance (MSED) of the constellation increases the performance in the presence of the latter. Hence, we conceive our constellation in this spirit. Fig. 3.6 shows the construction and the binary labeling of the proposed scheme, named four-dimensional two-amplitude ring-switched constellation with 64 symbols: 4D-2A-RS64. Two PSK modulations are considered: a QPSK with a radius R_1 and an 8-PSK with a radius $R_2 = 2R_1$. The 8-PSK on Y polarization (Pol Y) has a phase rotation of $\frac{\pi}{8}$ with respect to the QPSK on X polarization (Pol X) to optimize the distribution of the Euclidean distances between 4D symbol pairs as will be shown in Section 3.3.3. A 4D symbol, defined as two I-Q symbols multiplexed on two orthogonal polarization tributaries over one time slot, exists only when the two 2D symbols have the same color in Fig. 3.6, which means that two 2D symbols from different rings R_1 and R_2 are chosen, leading to

64 combinations. If a QPSK symbol with radius R_1 is chosen for Pol X, an 8-PSK symbol with radius R_2 is selected for Pol Y. This configuration contains 32 distinct 4D symbols. Inversely, the other configuration also contains 32 symbols.

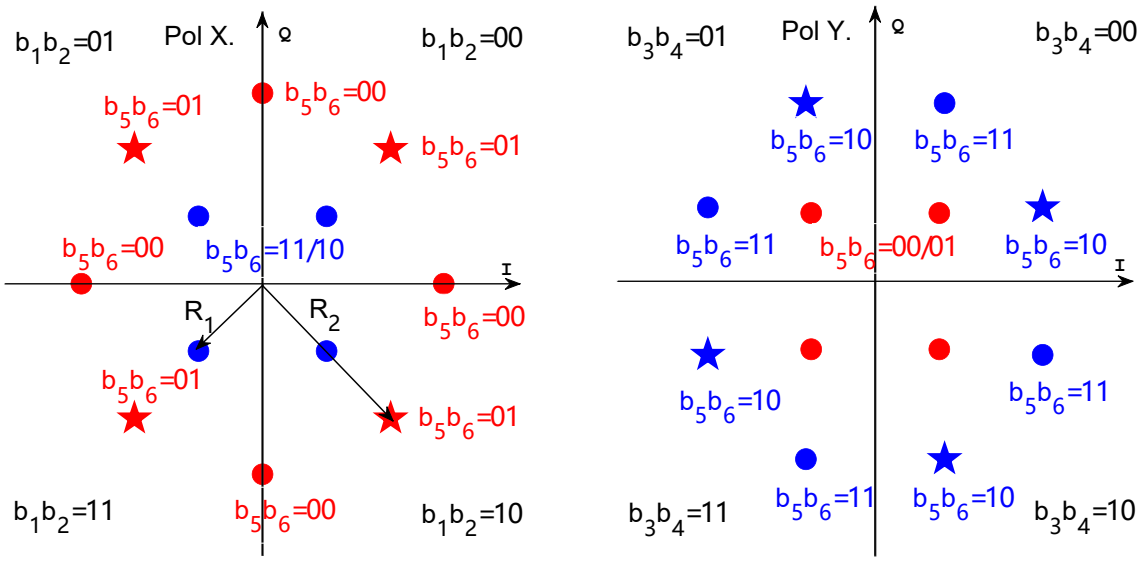


Figure 3.6: 2D-projections of the designed 4D-2A-RS64 modulation and associated bit-to-symbol mapping with $R_2 = 2R_1$. Red (resp. blue) circle markers and star markers on R_2 have the same labeled $[b_5, b_6]$ bits.

Now, we define the bit mapping rules to enhance the performance over an AWGN channel by maximizing the MSED and minimizing the bit error probability. The bit mapping is inspired by the one used in [29]. We decompose the 2D constellation in the I-Q plane into four areas defined by the four quadrants as shown in Fig. 3.6 and we associate the symbols on the main axes in Pol X to the clockwise neighboring quadrant. Each 4D symbol is labeled by a six-bit vector $[b_1, \ldots, b_6]$ as follows:

- The quadrants in each polarization are defined by 4 out of the 6 bits, namely, $[b_1, b_2]$ and $[b_3, b_4]$ for the first and second polarization tributary respectively as we show in Fig. 3.6. These 4 bits determine which out of the 16 quadrant combinations is to be chosen.
- In each quadrant combination, there are 4 possible combinations of 2D symbols. $[b_5, b_6]$ are assigned to those as shown with the labels in Fig. 3.6.

A last optimization step of the constellation can typically be a geometric shaping consisting of changing the ring ratio between R_1 and R_2, $r = R_1/R_2$, or the relative orientation of the QPSK and 8-PSK constellations. Based on an optimization of the achievable information rates taking into account the bit mapping, the authors in [29] and [192] improved the linear performance of 4D-64PRS and 4D-2A-8PSK over a given signal-to-noise ratio (SNR) value for a target pre-FEC BER. However, a trade-off exists between the complexity of the implementation of the geometrically

shaped constellation and the target performance. Overly precise optimization can be difficult to achieve in practice because it increases the cost and the footprint of the transmitter (high effective number of bits (ENoB) requirements at high symbol rates). In Fig. 3.7(a), we report the optimal ring ratio (dashed-line curve and right y-axis) for each SNR and find it to be between 0.47 and 0.51. SNR is defined as the symbol energy over the power spectral density of the AWGN noise. In what follows, the ring ratio of our scheme is set to 0.5, which is close to optimal over the studied SNR range.

In Fig. 3.7(d), we report the GMI of the three 4D schemes with optimized parameters at SNR= 8 dB and no implementation penalties, and compare them to the performance of PDM-8QAM-star. At an FEC overhead of 25%, corresponding to a net rate of 4.8 bit/4D-symbol, we see that 4D-2A-RS64, 4D-64PRS and 4D-2A-8PSK have 0.4 dB, 0.7 dB and 0.3 dB gain, respectively, over PDM-8QAM-star. The 4D-2A-RS64 performs better than 4D-2A-8PSK by 0.1 dB. Conversely, at the SNR required by PDM-8QAM-star to achieve a net rate of 4.8 bit/4D symbol, 4D-2A-RS64 achieves a GMI gain of 0.15 bit/4D symbol compared to PDM-8QAM-star.

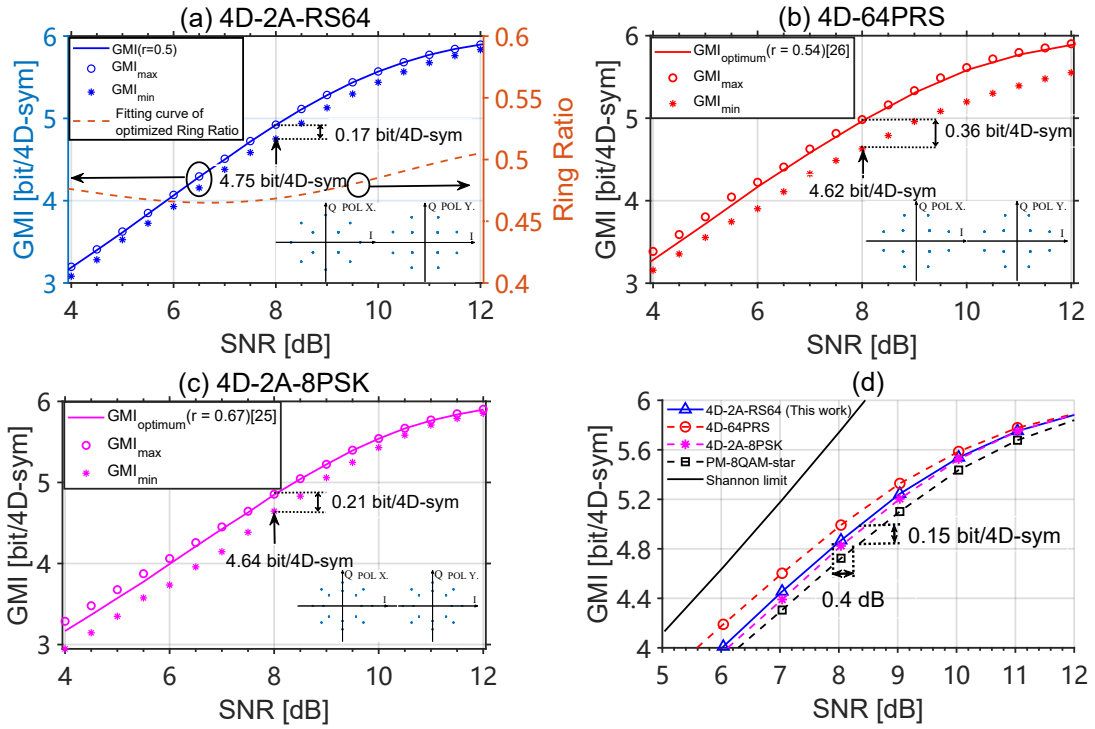


Figure 3.7: Solid line on (a), (b), (c): GMI (with optimized parameters for SNR=8 dB) versus SNR over AWGN channel; Dashed line on (a): Optimum ring ratio $r = R_1/R_2$ vs SNR. Asterisks: minimum GMI obtained by adding MZM imbalances of $\pm 15^\circ$ for I-Q quadrature and 1.7 dB for I-Q gain imbalances. Empty circles: maximum GMI when ring ratio is optimized for each SNR level. (d) GMI vs SNR for various modulation formats with 6 bit/4D-symbol. The insets within the figures show the 2D-projections of the three constellations. Ring ratios are 0.5, 0.54, and 0.67 for 4D-2A-RS64, 4D-64PRS, and 4D-2A-8PSK respectively.

3.3.2 Tolerance to Transmitter Impairments

Before further investigating the performance of our proposed modulation, we test its robustness to ring ratio variations and relative rotation angle variations between the two rings and compare it to the robustness of 4D-64PRS and 4D-2A-8PSK. These ring and phase variations or deviations from their optimal values can be introduced by imbalances at the Mach-Zehnder modulator (MZM) such as gain imbalances, quadrature imbalances or even skew between I and Q components of each polarization tributary [198, 200]. First, we fix the structure of the two-dimensional per-polarization projections of all three constellations using the optimized ring ratio for an SNR of 8 dB. In this case, the performance of each 4D modulation is given by the full-line curves in Fig. 3.7(a-c) for SNR values between 4 and 12 dB. The optimal ring ratio value is also given in the legend of each sub-figure. The curves plotted with empty-circle markers show the maximum GMI values of each scheme when using the optimal ring ratio for each SNR value. Our proposed scheme shows a near-optimal performance with a fixed ring ratio in the studied SNR range. Later, we emulate gain and quadrature imbalances at the MZM using the following model [198]:

$$\begin{aligned} E_x &= I_x + \alpha_x e^{j\theta_x} Q_x \\ E_y &= I_y + \alpha_y e^{j\theta_y} Q_y \end{aligned} \tag{3.11}$$

where I_x and Q_x (resp. I_y and Q_y) are the ideal in-phase and quadrature components of each polarization tributary, E_x and E_y are the generated complex amplitude at the outputs of each I-Q modulator, θ_x (resp. θ_y) are the I-Q phase angles for the x (resp. y) polarization tributary and $\alpha_x \leq 1$ (resp. $\alpha_y \leq 1$) is a gain discrepancy between the I and the Q components of each polarization tributary. When θ_x (resp. θ_y) are equal to 90°, the quadrature is perfect between the I and Q components. When α_x and α_y are equal to 1, no gain disparity is introduced between the two components. According to an Optical Internetworking Forum (OIF) agreement on 400G transceivers [204], that is used as a basis for interoperability, the maximum allowable fluctuations for the MZM should be in the range of $\pm 5°$ for the I-Q quadrature imbalance and 1 dB for the I-Q gain imbalance. To align with these requirements, these imbalances should be estimated and corrected through dedicated DSP algorithms and feedback loops. In our study, we chose the ranges for the considered imbalances similarly to previous works [198–200], with worst-case values going beyond the typical values. In Fig. 3.7(a-c), the worst-case GMI values are shown through the curves with asterisk markers. These are obtained for maximum deviations of θ_x (resp. θ_y) of $90° \pm 15°$ and a maximum gain imbalance of 1.7 dB between I_x and Q_x (resp. I_y and Q_y) components. With these simulated imbalances, at SNR= 8 dB, our 4D-2A-RS64 scheme has a minimum GMI floor of 4.75 bit/4D-symbol and GMI fluctuation of 0.17 bit/4D-sym. It shows better fault tolerance than 4D-64PRS and 4D-2A-8PSK, which have minimum GMI floors of 4.62 bits/4D-sym

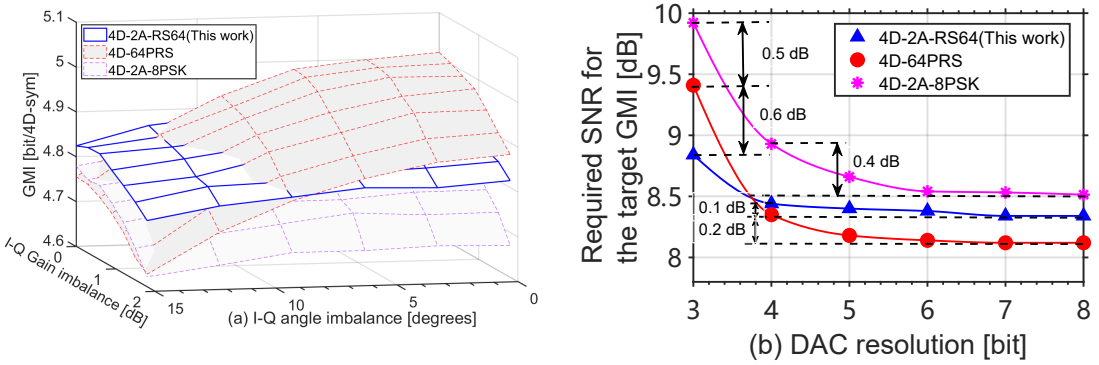


Figure 3.8: (a) GMI versus I-Q gain and quadrature imbalances at SNR=8 dB. (b) Required SNR versus DAC resolutions for different formats at a GMI of 5 bits/4D symbol (uniform quantization).

and 4.64 bits/4D-sym, and GMI fluctuations of 0.36 bits/4D-sym and 0.21 bits/4D-sym respectively.

To get a full view of the performance degradation of the three 4D schemes, we study the GMI variations for an I-Q gain imbalance within the range of 0 to 1.8 dB and an I-Q quadrature imbalance within the range of 0 to 15 degrees for each polarization tributary. In Fig. 3.8(a), we can see that, as the I-Q gain imbalance increases, the performance of 4D-64PRS deteriorates faster than the one of the proposed 4D-2A-RS64. The former becomes worse than the latter for I-Q angle imbalances higher than 12 degrees in the absence of an I-Q gain imbalance, and higher than 8 degrees for an I-Q gain imbalance of 1.8 dB. It is important to note that all three 4D modulation schemes possess symmetrical characteristics along the I and Q axes; consequently, we obtain similar performance degradation for I-Q gain imbalances between 0 to -1.8 dB and I-Q quadrature imbalances between 0 and -15 degrees. 4D-64PRS requires the generation of precise amplitude and phase values to achieve the geometric shaping gain as shown in Fig. 3.7(b). As the MZM imbalance increases, its performance gradually deteriorates until it falls below that of 4D-2A-RS64. 4D-2A-8PSK comprises two 8PSK constellations, hence the inner ring of 8-PSK necessitates a fine amplitude and phase mapping, leading to continuous performance degradation as the imbalances increase; moreover, its initial imbalance-free performance is lower than that of 4D-2A-RS64.

An ideal digital-to-analog converter (DAC) is used in [28] and [187]. However, it is interesting to take into account the quantization error stemming from the finite resolution of DACs when evaluating the performance of geometric-shaped (GS) constellations. Moreover, high-resolution DACs not only add higher hardware costs but also a higher energy consumption [117, 197]. To test the impact of DAC quantization, each transmitted sequence is oversampled to 2 samples per symbol, shaped using a root-raised cosine (RRC) filter with a roll-off factor equal to 0.1, and uniformly quantized by a DAC with a vertical resolution between 3 and 8 bits. Even though the

widely used DACs and ADCs in high-speed optical transceivers use 6 to 8 physical bits, the effective number of bits (ENoB) that take into account the signal distortions and noise added by the converters may drop below 4 bits [205, 206]. In our study, we consider a frequency independent quantization DAC. The back-to-back transmission performance is measured over an AWGN channel at a symbol rate of 90 Gbaud. In Fig. 3.8(b), we show the minimum required SNR to achieve a GMI of 5 bit/4D-sym. For DAC resolutions ranging from 5 to 8 bits, there is a fine mapping for each constellation symbol and the SNR penalties are practically negligible. In such cases, as illustrated in Fig. 3.7(d), 4D-2A-RS64 incurs a slight SNR penalty compared to 4D-64PRS. At 4 bits, the GMI penalties are 0.1 dB, 0.2 dB, and 0.4 dB for 4D-2A-RS64, 4D-64PRS, and 4D-2A-8PSK respectively. In the meanwhile, 4D-64PRS requires the lowest SNR value. At 3 bits, our 4D-2A-RS64 scheme shows a gain of 0.6 dB and 1.1dB relative to 4D-64PRS and 4D-2A-8PSK respectively. The penalty of the 4D-64PRS comes from its sensitivity to phase changes; indeed, the phases of the outer ring constellation points cannot be accurately achieved with 3-bit resolution DACs. The 4D-2A-8PSK penalty comes from the use of an 8-PSK on its inner ring which requires a higher resolution to generate the constellation accurately. Although our proposed scheme outperforms the 4D-64PRS only when the studied transceiver impairments are severe (I-Q imbalance of more than 10 degrees or a DAC resolution lower than 4 bits), it still shows an advantage with respect to 4D-2A-8PSK which was implemented in commercially available transceivers as mentioned in [207]. We believe that our proposed scheme merits to be considered as a relevant candidate among geometric-shaped (GS) 4D modulations as recent works are still looking into the design optimization of GS 4D schemes using machine learning tools [208] or simplified computation of AIR [209].

3.3.3 Performance Evaluation

Analysis of the Performance over An AWGN Channel

To easily compare the performance of the four formats, we list the following metrics and features for each format in Table 3.2:

- MSED: the Minimum Squared Euclidean Distance between 4D symbols. Each polarization tributary is normalized to a unit energy.
- n_d: the number of pairs of 4D symbols at MSED.
- The pairs $(d^2_{D_H=1}, n_{D_H=1})$ contain the values of Squared Euclidean Distances (SEDs) for which the Hamming distance (D_H) between the mapped bit patterns is equal to 1 and the number of symbols at those SEDs.

Table 3.2: Metrics for different 4D constellations

	PDM-8QAM-star	4D-2A-8PSK	4D-64PRS	**4D-2A-RS64**
MSED	0.84	0.88	0.69	**0.8**
n_d	192	128	32	**64**
$(d^2_{D_H=1}, n_{D_H=1})$	(0.84, 128) (3.15, 64)	(0.88, 128) (3.46, 64)	(0.69, 32) (0.90, 64) (0.98, 64) (5.50, 32)	**(0.8, 64) (0.94, 64) (1.39, 16) (3.2, 32) (5.46, 16)**
Gray Labeled	No	Yes	Yes	**Yes**
Constant Modulus	No	Yes	Yes	**Yes**
PAPR	1.58	1	1	**1**

- Gray Labeled indicates whether all the constellation pairs at MSED are at a Hamming distance of 1.
- Constant Modulus indicates whether the 4D-energy is constant and PAPR is the peak-to-average power ratio.

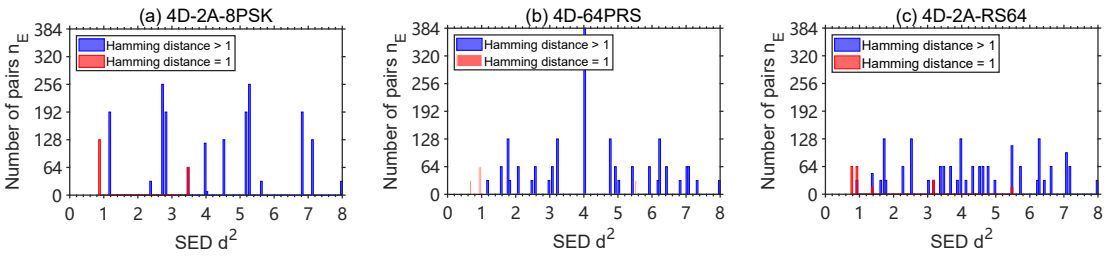


Figure 3.9: Histograms of SEDs in three 4D formats: (a) 4D-2A8PSK, (b) 4D-64PRS, and (c) Our 4D-2A-RS64. The red bars show the number of pairs with Hamming distance of 1 at the SED d^2. The blue bars show the number of pairs with Hamming distance greater than 1 at the SED d^2.

From Table 3.2, we see that PDM-8QAM has the worst linear performance because it has a moderate MSED and an excessive n_d, in addition to the absence of Gray labeling. Among the three 4D formats, 4D-2A-8PSK has the largest MSED thus obtaining the best performance for high SNRs, followed by 4D-2A-RS64, then 4D-64PRS. In addition, among the four schemes, 4D-64PRS has the smallest number of symbol pairs at MSED, n_d, thus obtaining the best performance at low SNR values, followed by 4D-2A-RS64, then 4D-2A-8PSK with an increasing n_d. All three 4D formats are Gray-labeled and have a constant 4D modulus. Finally, to complement Table 3.2, we show in Fig. 3.9 the SED distribution for the three 4D formats. From these histograms and the third row in Tab. 3.2, we see that, although the 4D-64PRS has the smallest MSED, only 32 pairs are located at this distance. Also, the pairs at the three smallest SEDs have a Hamming distance

equal to 1, which provides gains at low SNR values. The 4D-2A-RS64 constellation has a smaller MSED than the 4D-2A-8PSK, but only 64 pairs are located at MSED compared to 128 for the latter. In addition, for its 128 pairs with the smallest SED, 4D-2A-RS64 has 64 pairs at an SED of 0.94 that contribute to a small performance improvement compared to 4D-2A-8PSK at low SNR values. The loss of performance compared to 4D-64PRS is mainly due to the 32 symbol pairs at a SED of 0.94 and with a Hamming distance larger than 1 as can be seen in Fig. 3.9(c).

Performance over The Non-linear Channel

Over long-haul multi-span fiber transmissions, NLI induced by the Kerr effect will add distortions to the propagating signal in addition to the additive ASE noise from the optical amplifiers. This NLI can be reduced by forcing a constant energy constraint on the transmitted signal, hence minimizing the peak-to-average power ratio (PAPR) at the transmitter side which leads to lower NLI generated by both intra-channel and inter-channel propagation effects [185]. We test the performance gain over fiber propagation through numerical simulations. We compare the BERs, system margins, and GMI values of the three 4D formats to the conventional format PDM-8QAM-star, all having an entropy of 6 bit/4D. The transmitted signal consists of 11 WDM channels $\times$ 90 Gbaud, each modulated over dual-polarization. The channel spacing is 100 GHz and the central channel is located at 1550 nm. All channels are pulse-shaped with a root-raised-cosine filter (RRC) with a roll-off factor equal to 0.1. In order to better compare linear gain, nonlinear gain, and gain changes under MZM imbalance, we used an ideal laser with no phase noise.

The tested fibers are standard single mode fibers (SSMF) with a chromatic dispersion coefficient $D = 17$ ps/nm/km, a non-linearity coefficient $\gamma = 1.32$ $(\mathrm{W}\cdot\mathrm{km})^{-1}$, a fiber attenuation parameter $\alpha = 0.2$ dB/km. Polarization mode dispersion (PMD) is added with a PMD coefficient equal to 0.04 ps/$\sqrt{\text{km}}$. The link configuration was a homogeneous multi-span link with $N_{spans} \times 80$ km spans. The propagation was modeled through a split-step Fourier method (SSMF) implementing the non-linear Manakov equation with 50 wave plates per span to simulate the polarization effects [11]. Each span was followed by an erbium-doped fiber amplifier (EDFA) with a noise figure of 5 dB.

After propagation, the central channel is filtered and detected through a coherent receiver. We then apply on the detected signal a matched RRC filter with the same roll-off factor, and digitally compensate for the chromatic dispersion. Genie-aided linear channel equalization is then applied to compensate for the polarization changes prior to decoding (linear channel response is saved and fed to the frequency-domain MMSE channel equalizer). Finally, a genie-aided XPM-induced phase estimation filter with an averaging window of 64 symbols is used to compensate for the XPM-dominated phase rotation as in [210]. With these two perfect estimations, the measured performance is only related to the used modulation format (no penalty caused by imperfect channel

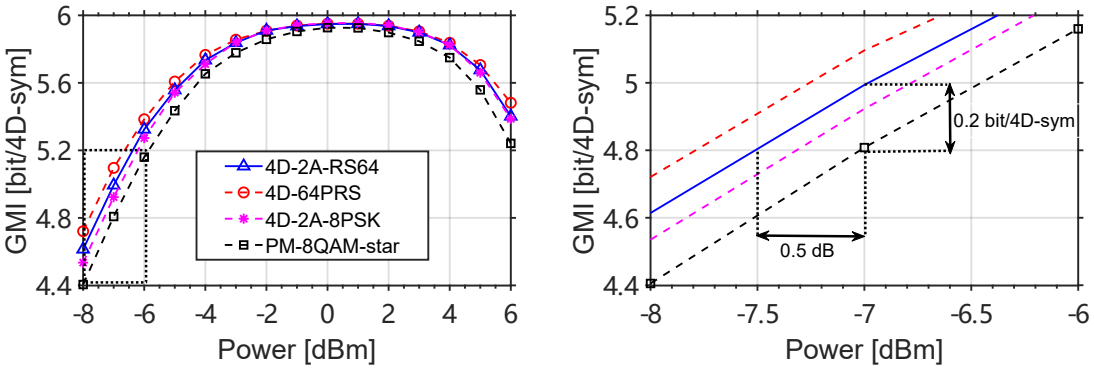


Figure 3.10: GMIs versus transmitted power after transmission over 20 spans of 80 km SSMF in the non-linear regime for 11 channels with 100 GHz spacing with a symbol rate of 90 Gbaud. Left: Overall trends. Right: Zoomed trends around a GMI of 4.8 bit/4D.

estimation). No additional non-linear compensation algorithms such as digital back-propagation were used at the receiver side.

Fig. 3.10 shows GMI curves versus single-channel launch power values after transmission over 20 spans of 80km SSMF. All three 4D formats outperform PDM-8QAM-star in the linear regime as well as in the non-linear regime. At an FEC overhead of 25% corresponding to a net rate of 4.8 bit/4D-symbol, we zoomed in on the corresponding area to better show the achieved gains. 4D-2A-RS64, 4D-64PRS and 4D-2A-8PSK show 0.5 dB, 0.73 dB and 0.25 dB gains, respectively, over PDM-8QAM-star. 4D-2A-RS64 performs better than 4D-2A-8PSK by 0.25 dB. While it does not beat the performance of 4D-64PRS, 4D-2A-RS64 does not require precise geometric shaping to maintain these gains and is less sensitive to implementation errors. At the SNR required for PDM-8QAM-star to reach 4.8 bit/4D-symbol, the GMI of 4D-2A-RS64 is 0.2bit/4D higher than PDM-8QAM-star. We obtain a higher gain in the fiber channel than in the back-to-back case because the 4D modulations are mitigating NLI in the presence of Kerr effect better than PDM-8QAM. It is also worth noting that in the single-channel test, we obtain nearly the same gains (not presented in Fig. 3.10), which is attributed to the constant modulus characteristics of the three 4D formats that effectively reduce the generation of self-phase modulation (SPM).

In Fig. 3.11(a), we show post-FEC and pre-FEC BER curves versus launch power after transmission over 20 spans. We used low-density parity-check (LDPC) codewords of length 64800 bits from the DVB-S2 standard [126] with a rate $R_C = 0.8$ and belief propagation (BP) decoding at the receiver side with 50 decoding iterations for a pre-FEC BER of 4×10^{-2}. The 4D-2A-RS64 gain in post-FEC BER compared to PDM-8QAM-star is in good agreement with the predictions of GMI gain of 0.5 dB in Fig. 3.10. At the optimal power operation point, all three 4D formats have a lower pre-FEC BER than PDM-8QAM-star. 4D-2A-RS64 and 4D-64PRS achieve the same minimum pre-FEC BER while 4D-2A-8PSK has a slightly lower pre-FEC BER value. While the

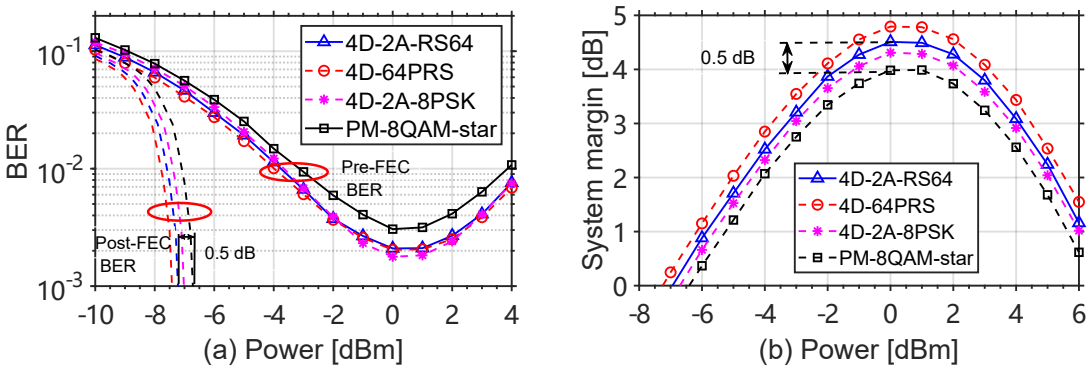


Figure 3.11: (a) Pre-FEC and post-FEC BERs versus transmitted power after transmission over 20 spans of 80km SSMF in the non-linear regime for 11 channels with 100 GHz spacing and a symbol rate of 90 Gbaud. (b) System margin (SNR gain compared to SNR with BER threshold = 4×10^{-2}) versus transmitted power after transmission over 20 spans of 80 km SSMF in the non-linear regime for 11 channels with 100 GHz spacing with a symbol rate of 90 Gbaud.

three 4D schemes have the same constant modulus property at the transmitter side, the lower minimum pre-FEC BER value of 4D-2A-8PSK is mainly due to its highest MSED value among the three schemes.

As the GMI of the three 4D formats saturates around the optimal power in Fig. 3.10, we also show in Fig. 3.11(b) the system margin versus launch power for a target pre-FEC BER of 4×10^{-2}. The system margin is defined as the difference between the required SNR_{elec} to achieve a given target pre-FEC BER and the SNR_{elec} at any launch power of the transmission system over the studied link configuration. SNR_{elec} is the SNR measured from the constellation at the end of the DSP chain at the receiver side (i.e., at the input of the decision circuit). The system margin captures both the linear gain and the non-linear gain brought by the reduced energy variations of the considered 4D schemes. 4D-2A-RS64, 4D-64PRS, and 4D-2A-8PSK show 0.5 dB, 0.8 dB, and 0.3 dB gain in system margin, respectively, over PDM-8QAM-star at the optimal launch power. Although chromatic dispersion will gradually increase the PAPR of the transmitted signal, we still measure gains when the signal has a constant-energy property at the transmitter side. We observe similar performance trends for the four schemes when propagating a single 90 Gbaud channel over the same link.

Finally, the solid lines in Fig. 3.12(a) show GMI curves of the central channel versus the transmission distance given as a number of spans at the optimal launch power for each distance when transmitting 90 Gbaud $\times$ 11 channels. For a net rate of 4.8 bit/4D-symbol, 4D-2A-RS64 increases the covered distance by 400 km relative to PDM-8QAM-star and more than 160 km relative to 4D-2A-8PSK, an increase of 10.6% and 4%, respectively. 4D-64PRS obtains the best performance, allowing transmission over two extra spans compared to 4D-2A-RS64. At the maximum transmission distance of PDM-8QAM-star, which corresponds to 47 spans, 4D-2A-RS64 offers a

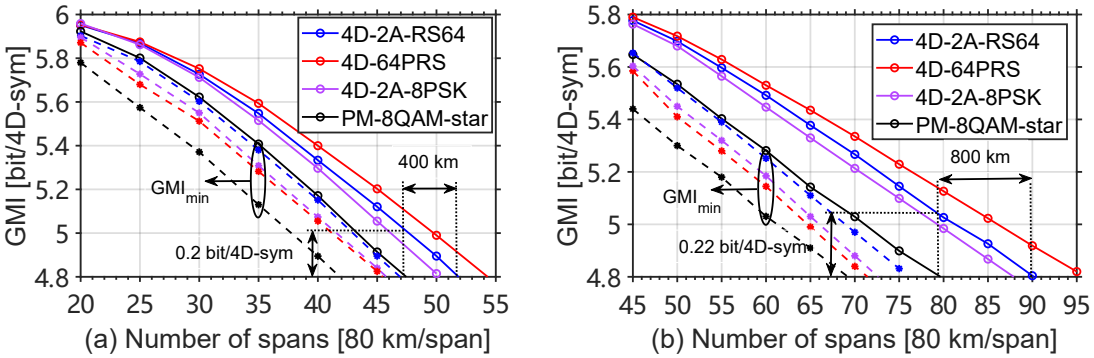


Figure 3.12: (a) GMIs versus the number of spans [80 km/span] for different 4D modulation formats with 6 bit/4D-symbol at 90 Gbaud. Solid lines with empty circles: GMI without MZM imbalances. Dashed line with filled circles: minimum GMI obtained by adding MZM imbalances of $\pm 15^{\circ}$ for the I-Q quadrature and 1.7 dB for the I-Q gain imbalances. (b) GMIs versus the number of spans [80 km/span] for different 4D modulation formats with 6 bit/4D-symbol at 45 Gbaud.

GMI gain of 0.2 bit/4D-symbol which is higher than the linear gain observed in Fig. 3.7(d). The relative additional non-linear gain in terms of GMI for the nonlinear channel with respect to the gain in the linear channel is $(0.2-0.15)/0.15 = 33\%$, which is higher than the relative gains of 4D-2A-8PSK and 4D-64PRS, $(0.13-0.1)/0.1 = 30\%$ and $(0.29-0.26)/0.26 = 11.5\%$ respectively. Next, we add the same worst-case simulated MZM imbalances as the ones used for the linear AWGN channel. The dashed lines in Fig. 3.12(a) show that our 4D-2A-RS64 scheme has the best performance, allowing the transmission over almost 2 extra spans compared to 4D-64PRS and 5 extra spans compared to PDM-8QAM-star.

We also test the performance at a lower channel rate. In Fig. 3.12(b), we show the results for 45 Gbaud $\times$ 11 channels. First, for a perfect transmitter (solid lines), at a GMI of 4.8 bit/4D-symbol, 4D-2A-RS64 increases the covered distance by 800km relative to PDM-8QAM-star and more than 160 km relative to 4D-2A-8PSK, an increase of 12.5% and 2.3% respectively. Since the dispersion-induced inter-symbol-interference grows more slowly over the propagation distance with symbols at 45 Gbaud than at 90 Gbaud, the PAPR has a slower increase over distance. Moreover, the optimal power will also be higher for higher baud rates, hence the non-linear effects will be magnified and will deteriorate the signal even more. If we compare gains in transmission distance for the three 4D constant modulus formats when moving from 45 Gbaud to 90 Gbaud, we notice that 4D-64PRS and 4D-2A-RS64 offer a gain of 8 spans (respectively 5) instead of 16 spans (respectively 10) with respect to PDM-8QAM, showing a gain reduction of 50% while the gain of 4D-2A-8PSK reduces from 8 to 3 spans, showing a gain reduction of 62.5%. Finally, with the simulated worst-case MZM imbalances, our 4D-2A-RS64 also shows the best performance, allowing transmission over 5 extra spans compared to 4D-64PRS and 7 extra spans compared to PDM-8QAM-star.

3.4 Summary

In this chapter, we designed a new 4D constellation, named 4D-2A-RS64, with an entropy of 6 bits/4D, which showed better linear performance and higher fiber non-linearity tolerance than conventional modulation formats PDM-8QAM-star or a recent 4D format, 4D-2A-8PSK, optimized for the non-linear channel. We demonstrated through numerical simulations that the maximum transmission distances are increased for both formats for different symbol rates. While the proposed modulation did not beat the performance of another 4D competitor, the 4D-64PRS format, we demonstrated that 4D-2A-RS64 imposes less stringent constraints on the transmitter compared to the other 4D formats. Hence, its generation using non-ideal components (extreme imbalances in MZMs and limited-resolution DACs) shows smaller performance degradation than state-of-the-art 4D formats such as 4D-2A-8PSK and 4D-64PRS in both linear and non-linear transmission regimes. In the design of future high-order modulation schemes with higher spectral efficiency, we can still optimize linear performance by increasing MSED and tailoring the distribution of all squared Euclidean distances. For the nonlinear part, it is crucial to maintain temporal energy fluctuations as low as possible. Furthermore, higher spectral efficiency often requires more symbols and more precise mapping, which means that sensitivity to MZM imbalance and DAC variations will increase. Extending the proposed solution to achieve even higher spectral efficiency by using more rings remains an open problem for future exploration. In addition, the design of signaling schemes that are optimized for the non-linear fiber channel using both multi-dimensional modulations and probabilistic shaping could help in achieving additional linear and non-linear gains as it has been suggested in [1, 2]. We will further explore this axis in the next chapter.

Chapter 4

Probabilistic Constellation Shaping with reduced NLI

4.1 Introduction

This chapter considers the evolution of Probabilistic Constellation Shaping (PCS) that leads to reduced nonlinear interference (NLI) during propagation, focusing on the groundbreaking probabilistic modulation schemes that have emerged, and further introduces the development of our two novel designs for PCS with enhanced tolerance to nonlinear Kerr effects.

Expanding on the progress achieved in our previous work and drawing inspiration from recent advances in distribution matching solutions for PCS, we have developed new distribution matching (DM) schemes. These schemes aims at mitigating the NLI induced by Kerr nonlinear effects in optical fiber transmission systems. Our design approaches revolve around reducing temporal energy variations in the transmitted signal to minimize nonlinear effects while prioritizing practicality. The solutions we have explored are based on standard QAM constellations multiplexed over multiple dimensions. Additionally, we prioritize practicality in encoding and decoding processes for modulation, shaping and sequence selection, aiming to avoid the reliance on lookup tables for encoding and decoding operations.

Firstly, we have devised a multi-dimensional probability shaping scheme called PCS-4D-128SP-16QAM, based on set partitioning and the Constant Composition Distribution Matcher (CCDM). This approach combines PCS, utilizing a Maxwell Boltzmann distribution, with multi-dimensional modulations. The goal is to enhance performance gains in both the linear and nonlinear domains of optical fiber transmission systems. By employing set partitioning and energy constraints on Four-Dimensional (4D) modulations in PDM transmissions, followed by a 4D PCS, we provide insights into constructing novel modulations that exhibit superior performance compared to conventional PCS constellations, where shaping is performed separately over each real dimension.

Our subsequent plan focuses on the Enumerative Sphere Shaping (ESS) technology which reduces rate loss, compared to CCDM, using short block lengths, thereby gaining nonlinear benefits from shorter block lengths as observed in [149]. Furthermore, we apply PCS over 64QAM constellations to boost superior spectral efficiency, however with a broader range of amplitudes and higher energy fluctuations compared to PCS-16QAM. When it comes to long-distance transmission, reducing the Kerr-induced NLI of PCS-64QAM aligns more closely with the advancements in practical applications, yielding more significant benefits than the ones observed until now on PCS-16QAM. We propose a 4D-energy-limited ESS for M-QAM signaling. Our approach is designed to minimize rate loss and significantly improve transmission performance in nonlinear WDM optical fiber systems. Our simulations, conducted using 64QAM, show that our proposed scheme outperforms conventional ESS by a remarkable 0.19 bits per 4D-symbol in terms of achievable information rate over a 205-km single-span link and a WDM transmission setup with five PDM channels, each operating at a net rate of 400 Gbit/s. Furthermore, we conduct a comprehensive analysis of the achieved performance across various shaping block lengths, revealing that the achieved gains do not scale well for multi-span systems.

In section 4.2, we present the state-of-the-art in distribution matching (DM) design for achieving both linear and nonlinear gains. Subsequently, in section 4.3, we introduce our design for the combination of multi-dimensional set-partitioning and CCDM, that yields the PCS-4D-128SP-16QAM scheme. In section 4.4, we introduce our design of a nonlinear tolerant ESS, the 4D band-limited ESS (4D-BL-ESS).

4.2 State-of-the-art of PCS Techniques with Reduced NLI

As we introduced in chapter 3, multi-dimensional modulations and geometric shaping (GS) were widely studied in the linear and nonlinear noise transmission regime. On the other hand, PCS was also proposed since 1993 in [164] as an alternative scheme to approach the capacity of an AWGN channel through the transformation of the symbol probability distribution into a Maxwell-Boltzmann (MB) distribution. PCS can easily achieve MB-like distributions and approach the Shannon limit, offering significant advantages over conventional 2D constellations when communicating over a linear AWGN channel.

We have seen in Chapter 2 that we can approach an MB distribution using CCDM [22], a competitive candidate to implement PCS over rectangular or square M-QAM, showing near-optimal linear shaping gains at long block-length [211]. However, the increased peak-to-average power ratio (PAPR) and higher standardized moments of PCS schemes will trigger more nonlinear Kerr-induced distortions compared to unshaped uniform modulations. More recently, several versions of PCS and DM were suggested to either further enhance the performance compared to CCDM

(in terms of rate loss or robustness to channel effects) or to make the shaped signal better suited to the conventional DSP chain while reducing the implementation complexity. For instance, truncated PCS based on PDM-64QAM was designed by suppressing the maximum-energy symbols to make the scheme better suited to the adaptive multiple-input multiple-output (MIMO) equalizer for polarization separation in [212].

Hybrid schemes employing both GS and PCS were also investigated by different research groups [152, 153]. In this section, we mainly focus on PCS and different recent implementations and adaptations to further improve linear performance and nonlinear performance by discarding, through additional design constraints, symbol sequences with high temporal energy variations. PCS can be applied independently on single dimensions or in a multidimensional space. Over a single dimension, a non-uniform channel input distribution can be obtained by transforming a uniform information sequence into a shaped amplitude sequence while satisfying a predefined condition (MB distribution or other channel-aware distribution [213, 214]). This operation is an invertible mapping that can be realized through different approaches. On the other hand, if we consider the transmission of N amplitudes where each amplitude is taken from a set $\mathcal{A}$, PCS can be applied over these N dimensions to select a subset of allowed sequences from $\mathcal{A}^N$. In [215], a multi-dimensional DM was applied to distribute the probability of amplitudes over multiple dimensions and enhance the linear performance compared to separate 1D distribution matching of amplitudes over the available dimensions.

As for the distribution matching based on ESS, several works aimed at improving the nonlinear performance by limiting the 1-dimensional (1D) amplitude variance or kurtosis of ESS schemes [216, 217], showing a nonlinear gain over single-span transmissions. We will explore some of the most promising CCDM-based and ESS-based schemes in the following subsections.

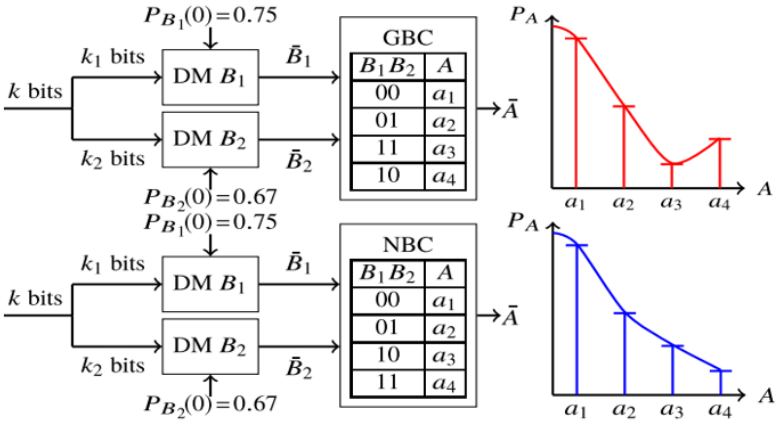


Fig. 4. GBC and NBC mappings used inside the BL-DM.

Figure 4.1: Comparison of Gray binary coding (GBC) and natural binary coding (NBC) mappings for the bit-level DM. Source: [30].

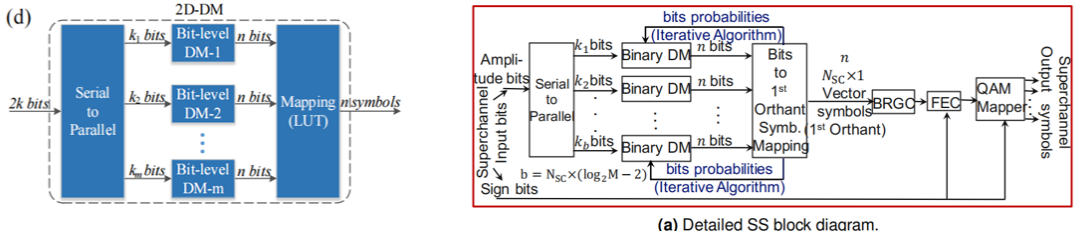


Figure 4.2: Multi-dimensional optimizations of bit-level DM: on the left, over 2 dimensions [31] and on the right, over 4 dimensions [32].

4.2.1 Bit-level CCDM

Bit-level CCDM, as referenced in [30, 218–220], is a special CCDM implementation tailored for each of the $m = \log_2(M)$ bits of an M-QAM format using natural binary coding, as illustrated in Figure 4.1. It is achieved through concurrent DM blocks with binary output alphabets, followed by a mapping of these binary outputs to symbols from the original non-binary alphabet. This approach of bit-level distribution matching significantly reduces rate loss in scenarios with short block lengths compared to symbol-level CCDM, thereby enhancing the overall throughput of the transmitter.

Within bit-level CCDM, 2D-DM structures with optimized probabilities and mapping [31] and 4D-DM structures with optimized probabilities and mapping [32, 221] were suggested. For this multi-dimensional version, an iterative algorithm is used to optimize each bit probability distribution and the bit-to-symbol mapping book while targeting the minimization of the root mean square (RMS) power of the multi-dimensional constellation. The probability distributions and the bit-to-symbol mapping book are stored as a look-up table for the implementation of the encoder and the decoder. Compared to the 2D bit-level DM and the 1D bit-level DM, the 4D bit-level DM improves linear and nonlinear performance by reducing RMS power and also reducing the rate loss. Block diagrams of these DM techniques are shown in Figure 4.2.

4.2.2 Modified Multi-set Partition Distribution Matching (MPDM)

In CCDM, each output sequence from the DM is characterized by a consistent composition C_0 and probability distribution $P_{\tilde{A}}$, hence all sequences belong to the same set. This is in stark contrast to the approach taken in MPDM, where output sequences belong to different sets with a specific probability distribution $P_{\tilde{A}_l}$, or equivalently a composition C_l. Each composition C_l has a complement $\bar{C}_l$ such that their combined average gives the target composition C_0. MPDM uses many of these pairs and arranges them in a binary tree such as the one shown on the left of Fig. 4.3. Such a methodical approach in MPDM not only minimizes the rate loss but also effectively maintains control over the average distribution.

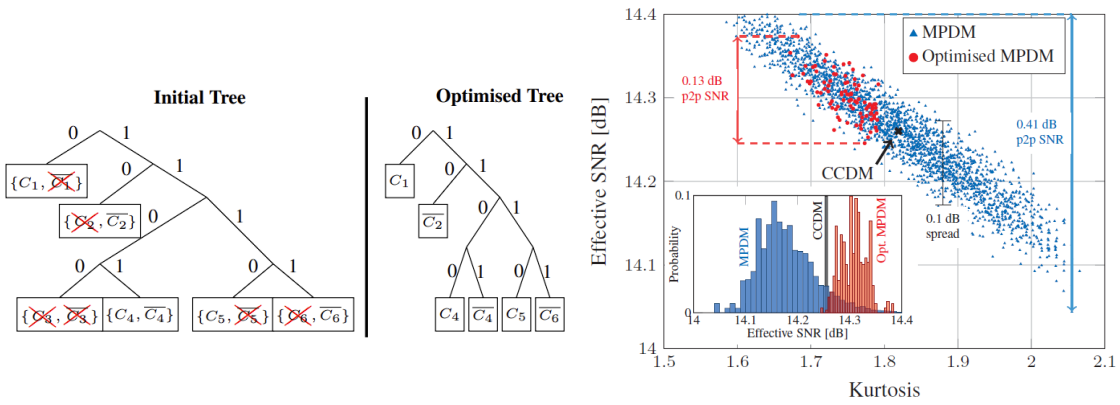


Figure 4.3: Comparison of MPDM and Optimised MPDM. Left: Illustration of MPDM and Optimised MPDM tree, the optimized tree is obtained by eliminating the crossed-out compositions of the initial tree that have high NLI. Right: Effective SNR versus kurtosis for CCDM, MPDM and the NLI-optimised MPDM. Inset: Weighted histograms. Source [33].

The same authors who designed MPDM suggested a modified version of it optimized for the nonlinear propagation regime [33]. on the left of Fig. 4.3, by eliminating the crossed-out compositions of the initial tree that have high kurtosis and high NLI. The new distribution matching slightly improves the average and worst-case SNR as shown on the right of Fig. 4.3. As a performance assessment tool, they have used the Enhanced Gaussian noise (EGN) model to demonstrate that the per-block SNR after fiber transmission varies significantly due to the variable-composition nature of the MPDM shaping scheme.

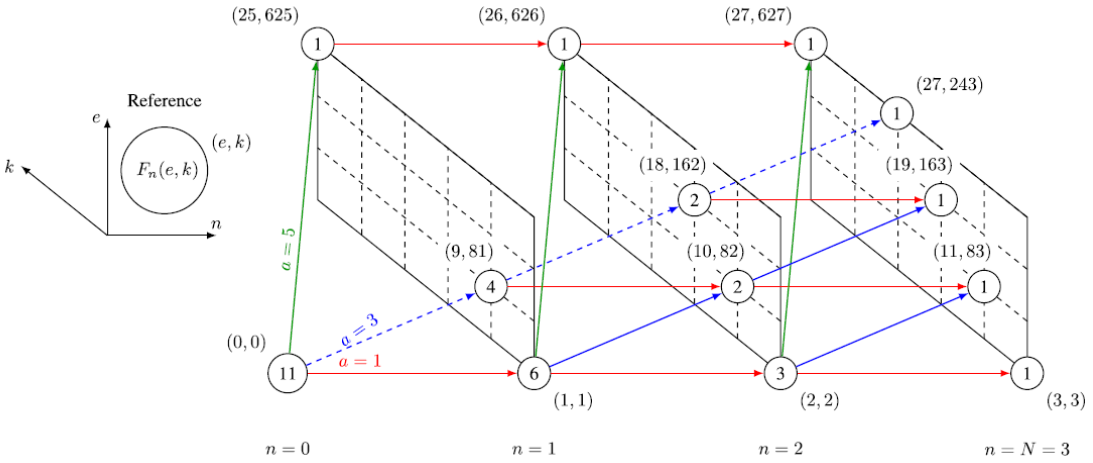


Figure 4.4: Bounded-energy and bounded-kurtosis enumerative amplitude trellis for 3 amplitude levels, $N = 3$, $E_{max} = 27$ and $K_{max} = 627$. Source: [34].

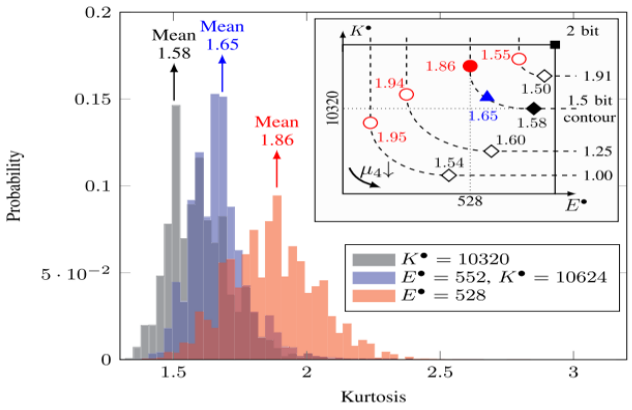


Fig. 3. Histograms of kurtosises of sequences for $N = 64$ with 64-QAM. Kurtosis of a sequence a^N is computed using (2) with the corresponding empirical distribution. **Inset:** The $E^\bullet K^\bullet$-plane with $k/N = 1.00, 1.25, 1.50$ and 1.91 b./amp. shaping rate contours. Here, the black square effectively represents uniform signaling where there is no constraint on energy or kurtosis, i.e., the signal space is effectively bounded by an N-cube.

Figure 4.5: Histograms of kurtosis distributions for different ESS schemes. Source: [34].

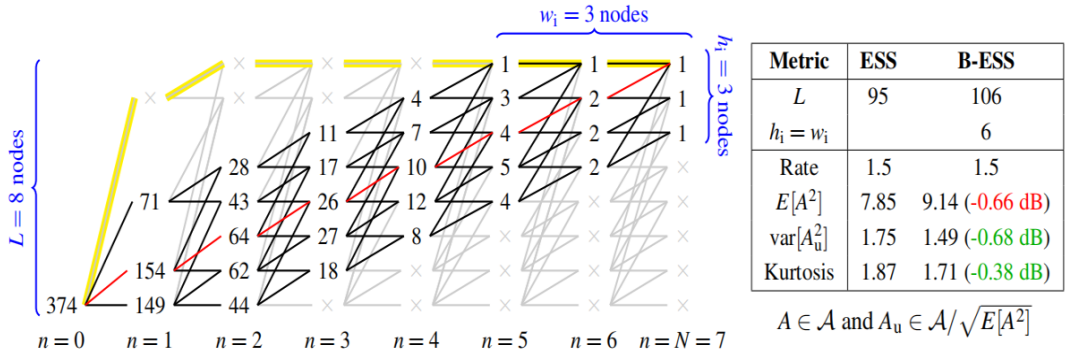
4.2.3 Kurtosis-limited ESS

The kurtosis-limited ESS (K-ESS) [34] limits the kurtosis of the generated ESS-shaped sequences to K_{max}, however, the average energy is increased in order to achieve the same spectral efficiency as the conventional ESS. Figures 4.4 and 4.5 show respectively the trellis representation of a K-ESS with specific values of the maximum energy E_{max} and maximum kurtosis K_{max}, and the histograms of the kurtosis values of ESS schemes with and without kurtosis limitations. When a constraint in K-ESS limits the kurtosis, as the maximum value of the kurtosis decreases, the maximum value of the symbol energy must increase to achieve the same rate (number of bits per amplitude). Thus, K-ESS has a penalty in the linear regime, but achieves a gain in the nonlinear regime. Finally, K-ESS cannot be used for PCS-16QAM schemes because of the one-to-one correspondence between energy and kurtosis for 16QAM modulation (the kurtosis does not bring additional a new degree of freedom to 16QAM to perform a sequence selection).

4.2.4 Band-trellis ESS

The band-trellis ESS (B-ESS) [35] reduces the energy variations (hence also the kurtosis) of the generated sequences by limiting the width and height of a band on the trellis representation of the ESS scheme as shown in Figure 4.6. Instead of adding a kurtosis limitation, the band-trellis ESS modulation reduces the power variations by directly limiting the trellis over which the scheme is defined by cutting the upper and lower parts of the trellis and keeping an energy band.

Compared to conventional ESS, K-ESS and B-ESS trade the linear gain for a nonlinear gain. They have demonstrated interesting trade-offs over short distances (up to 205 km) and for PCS-



Metric	ESS	B-ESS
L	95	106
$h_i = w_i$		6
Rate	1.5	1.5
$E[A^2]$	7.85	9.14 (-0.66 dB)
$\text{var}[A_u^2]$	1.75	1.49 (-0.68 dB)
Kurtosis	1.87	1.71 (-0.38 dB)

$A \in \mathcal{A}$ and $A_u \in \mathcal{A}/\sqrt{E[A^2]}$

Fig. 2. Using $\mathcal{A} = \{1,3,5,7\}$: **(Left)** Band-trellis for $E_{max} = 63$. **(Right)** ESS vs. B-ESS at $N = 108$.

Figure 4.6: Illustration of a band-trellis ESS scheme: on the left, trellis tree of a band-trellis ESS; on the right, a comparison between ESS and B-ESS. Source: [35].

64QAM modulations. B-ESS can be applied to a 16QAM-based transmission scheme but the rate loss is very high and the nonlinear gain is quite limited.

Finally, a single recent work was found in the literature addressing both digital multi-band (DMB) transmission and nonlinear shaping to mitigate NLI distortions. In [222], the nonlinearity-tolerant bisection-based-SS (BS-SS-NLI) is proposed based on the parallel bisection structure [223] achieving a lower average energy and rate loss than K-ESS at the same block length by selecting all preferential sequences. The demonstration is done using single-channel 16 subcarriers, 96 Gbaud per channel with dual-polarization 64QAM for a multi-span long-haul transmission over 2500km of SSMF. The new scheme can provide 0.28 dB SNR_{elec} gain and 0.1 bit/4D-symbol achievable information rate improvement compared to ESS.

4.2.5 Mapping of Distribution Matching Outputs over Multiple Dimensions

ESS-based distribution matching (or other DM techniques) can be generated and applied over 1, 2, 4, or even more dimensions. In Fig. 4.7 from [224], the authors show an illustration of 1D, 2D, and 4D symbol mapping strategies for ESS schemes. For 1D mapping, 4D symbols are shaped using 4 independent amplitude sequences (one ESS per dimension). For 2D mapping, each polarization, namely X and Y, uses one ESS-shaped amplitude sequence. For 4D mapping, a single-shaped sequence is used for a PDM wavelength channel. The latter mapping offers the advantage of reducing the probability of having equal amplitudes at the same time slot. Hence, high peak power values appear less frequently and lead to lower nonlinear distortions. This advantage was demonstrated in several works [216, 224, 225].

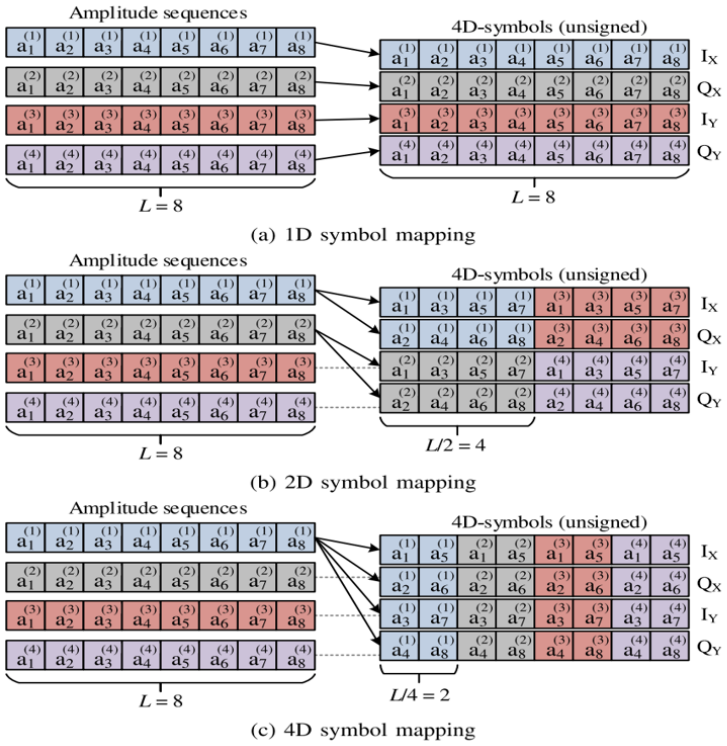


Fig. 3. Strategies for mapping amplitude sequences into modulated 4D symbols (corresponding signs are not reflected): (a) 1D symbol mapping, (b) 2D symbol mapping, (c) 4D symbol mapping.

Figure 4.7: Strategies for mapping PCS shaped sequences over multiple dimensions. Source: [2].

4.2.6 Discussion and Conclusions from the Different Shaping Strategies

CCDM with a long block length can be set as an upper bound during the comparison of the performance of different DM schemes in the linear noise regime. However, in practice, to limit the implementation complexity and harness gains in the nonlinear regime, short-block-length ESS becomes a more interesting option thanks to its low rate loss compared to CCDM.

Subsequently, we have seen that multidimensional distribution matching brings more degrees of freedom to optimize the probability distribution of amplitudes helping to harness both linear and nonlinear gains. All studied approaches target a reduction of energy variations to achieve nonlinear shaping gains.

ESS, MPDM, and bit-level DM are mostly beneficial in reducing the rate loss when using short block lengths (blocks of a few hundred amplitude values). By adding energy constraints to limit energy variations through the limitation of the kurtosis, the RMS power or the energy dispersion index (EDI) of the generated amplitude sequences defined in [107] (reminded in Eq. (5.2) and further explored in the next chapter), we can improve the linear and nonlinear performance. Almost all schemes showed gains over single-span links and/or single-wavelength transmissions.

The design challenge of game-changing PCS schemes is striking a balance between reduced rate loss and higher tolerance to Kerr effects for long-haul multi-span transmission systems.

It is worth noting that all the works from the literature emphasize the following factors that allow for the best trade-off between linear and nonlinear shaping gains:

- Shaping is most effective when performed over higher order modulation formats (more amplitude levels in 1D) and hence gains will be more easily reached over short distances where high optical signal-to-noise ratios can be achieved.

- For the mitigation of nonlinear phase noise (NLPN), shaping is best suited for single-span systems where NLPN dominates. Furthermore, in these systems, the correlation time that characterizes NLPN is too short for an adaptive phase recovery to be effective, but it is long enough to ensure the effectiveness of a shaping scheme [226, 227].

- For nonlinear fiber links in which the contribution of the modulation-dependent part of NLI is significant (those with in-line dispersion management (not covered in this chapter) or single-span links with high power), optimizations of the shaping scheme can be beneficial for a large nonlinear gain [1].

- Additionally, K-ESS and B-ESS are both limited to the 1D amplitude energy variation and kurtosis, but map the 1D amplitude by the 4D mapping. Instead of shaping on a per 1D amplitude or 2D symbol basis, constellation shaping over 4D or over several time slots to exploit the temporal correlations introduced by XPM can be an interesting approach to enhance performance.

Furthermore, there is no consensus on an adequate performance metric to evaluate the minimization of NLI when transmitting short-block-length shaped schemes apart from computing NLI through a complete split-step-Fourier-method (SSFM) based propagation [3]. This metric is accurate, albeit very complex to be considered for a real-time implementation in an optical transmitter. While the energy dispersion index (EDI) was mentioned earlier as a possible metric, windowed moment of the symbols were also proposed in other works. We will further discuss EDI and the windowed metrics in Chapter 5. However, we explain the motivation of using these metrics in a nutshell: since the statistical correlation of the shaped symbols cannot be explained by conventional models such as the EGN model that was designed for independent identically distributed (iid) signaling, nor by conventional moments of modulation schemes, several works focused on using windowed metrics such as windowed moments or EDI as metrics to evaluate the NLI-generation capability of a sequence in place of the conventional moments [222, 228]. Yet, these metrics were still computed at the transmitter side on the ideal undistorted sequences.

An observation that is also worth noting from [228] is related to the optimal symbol rate and optimal shaping block length for non-iid signaling for single-carrier and multi-carrier transmissions based on digital multibands (DMB). Indeed, the authors found that the optimal symbol rate for multi-span transmissions deduced from the EGN model is not valid for non-iid signaling. As a reminder, this rate is given by [14]:

$$R_{opt} = \sqrt{\frac{2}{\pi |\beta_2| L_s N_s}} \tag{4.1}$$

where β_2 is the group velocity dispersion parameter, L_s is the span length and N_s is the number of spans. Instead, the authors in [228] observed that the optimal symbol rate decreases when the block length increases. Besides, the optimal block length increases when the accumulated dispersion increases. Hence, they conclude that the best nonlinear-tolerant signaling should be designed by optimizing the temporal and spectral structure of the waveform for a given transmission link.

Finally, researchers also emphasized the importance of considering the interplay between the designed shaped schemes and the DSP algorithms that can partially remove NLI distortions such as the capability of CPE algorithms in partially mitigating NLPN as observed in [176–178].

4.3 4D-CCDM Design on Set-Partitioned Constellations

Based on the research summary in the previous section, we propose a first design that applies PCS over a 4-dimensional set-partitioned modulation defined according to energy constraints to target both enhanced spectral efficiency and reduced NLI. Set partitioning helps in achieving higher MSED and the chosen energy constraints will help limiting the PAPR of the constellation. The obtained simulation results inform us that a better trade-off must be found between the achievable linear and nonlinear gains in order to make the best of this new construction.

4.3.1 PCS Modulation Design over 4D

In a nut shell, the scheme consists in a probabilistically shaped 4D set-partitioned constellation. We start building the modulation scheme with a 4D modulation consisting of two square M-QAM symbols, one for each polarization tributary of the PDM signal. For the sake of low complexity, we can restrict the M-QAM constellations to rectangular or square QAMs defined over $\mathbb{Z}^2$. To increase the minimum squared Euclidean distance (MSED) of these QAM constellations, we apply set partitioning to the constellation.

We illustrate the construction with a square 16-QAM ($M = 16$) and define the set-partitioned modulation as follows: a 4D symbol is obtained by combining 2D symbols from two different

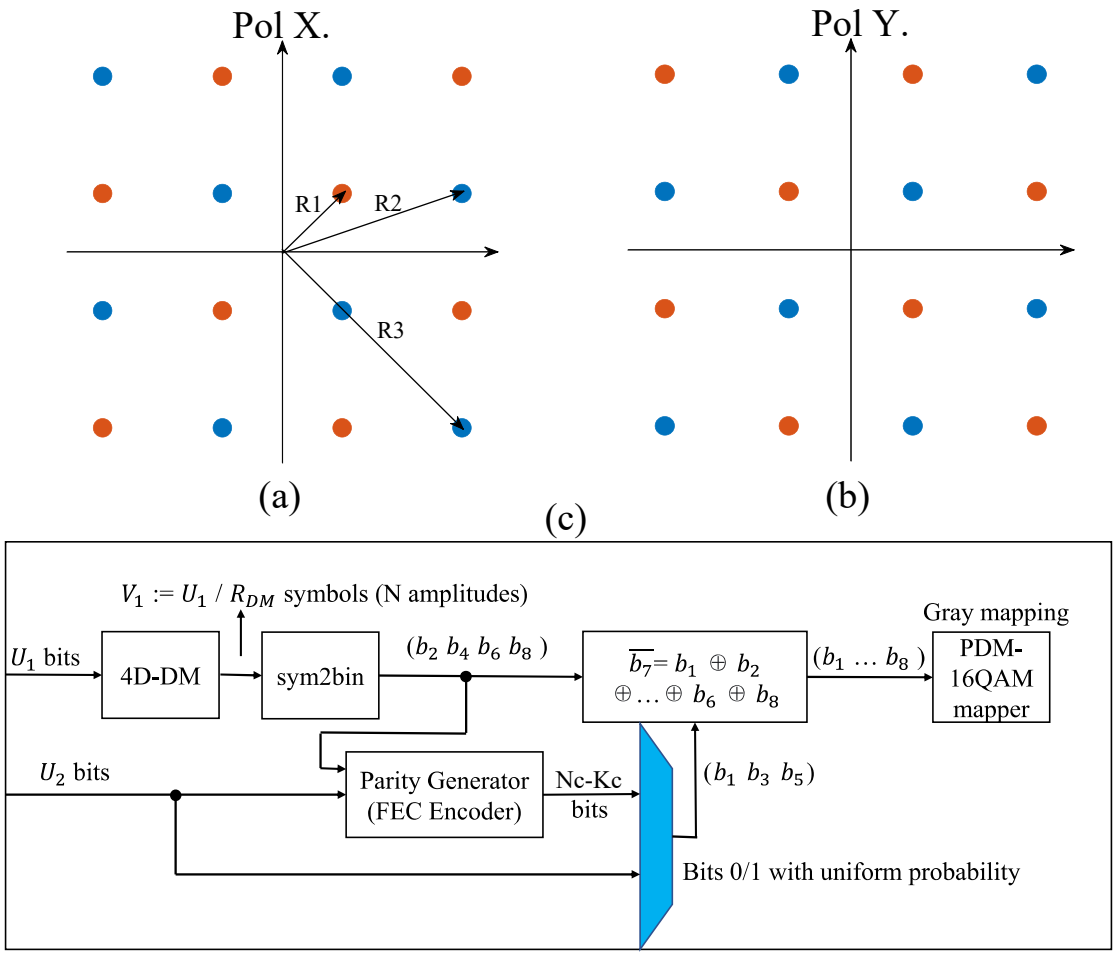


Figure 4.8: (a,b) 2D-projections of the 4D 128SP-16QAM in Pol X. (a) and Pol. Y (b). The three arrows represent the three radii values. (c) Block diagram of the PCS 4D CCDM transmitter with the FEC encoder

subsets, shown in different colors in Fig. 4.8(a-b), and assigning them to two orthogonal polarization tributaries. Each subset contains $8^2 = 64$ 4D symbols. All in all, we get 128 combinations of 4D symbols, and the MSED is two-fold increased compared to the one of PDM-16QAM. We call this new constellation 128SP-16QAM which carries $\log_2(128) = 7$ bits per 4D symbol. In the general case of a set-partitioned square M-QAM with a single level of partitioning, the number of bits per 4D symbol will be $2\log_2(M) - 1$.

From this 128SP-16QAM constellation, we can select a subset of symbols by applying constraints either on the desired 4D energy levels or on the desired combinations of 1D energy levels. In the former case, we can reduce the peak-to-average power ratio (PAPR) by removing high-energy symbols. In the latter case, we can remove selected 1D-energy combinations to reduce the number of neighbors at the MSED. For instance, following the first rule, we can define a constellation with 88 4D symbols by removing the 4D symbols containing a pair of 2D symbols laying both on the radius R_3 and the ones with a 2D symbol on the radius R_2 and another 2D symbol on

the radius R_3, where R_1, R_2 and R_3 are shown on the 2D projection in Fig. 4.8(a). This constraint reduces the number of 4D energy levels from five to three by removing the two highest levels. This leads to an entropy of $\log_2(88) \approx 6.46$ bits/4D symbol. We call this constellation 88SP-16QAM. On the other hand, following the second rule, we can remove all symbols of an intermediate 4D energy level or alternatively, all symbols corresponding to a given combination of 1D amplitudes.

As a next step in the design, we apply PCS over the 4D constellation using a Maxwell-Boltzmann distribution such that the probability of occurrence of the 4D energy values decreases from the lowest to the highest 4D energy. For the distribution matching of the conventional PCS scheme, referred to as PCS 2D in this section, we choose constant composition distribution matching (CCDM) [22] that involves a low-density-parity check (LDPC) encoder that encodes K_c input bits to N_c output bits. For a regular PCS 2D scheme with a square 16-QAM in each polarization, distribution matching can be implemented in parallel over each of the four real dimensions. For each real dimension, the output of the DM has 2 amplitude levels $\{1,3\}$ to which we associate two probabilities from the desired probability distribution as done in [22]. In Fig. 2.14 in Chapter 2, an implementation of a PCS transmitter for a single quadrature was described including a 1D distribution matcher and a forward error correction (FEC) encoder. The total net rate of the PCS 2D CCDM transmitter is $R_T = 4\left(U_1 + U_2\right)/V_1 = 4R_{DM} + 8R_C - 4$ bits per 4D symbol where U_1 and U_2 are bit streams at the input of each 1D transmitter, R_{DM} is the rate of the 1D distribution matcher (each 1D DM encodes U_1 bits into $V_1 = U_1/R_{DM}$ symbols), and $R_C = K_c/N_c$ is the rate of the embedded FEC encoder that encodes the bits coming from the U_2 bits and the Gray-mapped V_1 symbols. The formula for the total net rate does not take into account the rate loss introduced by the finite-length distribution matching, which for a practical implementation slightly modifies the length of U_1.

Now, we need to design a constant composition distribution matcher and a bit-to-symbol mapping for the 4D shaped 128SP-16QAM constellation and its variants (subsets of this constellation). To achieve this, we consider a bit mapping based on a Gray-mapped PDM-16QAM in order to avoid the use of a look-up table. We separate the 8 bits of a PDM-16QAM symbol into 4 sign bits (b_1, b_3, b_5, b_7) and four amplitude bits (b_2, b_4, b_6, b_8). In Fig. 4.8(c), we show the structure of the proposed PCS 4D transmitter that is inspired by the CCDM transmitter in Fig. 2.14. The major difference between the two transmitters is that the initial PCS transmitter for polarization division multiplexed systems consists of 4 parallel 1D PCS transmitters whereas our proposed PCS 4D transmitter performs distribution matching directly over the 4 dimensions.

The U_1 bits generate $V_1 = U_1/R_{DM,4D}$ amplitude combinations through a 4D distribution matcher (4D-DM) with rate $R_{DM,4D}$ bits per 4D symbol. Each of these V_1 amplitude combinations is then Gray-demapped into four amplitude bits, which generates $4V_1$ bits in total. For the different variants of the PCS 4D 128SP-16QAM constellations, the output of the DM can

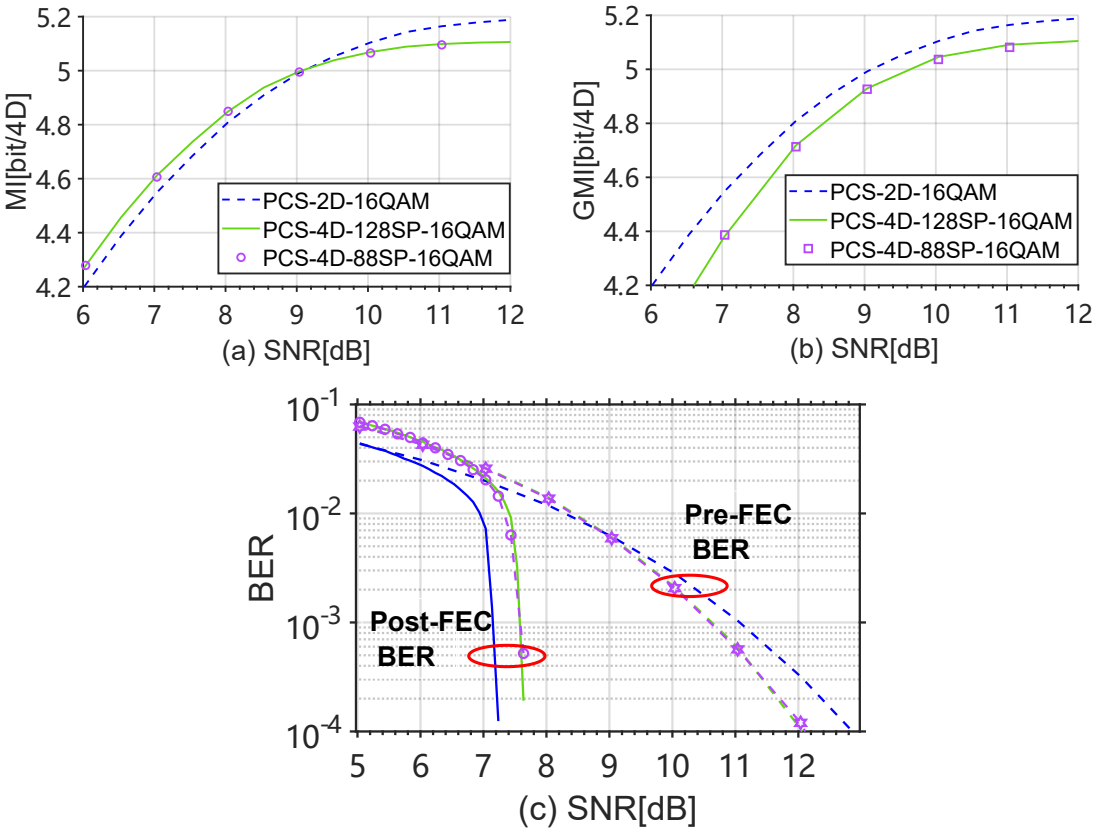


Figure 4.9: Performance over AWGN channel: (a) MI versus SNR for PCS schemes optimized for a net rate of 4.4 bits/4D with $R_C = 0.9$, (b) GMI vs SNR, (c) Pre-FEC BER and post-FEC BER vs SNR (PCS-2D-16QAM in blue, PCS-4D-128SP-16QAM in green, PCS-4D-88SP-16QAM in purple).

be one out of $N_{out} \leq 2^4 = 16$ combinations of 1D amplitudes $\{1,3\}$ with the desired probability distribution. In particular, the 128SP-16QAM constellation has $N_{out} = 16$ combinations of 1D amplitudes and the 88SP-16QAM has $N_{out} = 11$ combinations. The $4V_1$ bits are combined with U_2 uniformly distributed bits as the input of a (K_c, N_c) LDPC encoder. Then, the generated $N_c - K_c$ parity bits are combined with the U_2 bits to form the sequence of sign bits (b_1, b_3, b_5). Finally, b_7 is generated as $\overline{b_7} = b_1 \bigoplus b_2 \bigoplus \ldots b_8$ where $\bigoplus$ is the XOR operator, and the eight bits are mapped to a PDM-16QAM symbol. From the described structure, we find that $K_c = U_2 + 4V_1$ and $N_c - K_c + U_2 = 3V_1$ to have enough sign bits for the V_1 amplitude combinations. Hence, our transmitter receives $U_1 = \lfloor N_c R_{DM,4D}/7 \rfloor$ and $U_2 = \lfloor K_c - 4N_c/7 \rfloor$ bits at its input where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. The total net rate of this 4D transmitter is $R_T = (U_1 + U_2)/V_1 = R_{DM,4D} + 7R_C - 4$ bits per 4D symbol. As a last note, for the 4D 128SP-16QAM (without applying additional constraints), as we suppress none of the amplitude combinations, we can use the same 1D distribution matching as for the conventional PCS 2D.

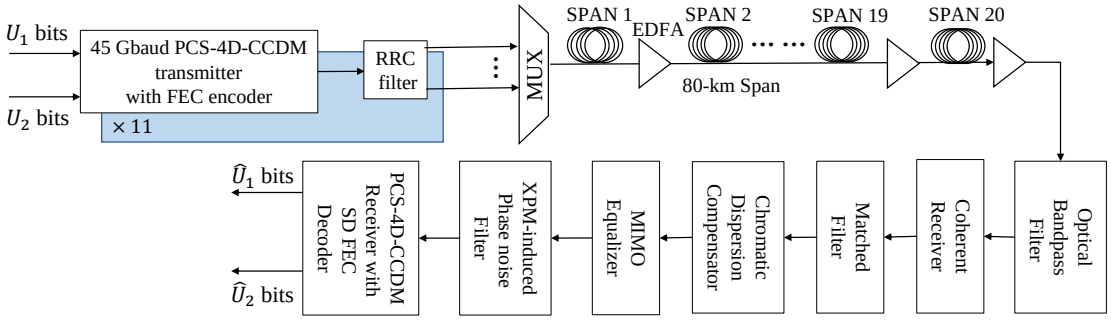


Figure 4.10: Schematic of the simulation setup with 4D PCS transmitter and receiver

4.3.2 Performance over an AWGN Channel

In Fig. 4.9, we show respectively in (a), the mutual information (MI) defined and computed as Eq. (1.31) that we adapt to 4D modulations, and in (b), the generalized mutual information (GMI) that takes into account the bit mapping, computed as Eq. (1.34) and that we also adapt to 4D modulations. These information rates are plotted versus the symbol-wise signal-to-noise ratio (SNR) for two PCS 4D schemes, the 128SP-16QAM (with five 4D energy levels) and the 88SP-16QAM (with three 4D energy levels) defined in the previous section. The performance is measured over an additive White Gaussian noise (AWGN) channel.

We compare the two 4D schemes to a conventional PDM PCS-2D-16QAM (blue curve). The shaping for the PCS-4D schemes was optimized by targeting a maximization of their generalized mutual information (GMI) at an SNR around 7 dB. The optimization is done by testing different combinations of R_C and $R_{DM,4D}$. The results in Fig. 4.9 are given for PCS 4D transmitters using an LDPC code of length $n = 64800$ bits from the DVB-S2 standard with a rate $R_C = 0.9$. For the PCS 4D schemes, the highest achievable GMI is 4.4 bits/4D-symbol for an SNR of 7 dB and it is obtained with a DM rate $R_{DM,4D} = 2.1$ bits/4D. The CCDM block length of PCS 4D is hence $V_1 = 9257$ symbols ($U_1 = 19360$ bits if the rate loss is zero; in practice for $n = 64800$, $U_1 = 19469$), $U_2 = 21291$ bits and the rate loss per 1D amplitude is 3.2×10^{-3} bit. The rate loss is computed as Eq. (2.10). For the conventional PCS 2D transmitter, we use the following parameters: $U_2 = 25920$ bits, $V_1 = 32400$ symbols (hence $U_1 = 9720$ bits if the rate loss is zero; in practice for $n = 64800$, $U_1 = 9732$), $R_{DM} = 0.3$ bits/1D and the rate loss is 4×10^{-4} bit per 1D amplitude.

At an SNR of 7 dB, we notice that the mutual information is indeed the highest for the PCS 4D schemes. However, if we look at their generalized mutual information, the GMI deteriorates compared to the conventional scheme. On one side, these behaviors can be explained by the increased MSED of the two PCS 4D schemes with a set-partitioned PDM-QAM compared to a conventional PCS 2D over a non-partitioned QAM. On the other side, the 4D modulations have a larger number of symbol pairs at MSED which cannot fit a Gray mapping for all the 4D neighbors. This sub-optimal bit mapping explains the degraded GMI.

In Fig. 4.9 (c), we validate the observations on the achievable information rates through bit-error-rate (BER) curves. To decode the FEC code, we use belief propagation (BP) decoding at the receiver side with 50 decoding iterations. The pre-FEC BER curves show the MSED advantage at high SNR and the post-FEC curves show the 0.4 dB penalty due to the sub-optimal bit mapping. This penalty can be suppressed by using joint iterative demapping and decoding or non-binary LDPC codes at the cost of increased complexity.

4.3.3 Performance over a Multi-span Nonlinear Fiber Link

After measuring the performance over an AWGN channel, we now evaluate the performance of the three schemes over a multi-span fiber link through numerical simulations, as shown in Fig. 4.10. The transmitted signal consists of 11 WDM channels, each modulated with one of the three schemes at 45 GBaud. The net rate per channel is 198 Gbits/s. The channel spacing is 50 GHz and the central channel is located at 1550 nm. All channels are pulse-shaped with a root-raised-cosine filter (RRC) with a roll-off factor of 0.1.

The tested fibers are standard single mode fibers (SSMF) with a chromatic dispersion coefficient $D = 17$ ps/nm/km, a nonlinearity coefficient $\gamma = 1.32$ $(\mathrm{W}\cdot\mathrm{km})^{-1}$, and a fiber attenuation parameter $\alpha = 0.2$ dB/km at 1550 nm. Polarization mode dispersion (PMD) is added to the simulated link with a PMD coefficient equal to 0.04 ps/$\sqrt{\text{km}}$. The link configuration is a homogeneous multi-span link of 20×80 km. We add an extra noise loading of -13 dBm at the receiver side to get a faster evaluation of low BER values in the dip of the BER curve. Each span is followed by an erbium-doped fiber amplifier (EDFA) with a noise figure of 5 dB.

At the receiver side, the central channel is optically filtered and coherently detected. The digital signal processing (DSP) consists of matched filtering, chromatic dispersion compensation, genie-aided linear MIMO channel equalization to remove PMD and polarization crosstalk, and finally, a genie-aided XPM-induced phase noise filter with an averaging window of 64 symbols as detailed in Sec. 1.6. We introduce no laser phase noise and no frequency offset between the signal carrier and the local oscillator. We choose a genie-aided DSP approach (perfect knowledge of the linear MIMO channel and optimal XPM-phase noise correction) to measure the intrinsic performance of the 4D-shaped constellations independently from the potential penalties that a sub-optimal DSP chain might add. The equalized symbols are then passed to the belief-propagation decoder for soft-decision FEC decoding. We do not perform iterations between the demapper and the decoder in order to avoid an increase in the complexity of the receiver.

In Fig. 4.11(a), we plot the BER curves, and in Fig. 4.11(b), the electric SNR curves as a function of the launched optical power. Electric SNR is the SNR measured from the noisy constellation points at the input of the decision circuit. As the electric SNR curves are almost the same with slightly better performance for the PCS 4D 88SP-16QAM, the BER enhancement of

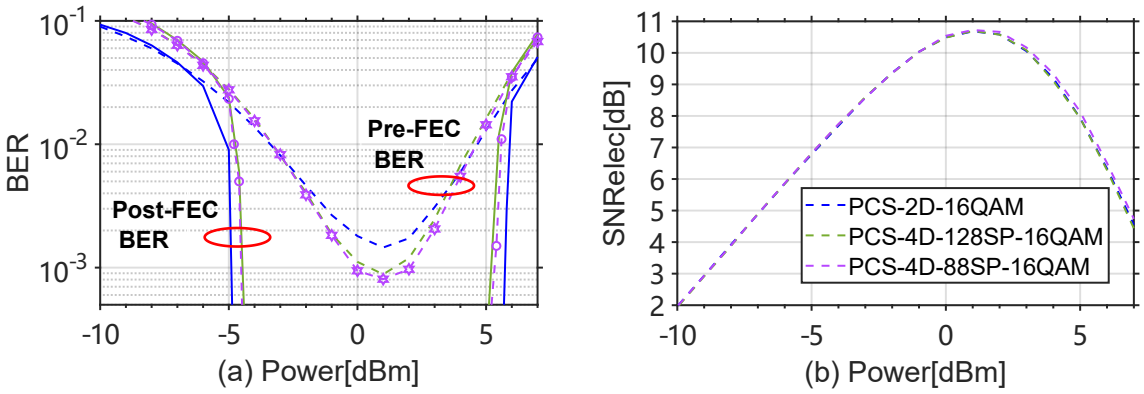


Figure 4.11: Performance over a multi-span link (same PCS schemes as in Fig. 4.9): (a) BER vs launched power/channel. (b) Electrical SNR vs launched power/channel.

the PCS 4D schemes, for BER values lower than 10^{-2}, stems from the increased MSED that brings performance gains in the linear regime.

4.3.4 Discussion and Conclusion

Through our numerical simulations, we have demonstrated that both linear and nonlinear gains can be achieved via probabilistic shaping of set-partitioned multi-dimensional modulations. In our proposed PCS-4D 88SP-16QAM scheme, we opted for probabilistic shaping across three 4D energy levels, in contrast to the five 4D energy levels utilized in the initial 128SP-16QAM constellation. Although using fewer energy levels in the constellation is less optimal for a linear AWGN channel, we observed that eliminating the two highest energy levels marginally improved performance in nonlinear conditions. It is important to note, however, that these gains were observed at BERs lower than 10^{-2} for a receiver employing sequential demapping and decoding. To extend these gains to higher BERs, the implementation of a joint demapper-decoder would be necessary, albeit at a higher complexity cost. While the gains noted with the two PCS-4D strategies were not substantial, the described design rule lays the groundwork for future research. This could lead to the development of multi-dimensional PCS schemes yielding more significant gains. Further enhancements are possible by refining the proposed construction, such as optimizing the selection process from the set-partitioned modulation and improving the bit mapping strategy.

4.4 4D-energy Limitation in ESS for NLI Reduction

As ESS with short block length was proven to have better performance than CCDM in both linear and nonlinear regimes, our next research direction mainly focuses on ESS with higher tolerance to nonlinear Kerr effects. We propose a 4D energy limit over ESS shaping to achieve either an improved tolerance to Kerr effects or a lower rate loss than state-of-the-art kurtosis-ESS

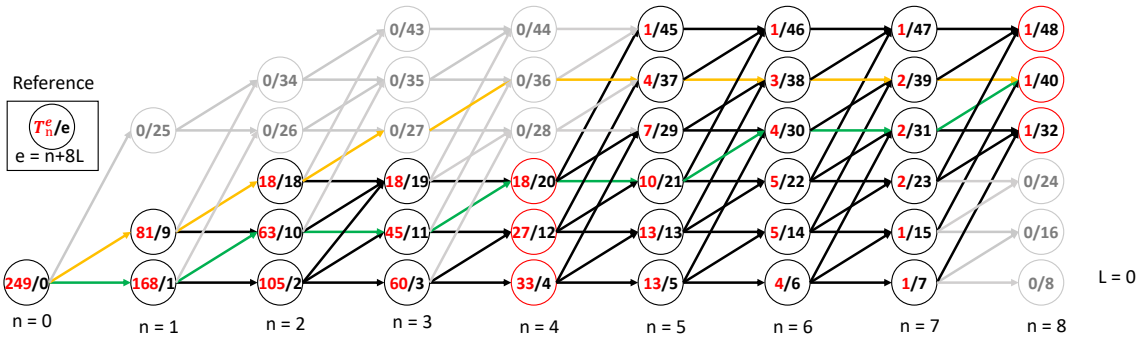


Figure 4.12: 4D-Band-limited trellis example using four amplitudes $\{1,3,5,7\}$, $E_{\max} = 48$, and block length $N = 8$.

(K-ESS) [216] and band-limited ESS (B-ESS) [217], and this is achieved by tuning the design parameters of the new scheme. Simulation results show that the proposed shaping outperforms conventional ESS in two different scenarios when applied over 64-QAM. We measure a gain of 0.19 bit per 4D symbol in the achievable information rate (AIR) in the same scenario scenario as for K-ESS and B-ESS, namely a 205-km single-span standard single-mode-fiber (SSMF) link and a WDM transmission of five channels with a net rate of 400 Gbit/s per wavelength at 50 GBaud and a block length of $N = 108$. We also tested a WDM transmission over a 5-span 400-km link with five 110 GBaud channels, for which smaller gains are observed even with larger block lengths.

4.4.1 Principle of 4D-BL-ESS

While both K-ESS [216] and B-ESS [217] (that we call 1D-BL-ESS in the following) use 4D mapping, they introduce energy constraints over a single dimension to limit the variations of each 1D amplitude. However, the strength of nonlinear effects depends on the variations of energy levels of the 4D symbols in consecutive time slots. Hence, by constraining energy variations of 4D symbols, we can improve the performance and achieve a better trade-off between linear and nonlinear gains. The following toy example illustrates this advantage. Consider an ESS with a block length $N = 8$ and two amplitude sequences $(3,3,3,3,1,1,1,1)$, highlighted in yellow in Fig. 4.12, and $(1,3,1,3,1,3,1,3)$, highlighted in green. These sequences have the same 1D average energy, variance, and kurtosis, and generate the two consecutive 4D-symbols: $[(\pm 3 \pm 3i, \pm 3 \pm 3i)^T, (\pm 1 \pm 1i, \pm 1 \pm 1i)^T]$ and $[(\pm 1 \pm 3i, \pm 1 \pm 3i)^T, (\pm 1 \pm 3i, \pm 1 \pm 3i)^T]$ respectively where the first symbol is attributed to the polarization X and the second symbol to the polarization Y in each pair, and $(\cdot)^T$ is the transpose of $(\cdot)$. Symbols from the first sequence show a large temporal 4D energy fluctuation from 36 to 4 while symbols from the second one have a constant 4D energy equal to 20.

In Fig. 4.12, we illustrate the new 4D-band-limited ESS (4D-BL-ESS) scheme through a trellis using 64-QAM that has 4 amplitudes in 1D: $\{1,3,5,7\}$, a block length $N = 8$, and a maximum

accumulated energy limit $E_{\max} = 48$. n and L are respectively the column and row indices of the nodes in the trellis. Each path in the trellis corresponds to a possible amplitude sequence. In each node, the number in black e is the intermediate accumulated energy up to that state, and the number in red T_n^e is the number of paths that lead from this state to a final state at $n = N$. The accumulated 1D energy in the 4-th and 8-th positions (red nodes) corresponds to the accumulated 4D energy in the 1-st and 2-nd time slots respectively. Sequences in grey are deleted based on a 4D-energy constraint instead of a 1D constraint as initially done in [217]. For this example, the constraint is: $4+28(i-1) \leq E_{4D,i} \leq 28+28(i-1)$ where $E_{4D,i}$ is the accumulated 4D energy in the i-th time slot. The construction used in the toy example can be generalized by tuning the constraints on the accumulated energy:

$$A + iK_1 \leq E_{4D,i} \leq \min(B + iK_2, E_{\max}) \tag{4.2}$$

where the parameters A, B, K_1, K_2 define the upper and lower limits of 4D energies in the trellis. By varying the values of these four parameters, we can control the width of the band as well as the start and end points of the band. At a fixed rate, the higher the maximum energy, the more sequences are possible, and the narrower the band width is to limit more sequences.

For ease of comparison between different schemes, we evaluate several metrics related to the linear and nonlinear performance using the same signaling parameters as in [216, 217], i.e. a shaping rate of 1.5 bits/amplitude, a block length $N = 108$, an LDPC FEC code with a rate of $r_c = 5/6$ and achieve a net bit rate of 8 bits/4D symbol which is calculated as in Sec. 2.5.3. For conventional ESS, the net bit rate for each amplitude is determined only by the chosen value of $E_{\max}$. However, for all three band-limited ESS and the K-ESS, we start with a larger energy limit $E_{\max}$, hence with a larger bit rate, and we then reduce the bit rate through constraints such as band limiting the energy or limiting the kurtosis until we reach the same net rate as the one of conventional ESS.

In [216, 217], the average energy of 1D amplitudes $\overline{E}_{1D}$ is used to assess the linear penalty, also the variance of 1D amplitudes σ^2_{1D} and the kurtosis of 2D amplitudes $\mu_{4,\text{2D}}$ are the used metrics to assess the nonlinear gain. Hence, in Table 4.1, we compute these metrics for five schemes: the conventional ESS [229], the kurtosis-ESS (K-ESS) [216], the 1D-BL-ESS [217] and two proposed 4D-BL schemes. We also add two 4D metrics: the energy dispersion index (EDI) defined in [107], and the windowed kurtosis defined in [108]. These two 4D metrics were shown to be better suited for assessing nonlinear gains of non-iid signaling schemes such as the short block-length PAS modulations [107, 108, 230]. The length of the averaging windows W_1 for the EDI and W_2 for the windowed kurtosis are computed as Eq. (2) in [107] and Eq. (7) in [108]. Their values in Table 4.1 are computed for the first transmission scenario over a single-span 205-km SSMF link simulated in the next section.

Table 4.1: Metrics for five different ESS schemes with $N = 108$, rate = 1.5 bits/1D amplitude

	ESS	1D-BL-ESS	4D-BL-ESS (Linear)	4D-BL-ESS (Nonlinear)	K-ESS
$E_{\max}$	860	996	948	996	1156
$\overline{E}_{1D}$	7.85	9.14	8.63	9.09	8.27
σ^2_{1D}	1.75	1.49	1.52	1.45	1.16
$\mu_{4,2D}$	1.87	1.71	1.69	1.64	1.57
EDI $(W_1 = 42)$	3.13	0.65	0.72	0.48	3.34
$\overline{\mu}_{4,E_{4D}}(W_2 = 54)$	3.63	2.8	2.84	2.69	3.17
Rate loss [bit/1D]	0.03	0.13	0.097	0.13	0.07

Here, we present the design and study the performance of two 4D-BL schemes: the first one, optimized for the linear regime, is denoted 4D-BL-ESS (Linear), $[A,B,K_1,K_2] = [4,92,34,33]$, and has a lower accumulated energy threshold $E_{\max}$ than its 1D-BL counterpart, thus a lower rate loss (linear gain) and almost the same energy variations as 1D-BL-ESS; and the second scheme, optimized for the nonlinear regime, is denoted 4D-BL-ESS (Nonlinear), $[A,B,K_1,K_2] = [4,84,36.5,35.2]$, and has the same accumulated energy threshold $E_{\max}$ as 1D-BL-ESS (hence, the same rate loss too) but it has lower 4D energy variations. The rate loss for each scheme is computed and reported in the last row of Table 4.1. We can see that all band-limited ESS schemes show lower energy variations than the conventional ESS (first column), but have a higher rate loss due to the addition of energy constraints that reduce the number of allowed amplitude combinations. At the same net rate, K-ESS has the lowest 1D energy variation, but the highest 4D energy variation, which hits at a worse nonlinear performance than the other schemes. This will be validated through the simulations in the next section.

4.4.2 Single-span 205-km 50GBaud Transmission

We evaluate the performance of the two proposed 4D-BL-ESS with $N = 108$ under the same conditions as the ones used in [216, 217]. The transmitted signals are 5 dual-polarization WDM channels, with a baud rate of 50 GBaud per channel. The raw data rate per channel is 600 Gbit/s. For all ESS schemes, we use an LDPC code of length $n = 64800$ bits from the DVB-S2 standard with a rate of $r_c = 5/6$. The achieved net bit rate is 8 bits per 4D symbol. The channel spacing is 55 GHz. Root-raised-cosine pulse shaping is applied with a roll-off factor of 0.1. The signal propagation is simulated through the split-step Fourier method based on the nonlinear Schrödinger and Manakov equations. SSMF fiber segments are simulated with a chromatic dispersion coefficient $D = 17\text{ps/nm/km}$, nonlinearity coefficient $\gamma = 1.3(\text{W}\cdot\text{km})^{-1}$, and an attenuation $\alpha = 0.2$ dB/km at $\lambda = 1550$ nm. Polarization mode dispersion (PMD) is added with 0.04 $\text{ps}/\sqrt{\text{km}}$. The link is a single span of 205 km. No laser phase noise is added at this point.

After propagation, the central channel is filtered. The signal undergoes matched filtering, chromatic dispersion compensation, and genie-aided multiple-input-multiple-output (MIMO) channel

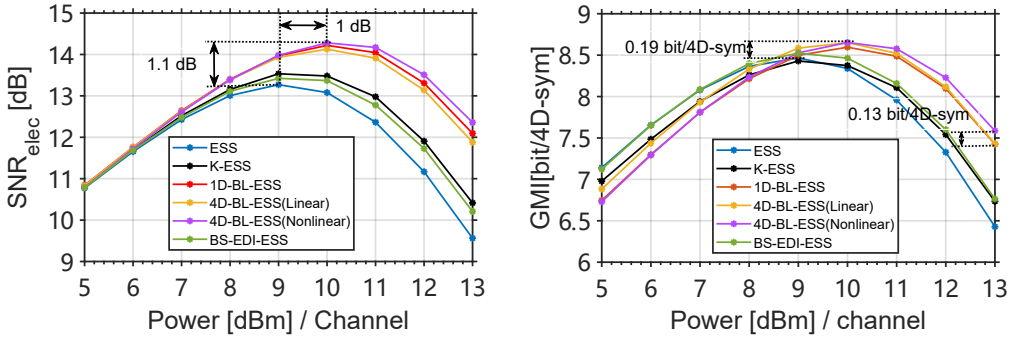


Figure 4.13: Left: Electrical SNR versus Power per channel in the nonlinear channel. Right: GMI versus Power per channel. Simulation of conventional ESS, K-ESS, 1D-BL-ESS, 4D-BL-ESS(Linear), and 4D-BL-ESS(Nonlinear) in 50 GBaud over 1×205 km SSMF link, 5 WDM channels, 55 GHz channel spacing, net rate per channel 400 Gbit/s.

equalization. Finally, we use a fully-data-aided phase filter with an averaging window of 64 symbols to compensate for the phase rotation induced by cross-phase modulation (XPM), as described in Sec. 1.6 and we compute the electric SNR and the GMI from the equalized signal. The electric signal-to-noise ratio $\mathrm{SNR}_{\mathrm{elec}}$ is the SNR computed from the constellation at the end of the DSP chain at the receiver side (i.e., at the input of the decision circuit) and the GMI is computed as Eq. (1.34). For each launch power, five different channel realizations are simulated.

In the left part of Fig. 4.13, we show the electric SNR versus the launch power per channel for the five ESS schemes. We see that the 1D-BL, 4D-BL (Linear), and 4D-BL (Nonlinear) ESS show respectively a 1 dB, 0.9 dB, and 1.1 dB increase in SNR at a power of 10 dBm compared to conventional ESS, and a 1 dB enhancement in optimal launch power. In the right part of Fig. 4.13, we show the results in terms of GMI computed as Eq. (1.34) per 4D symbol versus the optical power. The 4D-BL (linear) and (Nonlinear) show the best GMI. At a launch power of 10 dBm, the bit rate is increased by 0.07 bit/4D compared to the 1D-BL-ESS, and by 0.19 bit/4D compared to conventional ESS. 4D-BL-ESS (Nonlinear) also shows 0.13 bit/4D gain compared to the 1D-BL-ESS in highly nonlinear regime. While K-ESS shows the lowest 1D energy variation as can be seen from Table 4.1, its 4D energy variation is higher than those of the BL-ESS schemes, which is compliant with the measured SNR and GMI performance in Fig. 4.13.

4.4.3 5×80-km Multi-span 110 GBaud Transmission

Next, to assess the scaling of the linear and nonlinear gains, we compare four schemes: conventional, 1D-BL and 4D-BL ESS (Linear and Nonlinear) in a new transmission scenario with a higher baud rate and over a multi-span link configuration for different shaping block lengths N. We consider a 110 GBaud transmission over 5×80 km SSMF link, 5 WDM channels with a 112.5 GHz channel spacing. We also consider a laser linewidth of 100 kHz at each transmitter

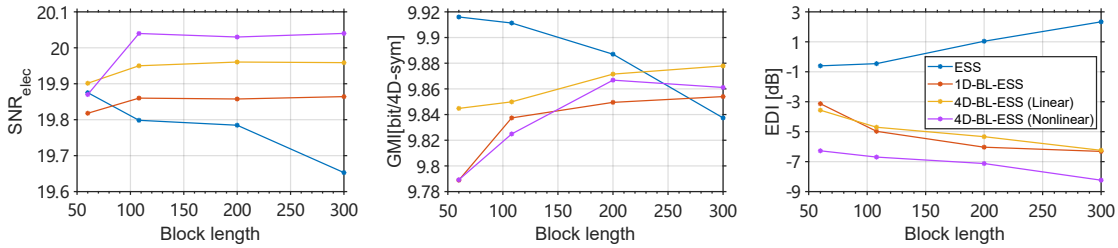


Figure 4.14: Simulation of conventional ESS, 1D-BL-ESS, 4D-BL-ESS(Linear), and 4D-BL-ESS(Nonlinear) for a 110 GBaud transmission over 5×80 km SSMF link, 5 WDM channels, 112.5 GHz channel spacing, net rate per channel: 838 Gbit/s. Electrical SNR (left) and GMI (middle) vs block length at optimal power (3 dBm). EDI (right) with $W_1 = 200$ versus block length.

and receiver side. The net rate per channel is 838 Gbit/s (880 Gbit/s with 5% pilot overhead). We compensate for the added phase noise using a realistic pilot-aided CPE [231] with a 5% overhead instead of a genie-aided method as in the previous section. The other DSP blocks are the same as the ones used in the previous section. We study the impact of the block length N on the performance in terms of electric SNR and AIR, both measured at the optimal launch power of 3 dBm.

In Fig. 4.14, we show the variation of the electric SNR, the AIR, and the EDI with respect to the shaping block length $N = \{60, 108, 200, 300\}$. For each block length, five different channel realizations are simulated. First, we can see, by comparing the leftmost and the rightmost figures, that the measured SNR is inversely correlated to EDI changes. However, we also notice that EDI only gives a rough prediction of the performance as it fails to capture small differences such as the gap between the performance of 1D-BL-ESS and 4D-BL-ESS (Linear). Second, we can see that, at a short block length $N = 60$, all BL-ESS schemes fail to achieve a significant gain in SNR as the nonlinear gains from energy limitation are still small compared to the conventional ESS. Third, when the rate loss decreases as the block length increases, and consequently the linear penalty decreases, the energy fluctuations of conventional ESS reduce the SNR while the SNR of the BL-ESS schemes remains stable up to $N = 300$. We did not increase the block length beyond 300 to maintain a low complexity for the distribution matching.

Finally, if we assess the GMI gains through the middle figure, we see that a net performance gain of BL-ESS schemes with respect to conventional ESS are observed beyond a block length of 200. However, all the GMI values achieved by the BL-ESS schemes remain lower than the GMI of conventional ESS for $N \leq 200$ due to their higher rate loss. Our explanation of the limited gains of the BL-ESS for longer distances and higher baud rates is the following: despite the optimization of the signal characteristics at the transmitter side for low NLI distortions through the addition of energy constraints, the chromatic dispersion gradually destroys these characteristics during the transmission. Hence, the band-limited ESS can only obtain very limited nonlinear gains over long-

distance and high baud-rate transmission (gains achieved over the first few spans), and the gain in GMI is also limited due to the higher rate loss.

4.4.4 Conclusion

We proposed a new design of the ESS scheme that constrains energy variations in higher dimensions (four dimensions in this work) and that can be tuned to boost either linear or nonlinear gains. The new scheme achieves higher gains than state-of-the-art schemes and is particularly interesting for short-distance links as the trade-off between the achieved linear and nonlinear gains becomes harder to meet for longer distances because of the accumulation of chromatic dispersion that increases the energy variations of the optical signal. Besides, our simulation results show that windowed 4D metrics of energy variations predict the nonlinear performance of ESS schemes better than 1D and 2D metrics; however, their prediction is not yet accurate enough and the search for better 4D metrics is still an interesting problem to solve.

4.5 Summary

In this chapter, we introduced two new PCS designs. The first, PCS-4D set-partitioned QAM, leverages PCS in multi-dimensional modulations, illustrating a slight performance enhancement in both linear and nonlinear regimes. This approach was exemplified using PCS-4D 88SP-16QAM modulation, which shows performance improvements through the utilization of fewer high energy levels. However, these benefits become evident primarily at lower error rates and necessitate more sophisticated DSP techniques. The second design, 4D-energy band-limited ESS, operates over four dimensions and aims at managing energy fluctuations to boost linear or nonlinear performance gains. This novel scheme surpasses existing ESS methods, especially in short-distance links.

The restriction of performance gains to short distances is attributed to the increasing complexity of balancing linear and nonlinear gains over longer distances, a challenge compounded by the accumulation of chromatic dispersion that intensifies energy variations in high-baud rate optical signals. Our analysis also revealed that 4D metrics are more effective in predicting the nonlinear performance of ESS schemes compared to 1D and 2D metrics. However, the current 4D metrics still lack sufficient precision, highlighting the need for more accurate 4D metric development as a key direction for future research and the achievement of nonlinear gains for multi-span transmission. We will delve into these two topics in the next and last chapter of the thesis.

Chapter 5

Selection of Probabilistic Constellation Shaped Sequences

5.1 Introduction

This chapter is dedicated to the exploration of metrics that assess Nonlinear Interference (NLI) or the potential of a sequence in generating NLI, which is essential for our task of designing nonlinear tolerant signalling schemes. We trace the development of sequence selection strategies based on a trial-and-error approach consisting in the evaluation of a metric for different sequences and the selection of the sequence that generates the lower amount of NLI. We introduce in this chapter our proposition of dispersion-aware metrics and use these for sequence selection showcasing an innovative approach that have a reduced complexity compared to other metrics used in recent prominent research works.

We have seen in the previous chapter notable methods for implementing PCS modulations including enumerative sphere shaping (ESS) [229] and constant composition distribution matching (CCDM) [22]. CCDM [22] has emerged as a competitive technique for implementing PCS with rectangular or square M-QAM constellations by delivering near-optimal linear shaping gains when large amplitude blocks are considered [211]. However, in the context of mitigating Kerr-induced nonlinear interference (NLI) in optical fiber transmission, long block length DM may exhibit a higher susceptibility to performance degradation, as highlighted in previous analyses [225, 232, 233]. These works showed that short block length DM can effectively reduce NLI in optical fiber transmission. Yet, it is crucial to consider that, as the block length decreases, the rate loss increases. Hence, finding a high-performance nonlinear tolerant shaping scheme involves striking a trade-off between nonlinear gain and rate loss [149]. Besides, simulation results in [234, 235] showed that carrier phase recovery (CPR) in the DSP chain at the receiver side can exhibit comparable effectiveness to finite-length PCS in mitigating NLI.

We have also reviewed in the last chapter the introduction of constraints in the trellis grid tree of ESS (kurtosis-limited ESS (K-ESS) with a maximum kurtosis limit on the output sequence, and band-limited ESS (B-ESS) which uses a specific subset of the trellis tree to limit 1D amplitude variance and kurtosis). While both K-ESS and B-ESS implement sequences in four-dimensional (4D) space while applying energy limitations over each separate dimension, we proposed a band-limited version of ESS with energy limitations in four dimensions (the two modulated polarization tributaries), achieving further throughput enhancement. However, for these three schemes, the rate loss was still significant.

Other design approaches of nonlinear tolerant signaling schemes were recently suggested. They consist in manipulating the input signal to the distribution matcher (DM) to generate multiple candidate sequences. Then, specific metrics are used to select “good" sequences, i.e., sequences generating low NLI. For example, the energy dispersion index (EDI), introduced in [107], is a metric that quantifies energy changes of the signal over a given temporal window. EDI showed a negative correlation with the effective Signal-to-Noise Ratio (SNR) of the CCDM-based transmission; however, **this negative correlation is observed in a specific single-polarization transmission**. As for ESS-based schemes, our study in [236] highlighted that the EDI, when calculated over 4D-energy levels, demonstrates a notable negative correlation for conventional ESS, B-ESS [217], and 4D-BL-ESS over 110 Gbaud, 400 km, 5 WDM transmission. However, this correlation appears to be inaccurate in the case of K-ESS [216]. This motivates us to look for more reliable metrics that can better predict the effective SNR.

In a different work [237], EDI was utilized for sequence selection in a scheme called List-Encoding CCDM (L-CCDM), achieving nonlinear gains with respect to conventional CCDM. Yet, for long-haul transmissions, CCDM is not the best distribution matcher, as similar nonlinear gains can be achieved through conventional ESS with shorter block lengths [149]. More recently, yet another approach, presented in [38], used EDI and a perturbation-model-based low-pass-filtered symbol-amplitude sequence (LSAS) metric [238] for sequence selection. These methods have substantially improved the nonlinear tolerance of ESS or CCDM. However, it is essential to highlight that **most significant gains in ESS schemes have been observed over short-distance single-span transmissions**.

The concept of **sequence selection (SS)** was first introduced in [36] and has been further studied in [239]. This innovative signaling approach provided valuable insights into the computation of a lower bound of the optical fiber channel capacity. In particular, the analysis clearly showed that there is considerable room for improvement in optical fiber transmission, especially for long-haul links. A recent study [3] from the same authors emphasized the importance of enhancing performance through sequence selection using sign-dependent metrics in long-distance transmission scenarios. The authors demonstrated that sign-independent metrics such as EDI, windowed

kurtosis, or LSAS do not yield significant throughput improvements over long-haul transmissions. They also showed that a sequence selection based on the computation of NLI through a numerical simulation (Split-Step Fourier Method or SSFM) of a noiseless single-channel propagation of sequences achieved throughput enhancements over long-distance transmissions even when an optimized CPR is applied. However, this selection method is not feasible in practice due to the high complexity of the SSFM algorithm.

we introduce a novel sign-dependent metric, the EDI of dispersed sequences, named D-EDI, which accounts for the influence of chromatic dispersion throughout transmission. Additionally, we show that D-EDI demonstrates a negative correlation with the effective SNR for high-rate multi-span scenarios. By negative correlation, we mean that the two metrics vary in opposite directions. Next, we utilize EDI and D-EDI as metrics for sequence selection, with ESS serving as the DM. The approach that employs sign-independent EDI for sequence selection is termed E-SS, and the one using sign-dependent D-EDI is designated as D-SS[1]. We thoroughly investigate and optimize the performance of these schemes in both single-span and multi-span transmission scenarios with optimized CPR. Our D-SS scheme demonstrates superior performance compared to ESS without sequence selection across various block lengths and complexity levels, in both single-channel and WDM transmission scenarios. Remarkably, it performs on par with the ideal SSFM-based sequence selection method, however with a reduced complexity.

The chapter is structured as follows: in Section 5.2, we introduce the evolution of different sequence selection methods, using sign-independent and sign-dependent metrics. in Section 5.3.2, we describe the simulated transmission systems, then, in Section 5.3.3, we define the new metric D-EDI. In Section 5.3.4, we outline the architecture of the E-SS and D-SS transmitter schemes. Section 5.3.5 focuses on the performance of a single-span transmission system and gives insights into the optimization of the proposed methods in this context. In Section 5.3.6, we study the performance over multi-span single-channel and WDM transmission systems with CPR at the receiver side. Finally, Section 5.4 summarizes key findings and conclusions drawn from our investigations.

5.2 From Sign-independent to Sign-dependent Sequence Selection

5.2.1 Sequence Selection

The precise capacity of fiber channels in the nonlinear regime remains an elusive target, with tight bounds not yet known, as indicated in references [240–242]. However, it is widely accepted that this capacity is bounded. The upper limit is constrained by the linear capacity of the channel [243], representing the maximum capacity achievable under nonlinear and noisy channels. On the other

[1]For a PCS system employing sequence selection based on either EDI or D-EDI and comprising n cascaded DMs, where each DM includes ν flipping bits, we name the scheme E-SS$_n^\nu$ for the EDI-based approach, and D-SS$_n^\nu$ for the D-EDI-based approach. The schemes will be detailed in section 5.3.4.

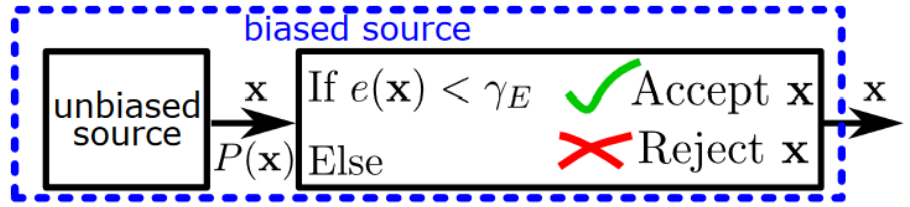


Figure 5.1: The biased source, obtained with a selection procedure from an unbiased source. Source: [36].

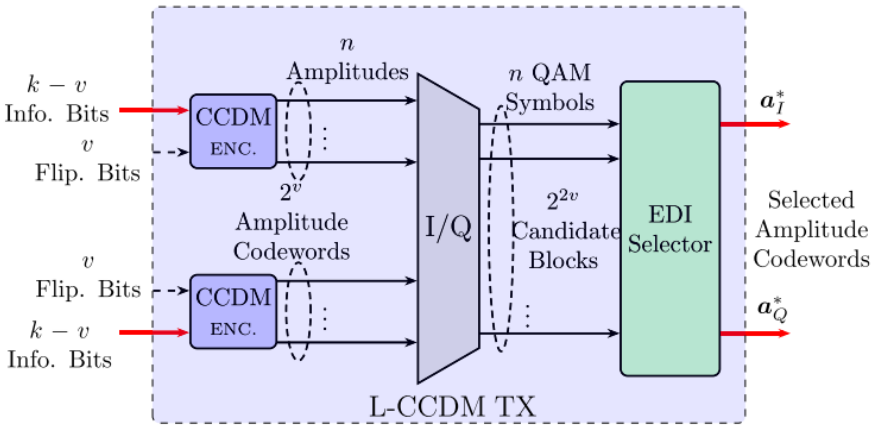
hand, the lower limit is defined by numerous bounds provided in [90, 244, 245], which offer a range of theoretical and empirical limitations on the capacity.

These bounds are typically established by computing an Achievable Information Rate (AIR) while considering a fixed simple input distribution such as independent and identically distributed (i.i.d.) samples with a Gaussian distribution [246–249]. Alternatively, unbounded capacity with an increasing SNR can be approached using continuous constellations, such as discretizing a 2D Gaussian constellation in magnitude, thus creating concentric rings constellation [90]. The optimization of decoding metrics within these distributions contributes to the estimation of AIR and, consequently, to the understanding of the capacity bounds.

Understanding these bounds is crucial for designing and optimizing optical fiber communication systems, as it provides insights into the limits of transmission rates and efficiencies under nonlinear conditions. While the exact capacity might not be known, these upper and lower limits serve as critical guidelines for the theoretical and practical exploration of fiber optic communications in the nonlinear regime.

The basic principles of sequence selection are shown in Fig. 5.1. Sequence selection uses a certain metric to select 'good' sequences for transmission and reject 'bad' sequences. A random sequence of N symbols $\mathbf{X}$ is initially drawn from an unbiased source with probability distribution $P(\mathbf{X})$. Each sequence is then evaluated against a chosen measure. If it meets the criteria, it is accepted as the actual transmission source; otherwise, it is rejected and discarded. This method creates biased sources by selecting unbiased sources and then generates symbols for the channel inputs. The unbiased source can be chosen arbitrarily. For a signal source with N symbols, it can be a continuous constellation (such as Gaussian constellation), or a discrete constellation (such as QAM). The symbols can be independent or correlated. In our research work, the symbols are the output symbols of a PCS transmitter obtained by a finite-block-length distribution matcher.

In the broader context of sequence selection strategies, methods such as K-ESS [107], B-ESS [35], and 4D-BL-ESS [236] detailed in Chap. 4 can be categorized as types of crude sequence selection. These methods primarily involve changing the set of sequences generated by DM based on certain criteria. For instance, sequences with higher kurtosis are discarded in K-ESS. At the same time, B-ESS and 4D-BL-ESS involve trimming the trellis tree directly to force the generation



(a) L-CCDM in the PAS transmitter.

Figure 5.2: List CCDM in the probabilistic amplitude shaping transmitter. Source: [37].

of sequences with smaller energy fluctuations. K-ESS has a lower additional rate loss because of the lower number of rejected sequences. However, it is important to note that solely relying on one-dimensional amplitude kurtosis may not be sufficient to select a 'good' sequence, leading to limited improvements in nonlinear performance. In contrast, B-ESS and 4D-BL-ESS focus more on managing energy fluctuations but may not provide a selection process as fine as K-ESS.

Moving towards more refined sequence selection strategies, one approach involves generating multiple candidate output sequences by varying the input source bit sequence of the DM, rather than altering the DM itself. These candidates can then be evaluated based on their Energy Dispersion Index (EDI) using List-CCDM [237] or through the Lowpass-Filtered Symbol-Amplitude Sequence (LSAS) metric using LSAS-ESS [38]. The optimal sequence for transmission is selected based on comparative assessments of these sign-independent metrics. They are called sign-independent because these metrics focus on the symbol energies and discard their signs. On the other hand, calculating the intensity of nonlinear noise for candidate sequences through split-step Fourier method (SSFM) simulations [3] provides an upper bound on the performance that current sequence generation methods can achieve. This calculation offers a realistic gauge of the potential improvements and limitations inherent in these sequence selection strategies by sign-dependent metric, guiding further development and optimization in the field of optical communications.

5.2.2 Sign-independent Sequence Selection Metrics

List-CCDM: EDI-based Sequence Selection

List-CCDM [37, 250] is an evolution of conventional CCDM consisting in the addition of v flip bits to change the DM output as shown in Fig. 5.2, followed by the selection of a subset of sequences

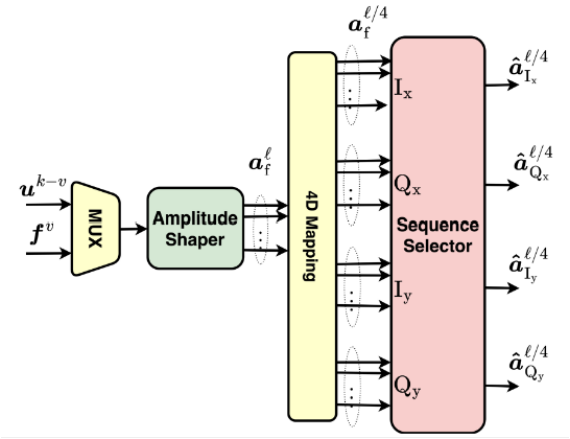


Figure 5.3: Block diagram for 4D-mapping followed by sequence selection. Source: [38].

based on a metric optimization, for instance, the minimization of the energy dispersion index (EDI) which is computed as follows. Let G_i^W be a moving window statistic that measures the windowed energy fluctuation in a sequence of symbols $\mathbf{X}$ defined in [37] as:

$$G_i^W = \sum_{j=i-W/2}^{i+W/2} |X_j|^2 \tag{5.1}$$

where W is an even positive integer and X_j are symbols from the sequence $\mathbf{X}$. Based on this windowed energy process, EDI is defined as the ratio of the variance of the windowed energy to the mean of the windowed energy, both averaged over $i = \{1, \cdots, N\}$ where N is the block length.

$$\text{EDI} = \frac{\frac{1}{N}\sum_{i=1}^{N} \overline{\text{var}}(G_i^W)}{\frac{1}{N}\sum_{i=1}^{N} \overline{\text{E}}(G_i^W)} \tag{5.2}$$

In this section, we only briefly introduce EDI. In order to understand D-EDI more coherently, we will explain the calculation process of EDI in detail in Sec. 5.3.3.

LSAS-based Sequence Selection

Another method for selecting shaped sequences that minimizes NLI is the LSAS method, which allows the calculation of the nonlinear tolerance of a finite-length PAS for a given system configuration (such as link length and baud rate) by using a frequency-domain representation of an effective linear channel [38]. LSAS stands for Lowpass-Filtered Symbol-Amplitude Sequence. The proposed metric, denoted as λ_{LSAS}, is expressed as:

$$\lambda_{LSAS} = \sum_{n=0}^{\frac{l}{d}-1} \sum_{p \in P} \left| s_p^{(c)}(n) \sum_{p' \in P} \sum_{c' \in \widetilde{C} \subseteq C} \Delta D_{p,p'}^{(c,c')}(n) \right|^2 . \tag{5.3}$$

The dimensionality of the mapping is denoted as d, and $\widetilde{C}$ represents a chosen subset of all possible sequence candidates C. The length of each block in this context is represented by l. A crucial element of the λ_{LSAS} metric is the distortion variation term, $\Delta D_{p,p'}^{(c,c')}$, as defined in Eq. (7) of [38]. Notably, for this metric, only the central l/d convolution outputs are considered. This specific focus ensures that the λ_{LSAS} metric retains its independence across consecutive shaping blocks. The λ_{LSAS} metric is particularly designed to capture the nuances of nonlinearity as expressed through symbol amplitudes and to account for the filtering effects imposed by the fiber channel. By considering the symbol magnitude and how sequences are filtered through the channel, λ_{LSAS} provides a comprehensive measure of sequence quality with respect to nonlinear distortion.

The block of LSAS sequence selection operates similarly to L-CCDM, but it adopts ESS as its DM method, resulting in superior performance compared to the EDI metric. However, it is important to note that while both EDI and LSAS sequence selection methods offer gains in transmission performance, these improvements are mostly evident in single-span transmissions. The benefits tend to disappear over long-haul transmissions, where the accumulation of dispersion and other impairments over extended distances limits the efficacy of these sequence selection strategies. Consequently, while EDI and LSAS provide notable advantages in certain transmission scenarios, their impact over long distances is relatively modest.

Finally, another sign-independent metric, the nonlinear phase noise (NPN) metric was proposed in [251], which is based on the frequency-resolved logarithmic perturbation model. Compared with LSAS, it can further explain the interaction between CPE and PAS.

5.2.3 Sign-dependent Sequence Selection Metric: The SSFM

To effectively gauge the potential benefits of nonlinear shaping and quantify the achievable improvements, a sophisticated approach involves employing the average NLI metric as detailed in [36]. This metric necessitates transmitting the candidate sequences through a single-channel noiseless simulation to estimate the NLI it incurs. Crucially, this estimation process involves averaging the impacts of adjacent sequences, thereby accounting for the inter-block NLI effects. By leveraging this average NLI metric, one can assess not only the immediate nonlinear distortions impacting a given sequence but also how the sequence might influence or be influenced by its temporal neighbor sequences in a transmission stream.

The performance of a sign-dependent metric and a sign-independent metric are compared in Fig. 5.4 by three different scenarios.

- Ideal sequence selection with shaped symbols: In the first case, both amplitude and sign are shaped according to the average nonlinear interference (NLI) metric, that is, SSFM-based sequence selection. The method uses a comprehensive shaping strategy that considers the amplitude and phase characteristics of the signal to minimize NLI.

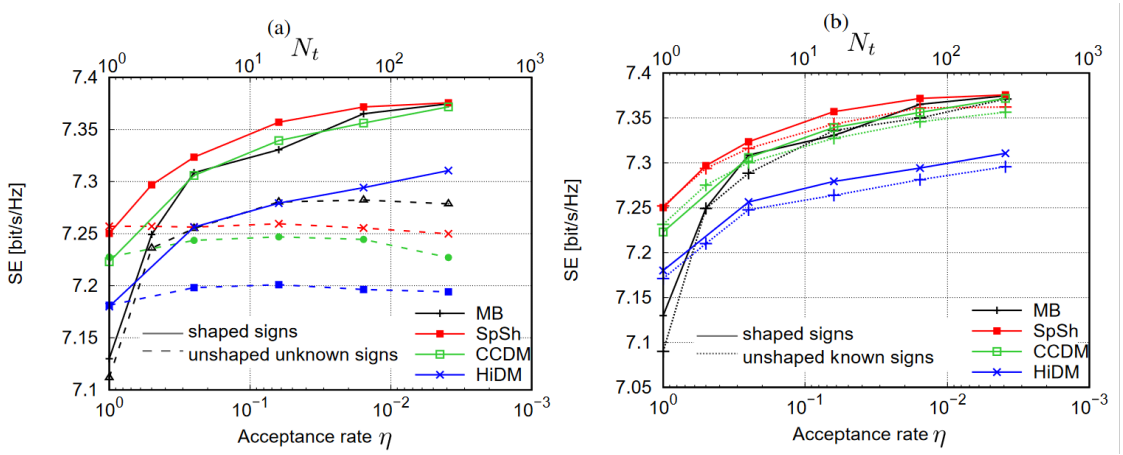


Figure 5.4: Optimal SE versus acceptance rate η with ideal sequence selection with (a) shaped signs (solid) and unshaped unknown signs (dashed), and (b) shaped signs (solid) and unshaped known signs (dotted). $5\lambda \times 46.5$ GBd with PCS-64QAM in 9.2bits/4D-symbol over 30×80 km link. MB: i.i.d. Maxwell Boltzmann distribution symbols. SpSh: Enumerative sphere shaping. CCDM: Constant composition DM. HiDM: Hierarchical DM Source: [3].

- Ideal Sequence Selection with Unshaped and Unknown Signs: The second case assumes that only the amplitude is shaped. Here, the metric has nothing to do with the sign of the symbol, in other words, it is only related to the energy change of the sequence and has nothing to do with the phase change of the sequence. This situation corresponds to sign-independent metrics such as EDI, NPN, LSAS, or kurtosis.

- Ideal sequence selection with unshaped and known signs: The third case is an intermediate case where the sign of the symbol is known and SSFM-based sequence selection is used. However, unlike the first case, the sign is fixed and cannot be changed by the sequence selection procedure. This approach recognizes sign information but does not allow its optimization as part of sequence selection.

By comparing the three situations we can draw the following conclusions:

- An effective sequence selection strategy necessitates the use of sign-dependent selection metrics. In contrast, sign-independent metrics like EDI, LSAS, NPN, or kurtosis may not yield significant benefits unless the unbiased source (prior to selection) is specifically optimized to reduce intensity fluctuations. Without such prior optimization, relying on sign-independent metrics alone could lead to minimal or no gain in system performance.

- For any given sign sequence, There are ‘good’ and ’bad’ amplitude sequences that can be identified and appropriately selected to improve performance.

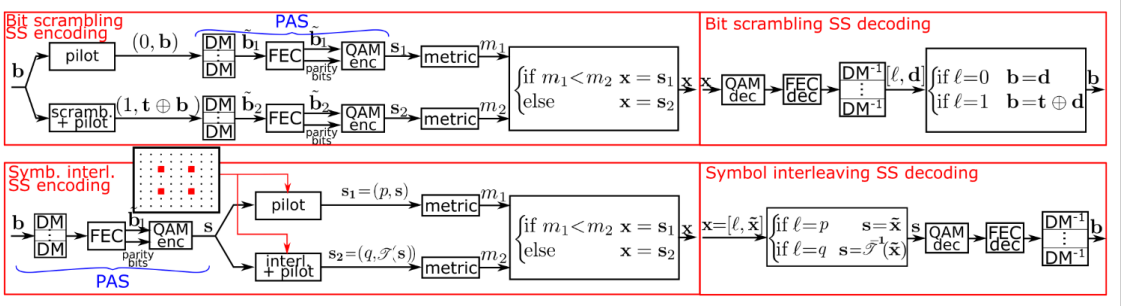


Figure 5.5: FEC-independent BS sequence selection (top) with single block (SB) (bottom left) and multi block (MB) (bottom right) implementations. Source: [3].

- Sequence selection with an unshaped but known sign nearly matches the performance of sequence selection with a shaped sign. This finding demonstrates the high complexity of simultaneous sign and amplitude selection. However, it is enough to maintain an i.i.d. sign distribution while selectively choosing the amplitude using a sign-dependent metric, which reduces a large part of the complexity and maintains performance improvements.

In their study, the authors explored two distinct methods for generating candidate sequences: Bit scrambling Sequence selection encoding (BS-SS) and Symbol Interleaving Sequence selection encoding (SI-SS) as shown in Fig. 5.5. BS-SS employs XOR operations to variate the input bit sequences. SI-SS method focuses on generating different candidate sequences by interleaving pilot symbols at various levels. These methods reorder the symbols in a sequence to create variations, aiming to identify sequences with lower nonlinear interference. Among these two, BS-SS has a superior performance compared to SI-SS. The effectiveness of BS-SS is likely attributed to its ability to produce a more diverse set of sequence variations through bit scrambling, providing a broader selection pool to find sequences with minimized nonlinear distortions.

5.3 Our proposal: Dispersion-Aware Sequence Selection

In this section, We introduce a novel sign-dependent metric: the energy dispersion index (EDI) of sequences that endured chromatic dispersion, denoted as D-EDI, which exhibits a more accurate opposite variations with the transmission performance compared to the standard EDI metric. Then, by applying D-EDI and EDI to the sequence selection (SS) process, we present two signaling approaches denoted as D-SS and E-SS respectively. These approaches are designed to minimize rate loss and enhance transmission performance in nonlinear optical fiber transmission systems, catering to both short-distance and long-haul scenarios. With enumerative sphere shaping (ESS) as distribution matcher (DM), our simulation results reveal significant performance gains over ESS without sequence selection (SS), with improvements up to 0.4 bits/4D-symbol. These improvements were observed over a 205-km single-span standard single mode fiber link in WDM

transmission, with five dual-polarization channels, each operating at a net rate of 400 Gbit/s. Furthermore, we demonstrate that D-SS surpasses ESS without SS by 0.03 bits/4D-symbol in achievable information rate over a 30×80 km link in a single-wavelength, with 8 discrete multi-band (DMB) transmission, and an 880 Gbit/s net rate. Notably, our proposed D-SS scheme achieves similar performance to a sequence selection based on a full split-step Fourier method (SSFM) simulation and it consistently delivers throughput enhancements across various block lengths and selected sequence lengths.

5.3.1 Our Mapping Strategy and Multi-block FEC-independent Sequence Selection

Mapping Strategy

Distribution matching can be applied across different dimensions, including 1D, 2D, 4D, or even more. Fig. 4.7 provides a visual representation of the mapping strategies employed by ESS schemes up to four dimensions. In the case of 1D mapping, 4D symbols are shaped using four independent amplitude sequences, essentially employing one ESS per dimension. In 2D mapping, each polarization, namely X and Y, utilizes a distinct ESS-shaped amplitude sequence. In contrast, 4D mapping employs a single-shaped sequence for a dual-polarization symbol sequence. For example, assuming the amplitude sequence is $[a_1, a_2, \cdots, a_8]$, the dual-polarization symbol sequence over two time slots employing 4D mapping will be:

$$\begin{bmatrix} S_{Pol.X,T_1} & S_{Pol.X,T_2} \\ S_{Pol.Y,T_1} & S_{Pol.Y,T_2} \end{bmatrix} = \begin{bmatrix} \pm a_1 \pm j a_2 & \pm a_5 \pm j a_6 \\ \pm a_3 \pm j a_4 & \pm a_7 \pm j a_8 \end{bmatrix} \tag{5.4}$$

This last approach offers the advantage of reducing the likelihood of having equal amplitudes at the same time slot, resulting in less frequent occurrences of high peak power values and leading to lower levels of non-linear distortions. This advantage has been demonstrated in various studies [216, 224, 225]. Hence, we use 4D mapping of the amplitudes at the output of the distribution matcher.

Multi-Block FEC-Independent Sequence Selection

The goal of the multi-block FEC-independent SS scheme is to reduce the complexity by moving the FEC from before to after the sequence selection. The implementation of multi-block FEC-independent sequence selection was proposed in [3]. As shown in Fig. 5.6, the parity bits of each FEC-independent D-E-ESS block (a block that will be presented in Fig. 5.7 of the next section), are assigned to a portion of the signs in the subsequent block. In the process of selecting a specific sequence, these signs are predetermined by the parity bits from the preceding block. These predetermined signs are required for the calculation of a sign-dependent metric. It is important to note

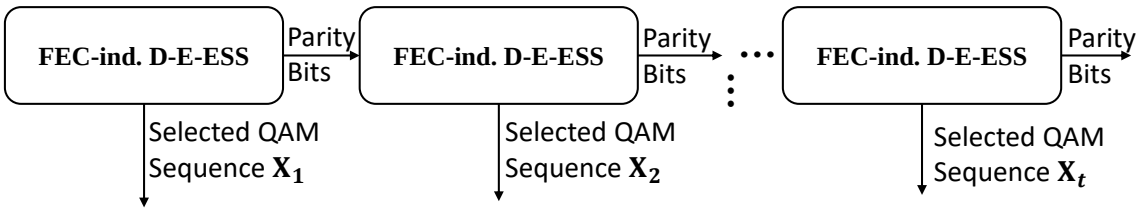


Figure 5.6: Multi-block FEC-independent D-E-ESS sequence selection implementations. The parity bits from a current FEC-independent D-E-ESS block are allocated to a subset of the signs in the subsequent block.

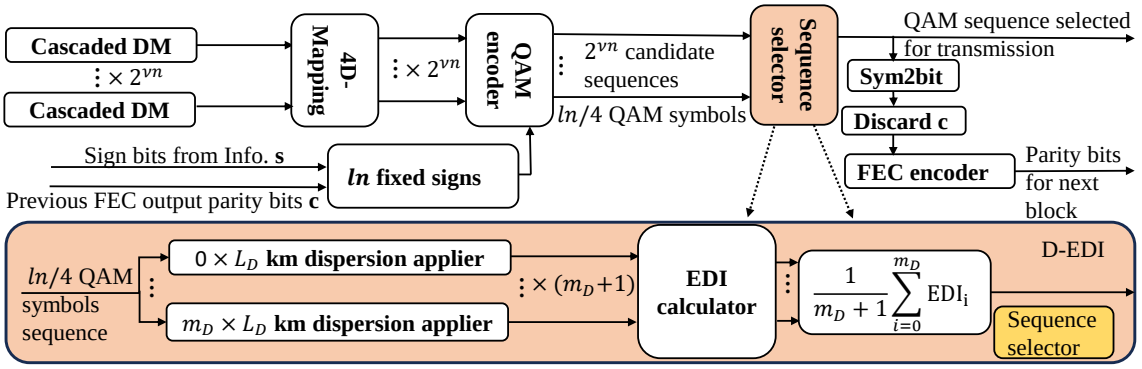


Figure 5.7: Block diagram of FEC independent D-E-ESS in the PCS transmitter, EDI of dispersed sequences (D-EDI) is determined by averaging the calculated EDI at multiple locations along an ideal dispersive fiber. The dispersion is applied at 1 sample per symbol. EDI is a special case of D-EDI where $m_D = 0$. The sign bits are fixed through a multi-block FEC-independent sequence selection process as shown in Fig. 5.6.

that, once the candidate sequences have been generated, their bits become fixed and unalterable. Consequently, if the FEC encoding is applied after the sequence selection (SS) process, it becomes necessary to shift the parity bits to the subsequent block.

5.3.2 System Description

We study the performance of E-SS and D-SS in two scenarios: first, a single-span transmission scenario similar to the one in [216, 217]; second, a multi-span long-haul transmission using digital-multi-band (DMB) format. Through the two studies, we observe the effects of nonlinear distortions with various signaling schemes. NLI distortions are dominant at high power levels, therefore, they primarily occur over the first kilometers of the span in a single-span transmission, and occur at the start of each span in a multi-span scenario with discrete optical amplification.

Our simulated single-span scenario consists of a transmission over 205 km of standard single-mode fiber (SSMF), to perform a consistent comparison with [216, 217] in which the kurtosis-limited (K-ESS) and band-limited (B-ESS) sphere shaping were introduced. We perform probabilistic shaping using PDM 64-QAM modulation with a block length $l = 108$ to align with the conditions described in [216, 217]. We also consider the FEC-independent sequence selection pro-

posed in [3] and ESS [229] for distribution matching. We also use a 4D mapping strategy [224] of our shaped amplitudes which has usually shown better performance than 1D and 2D mapping [38]. Our transmitted signal comprised 5 wavelength-division multiplexing (WDM) channels, each operating at 50 GBaud, resulting in a raw data rate of 600 Gbit/s per channel. For all the tested ESS schemes, we employed a low-density parity-check (LDPC) code with a length of 64800 bits, following the DVB-S2 standard, and a code rate $r_c = 5/6$. Consequently, the achieved net bit rate amounted to 8 bits per 4D symbol, resulting in a net bit rate of 400 Gbit/s per channel. The channel spacing is set to 55 GHz, and we apply root-raised cosine (RRC) pulse shaping with a roll-off factor of 0.1 to each channel. SSMF is simulated with an attenuation coefficient $\alpha_{\text{dB}} = 0.2$ dB/km, chromatic dispersion coefficient $D = 17$ ps/nm/km, polarization mode dispersion (PMD) of 0.04 ps/$\sqrt{\text{km}}$, and a nonlinear parameter $\gamma = 1.3$ $(\text{W}\cdot\text{km})^{-1}$ at $\lambda = 1550$ nm. We do not add any laser phase noise in the simulation to simplify our analysis. Following propagation, the central channel underwent several signal processing steps, including optical filtering, matched filtering, chromatic dispersion compensation, and genie-aided multiple-input-multiple-output (MIMO) channel equalization. Subsequently, for carrier phase recovery (CPR), we applied a fully-data-aided phase filter with a window averaging over 64 symbols to compensate for phase rotation induced by cross-phase modulation (XPM), as elaborated in [210]. Finally, we measured the electric SNR denoted SNR_{elec} from the equalized constellations and computed the generalized mutual information (GMI) using Eq. (10) from [252].

For the higher baud rate long-haul link scenario, we evaluate the potential advantages of our E-SS and D-SS schemes employing digital-multi-band (DMB) signals known for their superior resistance to Kerr effects compared to single-carrier schemes [108, 253]. The baud rate in this scenario is set at 110 Gbaud, with 8 digital subcarriers, first in a single-wavelength configuration, then in a 5 WDM-channel configuration. The link is made of 30×80 km SSMF spans with one 5 dB-noise-figure EDFA amplifier per span. The achieved net data rate per channel is 880 Gbit/s. The digital signal processing (DSP) blocks employed in this scenario are consistent with those detailed in the preceding paragraph.

5.3.3 From EDI to D-EDI

EDI is a sign-independent metric proposed in [107]. We denote it as $\Psi[\mathbf{X}]$ in Eq. (5.5) defined as the ratio of the variance to the mean of the windowed energies in the sequence $\mathbf{G}^w$ defined in Eq. (5.6). For a given window length w and a $2 \times L_s$ sequence $\mathbf{X}$ defined in Eq. (5.7) as L_s 2D complex-valued column vectors or equivalently two L_s-long row vectors, each element G_i^w of $\mathbf{G}^w$ is computed as the sum of the energies of $w+1$ 2D symbols centered on $\mathbf{x}_i$ as shown in Eq. (5.8). $\text{Var}[\cdot]$ and $\text{E}[\cdot]$ are the variance and expectation of $\cdot$ respectively.

$$\Psi[\mathbf{X}] \triangleq \frac{\mathrm{Var}[\mathbf{G}^w]}{\mathrm{E}[\mathbf{G}^w]} \tag{5.5}$$

$$\mathbf{G}^w = \left[G^w_{\frac{w}{2}+1}, G^w_{\frac{w}{2}+2}, \cdots, G^w_i, \cdots, G^w_{L_s-\frac{w}{2}}\right] \tag{5.6}$$

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{L_s}] = \begin{bmatrix} \mathbf{x}_{\mathrm{pol}_1} \\ \mathbf{x}_{\mathrm{pol}_2} \end{bmatrix} \tag{5.7}$$

$$G^w_i = \sum_{k=i-\frac{w}{2}}^{i+\frac{w}{2}} \|\mathbf{x}_k\|^2, \quad i : \frac{w}{2}+1 \rightarrow L_s - \frac{w}{2} \tag{5.8}$$

The optimal EDI averaging window length w (where $w \geq 2$ and is an even number), expressed as a number of symbols, is directly linked to the two-sided channel memory, denoted $2M$ and given by $2M = 2\lfloor \pi |\beta_2| B R_s L \rceil$ where β_2 stands for the group velocity dispersion, B is the channel spacing, R_s is the symbol rate, L is the transmission distance, and $\lfloor \cdot \rceil$ denotes the rounding to the nearest integer. As observed in previous works [107] (specifically in their Fig. 9), the optimal window length w, defined as the length where the correlation between EDI and SNR is maximized, is typically less than $2M$.

To further capture the energy variations within a finite-length sequence propagating over a link, we introduce the EDI of dispersed sequences, abbreviated as D-EDI and defined as the average of EDI values computed at multiple points along a linear dispersive and lossless fiber. This is shown in Fig. 5.8 and in Eq. (5.9) in which D-EDI is denoted as Ψ_D and $\mathbf{X}^{(N)}$ designates the symbol sequence that has propagated over $N \times L_D$ kilometers of an ideal dispersive fiber obtained using Eq. (5.10) and (5.11), $\mathrm{m_D}$ is the maximum number of dispersion operations and $\Psi[\cdot]$ is the EDI operator defined in Eq. (5.5).

$$\Psi_D \triangleq \frac{1}{m_D+1} \sum_{N=0}^{m_D} \Psi\left[\mathbf{X}^{(N)}\right] \tag{5.9}$$

$$\mathbf{X}^{(N)} = \mathfrak{D}_{D,N\times L_D}[\mathbf{X}] \tag{5.10}$$

$$\mathfrak{D}_{D,z}[\mathbf{X}] \triangleq \begin{bmatrix} \mathfrak{F}^{-1}\left(\mathfrak{F}\left(\mathbf{x}_{\mathrm{pol}_1}\right) \odot \exp\left(j\frac{D\lambda^2\pi\Delta\mathbf{f}^2 z}{c}\right)\right) \\ \mathfrak{F}^{-1}\left(\mathfrak{F}\left(\mathbf{x}_{\mathrm{pol}_2}\right) \odot \exp\left(j\frac{D\lambda^2\pi\Delta\mathbf{f}^2 z}{c}\right)\right) \end{bmatrix} \tag{5.11}$$

The dispersion operator in Eq. (5.11) consists in a chromatic dispersion filter applied in the frequency domain and capturing the dispersion of a fiber with dispersion coefficient D and length z km. $\mathfrak{F}$ and $\mathfrak{F}^{-1}$ are the fast Fourier transform (FFT) and inverse fast Fourier transform respectively, $\odot$ the element-wise multiplication, $j = \sqrt{-1}$ the imaginary unit, λ the central wavelength

of the signal, c the speed of light in vacuum and $\mathbf{\Delta f}$ the vector of frequency bins over which the FFT is computed. Dispersion is digitally applied over the perfectly-synchronized unfiltered symbol sequence sampled at one sample per symbol. We clearly see that, unlike EDI, D-EDI requires the knowledge of sign bits to compute the form of a dispersed sequence. EDI is its special case when $m_D = 0$.

In a single-span link, the signal power is substantially high at the start of the span, leading to significant NLI. To capture waveform variations within the first few kilometers over which the power loss is relatively minimal, we can compute the EDI multiple times shortly after the transmission begins within this span to derive the D-EDI. In what follows, to keep the computational complexity low, for a single-span link, we add up two EDIs, the one computed over the sequences at the transmitter output and the one computed after propagation over the effective length of the fiber span $L_{D(\text{single span})} = L_{\text{eff}} = \frac{1-\exp(-\alpha L)}{\alpha}$ [77] where α is the fiber attenuation in km^{-1} and L is the length of the fiber span in km.

For a multi-span link consisting of m spans of L kilometers each, we may compute EDI several times within each span to capture the evolution of the waveform distortion. In this paper, to limit the complexity of the metric computation, we calculate the EDI of the sequence at the beginning of each span and combine them to obtain the D-EDI, hence $m_D = m - 1$ and $L_D = L$. We may reduce the complexity by computing EDIs less often, i.e. reducing m_D and increasing L_D as shown in Sec. 5.3.6.

Compared to the SSFM method [3], D-EDI is less complex. For a simple complexity discussion, we can compare the number of FFTs used in each span. For our D-EDI, we only need to apply one FFT and one IFFT per span, whereas SSFM needs to add dispersion and nonlinearity progressively through several steps per span, requiring one FFT and one IFFT in each step, and for a 1-km step, the transmission of 80 km per span requires at least 80 FFTs and 80 IFFTs each, which results in a much higher complexity than D-EDI.

5.3.4 Implementation of EDI and D-EDI: E-SS and D-SS design

The fundamental concept behind E-SS and D-SS draws inspiration from the list-CCDM [237] and the sequence selection in [3]. Our approach extends the concept for the use of various distribution matchers (DMs) and we concatenate the output of several DMs to generate a longer candidate sequence. In this study, we only use ESS for distribution matching.

In Fig. 5.8, we plot the encoding process where input bits are fed into a 1D ESS distribution matcher (DM) with block length l that generates l amplitudes. The input bit stream is composed of two parts: $k - \nu$ information bits and ν prefix flipping bits. Altering these ν prefix flipping bits can cause pronounced changes in the indices of the sequences during the ESS encoding process. Such changes are reflected as considerable alterations in the amplitudes across the block. Conveniently,

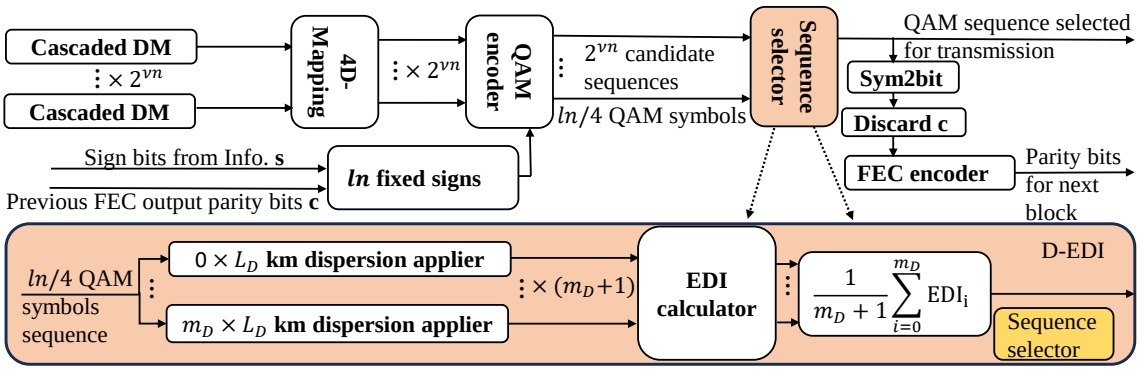


Figure 5.8: Block diagram of FEC independent D-SS in the PCS transmitter: EDI of dispersed sequences (D-EDI) is determined by averaging the calculated EDI at multiple locations along an ideal dispersive fiber. The dispersion is applied at 1 sample per symbol. EDI is a special case of D-EDI where $m_D = 0$. The sign bits are fixed through a multi-block FEC-independent sequence selection process as shown in [3].

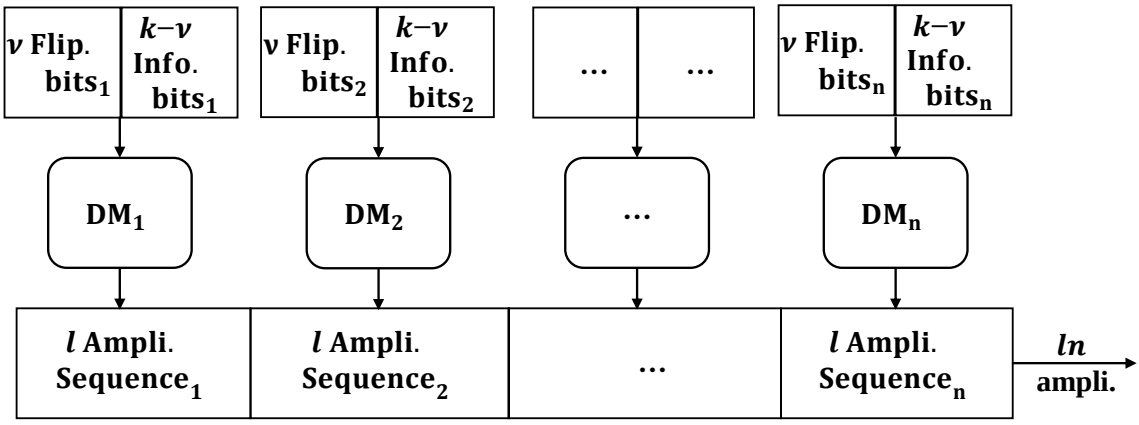


Figure 5.9: Structure of cascaded distribution matcher (DM). The output amplitude sequences of n DMs are concatenated to form a longer sequence.

these flipping bits are discarded after the decoding process at the receiver side. The outputs from n 1D-DM can be sequentially concatenated, from head-to-tail, to create extended sequences, which can then undergo a selection process as shown in Fig. 5.9. This was already mentioned in [3], in which the authors precise that the length of the candidate sequence and the block length can be optimized almost independently. This resulting sequence will have a length of ln. Cascading DMs have two advantages: they permit obtaining more candidate sequences while maintaining the same rate loss, and at the same time, we maintain the use of short-block-length DMs that showed a clear advantage against long-block-length DM with respect to nonlinear effects. For an E-SS (also true for D-SS) with n cascaded DMs, where each DM includes ν flipping bits, we can generate $2^{\nu n}$ different candidate sequences. Subsequently, the ln amplitudes are mapped to $L_s = ln/4$ 4D-QAM symbols after 4D-mapping [224].

E-SS employs EDI as a metric for sequence selection, hence the assigned signs to the generated

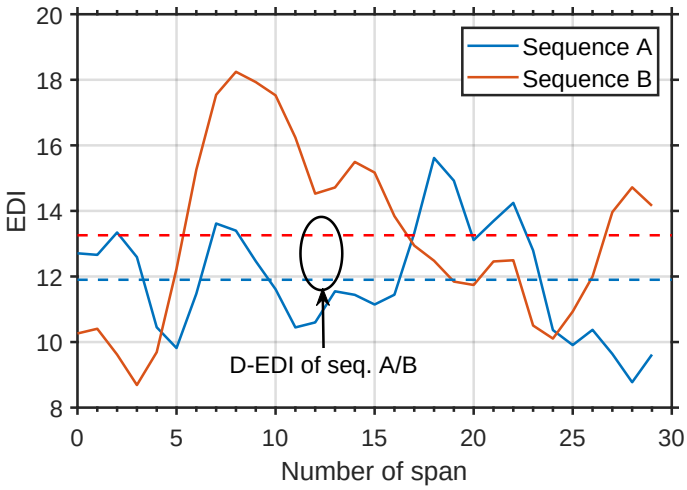


Figure 5.10: Simulation in single channel, 8 DMB subcarriers, 110 GBaud transmission over 30 × 80 km SSMF link, with a net rate of 880 Gbit/s. Sequences A and B are two random sequences generated by single ESS with block length $l = 108$ over the 4^{th} subcarrier (i.e. central subcarrier), each consisting of 108 4D 64-QAM symbols. Solid lines represent the evolution of EDI at the beginning of each span, dashed lines represent D-EDI computed as an averaged EDI over all spans. D-EDI gives a more accurate estimation of NLI generated by each sequence after transmission.

amplitudes are not necessary for the evaluation of EDI. The sequence selector then identifies the $2^{\nu n}$ different candidate sequences and transmits the sequence with the lowest 4D-energy EDI.

D-SS employs D-EDI as a metric for sequence selection, which requires the assigned signs for the metric evaluation and selection. The sign bits can be fixed through a multi-block FEC-independent sequence selection process as proposed in [3]. In this scenario, the sign bits are chosen in two parts: a part from the information stream of the current D-SS block, denoted **s**, and a part from the FEC output parity bits of the previous D-SS block, denoted **c** as shown in Fig. 5.8. Once the QAM sequence is selected, it is demapped into a bit sequence and sent into an FEC encoder, after discarding the parity bits from the previous D-SS block **c**, to generate parity bits for the next D-SS block.

The sequence selection process introduces an additional rate loss [36, 239], which is $\frac{\nu}{l}$ accounting for the effect of the flipping bits as in [38]. The total rate loss R_{loss} of E-SS and D-SS in each dimension is:

$$R_{\text{loss}} = H(P_a) - \frac{k-\nu}{l}. \tag{5.12}$$

where $H(P_a)$ is the entropy linked to the distribution P_a and P_a is the marginal distribution of amplitudes of 1D PAM symbols. As ν increases, the number of candidate sequences also increases, resulting in higher nonlinear shaping gain. However, this comes at the cost of an increased rate loss. As a result, an efficient sequence selection mechanism should strike a balance between linear and nonlinear shaping gains.

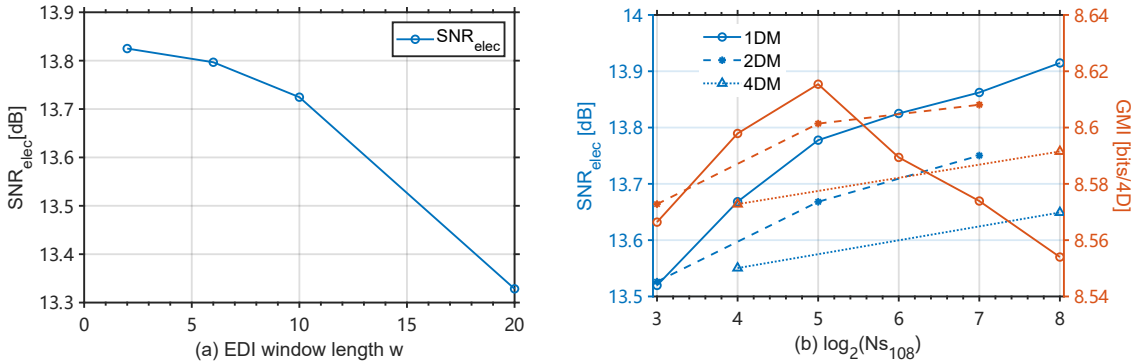


Figure 5.11: Simulation of 50 GBaud E-SS transmission over 205 km single span SSMF link, 5 WDM channels, 55 GHz channel spacing, net rate per channel 400 Gbit/s. (a): Electrical SNR versus the EDI window length w used in the sequence selector to evaluate the EDI of sequences. $\nu = 3$ flipping bits and a single 1D-DM ($n = 1$) is used. The power per channel is 9 dBm. (b): Electrical SNR and GMI at 9 dBm versus $\log_2(N_{S_{108}})$ when cascading 1, 2 or 4 DMs, where $N_{S_{108}}$ is the number of required sequence selections to generate 108 4D-symbols. E-SS$_1^3$ exhibits the highest performance in terms of GMI. (ESS without SS exhibits 13.27 dB electrical SNR value and 8.46 bits/4D GMI value).

In Fig. 5.10, we illustrate the evolution of the EDI metric at the beginning of each span when transmitting two different random sequences over a 30×80 km link with a single channel, 8 DMB subcarriers, 110 GBaud transmission, and a net rate of 880 Gbit/s. Each of these sequences comprises 108 random 4D 64-QAM symbols over the 4$^{\text{th}}$ subcarrier. EDI is calculated at the beginning of each span and both EDI and D-EDI employ a window length of 2 symbols ($w = 2$), that is 1 symbol before and 1 symbol after the current symbol. Before transmission, sequence B exhibits a lower EDI compared to sequence A. This implies that sequence B will generate less NLI over the first span than sequence A. However, as the sequences propagate through the dispersive fiber, their characteristics change continuously due to the influence of dispersion. After only 4 spans, the EDI of sequence B surpasses that of sequence A. Ultimately, when the D-EDI is computed by averaging EDI over all spans, sequence A yields a smaller D-EDI even though it had a larger initial EDI. This suggests that sequence A maintains a higher average nonlinear shaping gain throughout the entire transmission. This outcome emphasizes the importance of considering sign-dependent metrics, such as D-EDI, when assessing the nonlinear shaping gain performance of sequences. Sign-independent metrics may not provide a complete picture of NLI manifestation, as the influence of dispersion can cause initially favorable sequences to show deteriorated performance after a certain propagation distance, while initially unfavorable sequences may show an improved performance. Hence, D-EDI allows for a more accurate estimation of NLI generated by each sequence throughout a transmission.

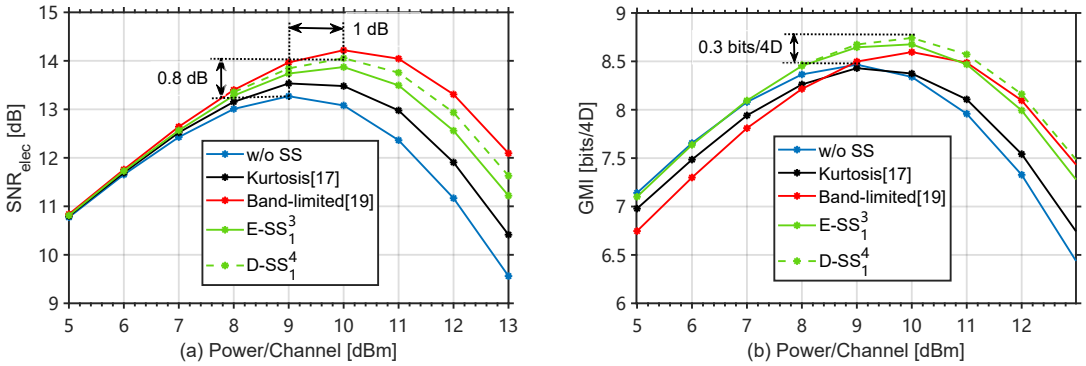


Figure 5.12: Simulation of ESS without SS, K-ESS, B-ESS, E-SS$_1^3$, and D-SS$_1^4$ (best-performing scheme) with D-EDI computed after applying a single 21 km dispersion filter. Same system configuration as for Fig. 5.11. Left: Electrical SNR versus power per channel. Right: GMI versus power per channel. Both E-SS$_1^3$ and D-SS$_1^4$ demonstrate superior performance when compared to state-of-the art schemes.

5.3.5 Performance assessment: Single-span transmission

The single-span transmission scenario was introduced in Sec. 5.3.2. First, we focus on the E-SS scheme. In Fig. 5.11 (a), we plot the $\mathrm{SNR}_{\mathrm{elec}}$ for various EDI window lengths w used by the sequence selector. For this test, we used a single 1D-DM ($n = 1$) and $\nu = 3$ flipping bits and denote the scheme E-SS$_1^3$ (or in general E-SS$_n^\nu$). This setup allows us to select the most suitable sequence from eight possible candidates ($2^3 = 8$). We calculate $\mathrm{SNR}_{\mathrm{elec}}$ using equalized symbols after transmitting over a single span with a channel power of 9 dBm, which is near the optimal power level. Our findings show that as w increases, the $\mathrm{SNR}_{\mathrm{elec}}$ gradually declines. We find that a window length $w = 2$ yields the best performance, and is much lower than the channel memory ($2M = 76$ when considering the total length of the span) as the NLI is generated mostly at the beginning of the single-span transmission. Therefore, we proceed with $w = 2$ in subsequent single-span simulations.

In Fig. 5.11 (b), we analyze the performance of E-SS by varying the number of sequentially cascaded DMs n, where each DM has ν flipping bits. A single DM with a block length $l = 108$ produces 27 4D-symbols after 4D-mapping. When cascading $n = 1, 2$ or 4 DMs, the candidate sequence lengths will be $27, 54$ and 108 respectively. As a complexity metric, we designate the number of tested candidate sequences per 4 DMs, i.e. per 108 4D-symbols as $\mathrm{N}_{\mathrm{S}_{108}}$. We choose 4 DMs as 4 is the least common multiple between the three studied values of $n = \{1, 2, 4\}$. To achieve a total sequence of 108 4D symbols, the E-SS transmitter needs to be called $\frac{4}{n}$ times, resulting in $4, 2$ and 1 calls for $n = 1, 2$, and 4 cascaded DMs respectively. Hence, $\mathrm{N}_{\mathrm{S}_{108}}$ is given by:

$$N_{S_{108}} = \frac{4}{n} 2^{\nu n} \tag{5.13}$$

We see from Fig. 5.11 (b) that, as $N_{S_{108}}$ increases, SNR_{elec} also increases for $n = \{1,2,4\}$. However, as the number of flipping bits increases, the rate loss likewise increases. The maximum GMI is hence attained when achieving a trade-off between rate loss and nonlinear gain. Specifically, the largest GMI is achieved with a single 1D-DM at $\log_2(N_{S_{108}}) = 5$ (32 candidate sequences), corresponding to the E-SS$_1^3$ configuration.

Next, we also evaluated the performance of the D-SS scheme over the same link by computing D-EDI as the sum of the EDI of the sequence at the transmitter output and its EDI after propagating over a linear dispersive fiber without attenuation of length L_{eff} over which nonlinear effects remain significant. $n = 1$ and $v = 4$ (D-SS$_1^4$) yields the best GMI gains through a similar optimization as the one shown in Fig. 5.11(b) for E-SS. In Fig. 5.12 (a), we show SNR_{elec} versus the launch power per channel for five different ESS schemes. The D-SS, E-SS, band-limited ESS (B-ESS), and kurtosis-limited ESS (K-ESS) exhibit SNR gains of approximately 0.8 dB, 0.6 dB, 1 dB, and 0.3 dB respectively, compared to the ESS without sequence selection (SS) at a power level of 10 dBm. Additionally, they demonstrate up to 1 dB improvement in the optimal launch power. In Fig. 5.12 (b), we present the results in terms of GMI per 4D symbol. Notably, the D-SS scheme yields the highest GMI. Specifically, at a power level of 10 dBm, it achieves a GMI increase of 0.15 bits/4D compared to B-ESS and a 0.3 bits/4D improvement compared to ESS without SS. The E-SS scheme also delivers a GMI gain of 0.1 bits/4D (respectively 0.2 bits/4D) in comparison to B-ESS (respectively ESS without SS). It is important to mention that both EDI and D-EDI based shaping approaches exhibit nearly the same linear performance as unconstrained ESS due to their small rate loss. E-SS and D-SS exhibit similar performance largely because the initial shape of the waveform at the start of the link plays a crucial role in the generation of NLI for single span links. However, a slight performance improvement is achievable by selecting sequences that also exhibit low EDI after the effective length L_{eff}. In terms of rate loss, D-SS demonstrates superiority over B-ESS and K-ESS, with a rate loss of 0.26 bits/4D compared to 0.5 bits/4D and 0.27 bits/4D respectively. The rate losses are not discernible in Fig. 5.12 (b) between 5 dBm and the optimal power level because nonlinear gains are already occurring within these power ranges. D-SS also outperforms E-SS in the nonlinear regime. Overall, it outperforms B-ESS in both linear and nonlinear regimes. Finally, we remind that D-SS requires a slightly higher complexity compared to E-SS since it necessitates knowledge of sign bits and computation of dispersed sequences at the transmitter side.

5.3.6 Performance assessment: Multi-span transmission

Optimization and Transmission Performance

The multi-span transmission scenario was introduced in Sec. 5.3.2. For the DMB scheme, we select sequences for a single digital band and then distribute them over the eight bands. Perform-

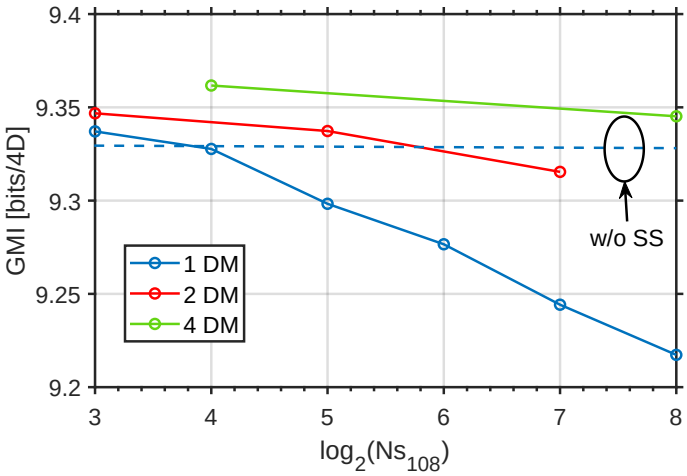


Figure 5.13: Simulation of D-SS with D-EDI computed after applying dispersion successively over 30×80 km, single channel, 8 DMB subcarriers, 13.75 GBaud per subcarrier transmission, net rate per channel 880 Gbit/s. GMI versus $\log_2(N_{S_{108}})$ at the optimal power (4.5 dBm), where $N_{S_{108}}$ is the number of tested sequences per 108 4D-symbols. ($w = 2$ for all multi-span tests, corresponds to the memory of a 13.75 Gbaud transmission over a single 80 km span). D-SS$_4^1$ exhibits the best performance.

ing sequence selection jointly for all digital bands would introduce significant complexity that we avoid in this work. In Fig. 5.13, we evaluate the performance of our D-SS scheme for different cascaded DMs and different numbers of flipping bits. In long-distance links, accumulated chromatic dispersion becomes apparent as its induced inter-symbol interference involves hundreds of symbols. For these links, it becomes more important to cascade multiple 1D-DMs to generate extended candidate sequences over which we will compute D-EDI. By doing so, we account for the impact of the accumulated channel memory on the waveform, thus making a better selection process. In addition, for the same complexity level as a single DM, this approach helps in reducing rate loss. As we did in the previous section, the complexity of the scheme is evaluated as $N_{S_{108}}$, the number of tested candidate sequences for the generation of 108 4D-symbols. We calculate D-EDI by summing up EDI values of the sequences at the input of each of the 30 spans constituting the 30×80 km link ($w = 2$ corresponding to the memory of a 13.75 Gbaud transmission over a single 80 km span). We observe that the largest GMI occurs in the configuration where four 1D-DMs are arranged in series, and each 4D-DM has one flipping bit, denoted D-SS$_4^1$, and corresponding to $\log_2(N_{S_{108}}) = 4$ (i.e. 16 candidate sequences). The GMI curve for a single DM exhibits a significantly steeper slope compared to the one for four DMs, as the rate loss increases at a slower pace in the case of cascading four DMs.

Next, in Fig. 5.14, we compare the performance of four single-wavelength transmission schemes: ESS without SS over a single carrier (SC) serving as a baseline, ESS without SS over DMB using 8 subcarriers (8DMB), D-SS$_4^1$ (optimal case) and E-SS$_4^1$ schemes, both in 8DMB mode. The D-SS$_4^1$ shows a 0.03 bits/4D-symbol gain in GMI at a power of 4.5 dBm compared to ESS without SS

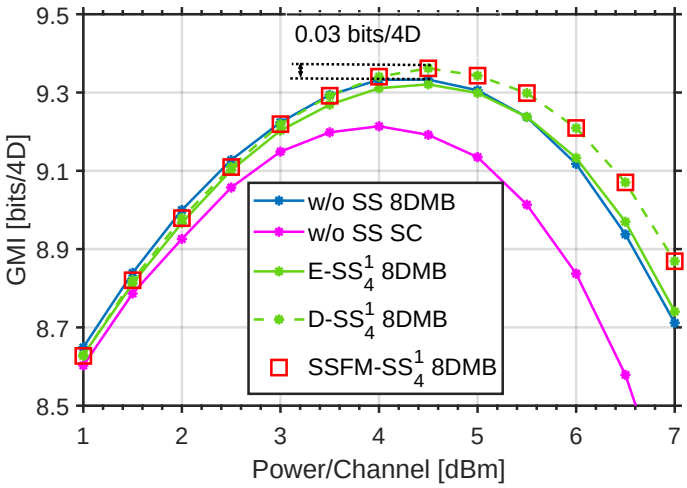


Figure 5.14: Simulation of ESS without SS, E-SS$_4^1$, D-SS$_4^1$ and SSFM-SS$_4^1$ with 30×80 km dispersion applier over 1 channel 8 DMB subcarriers, and ESS without SS with 1 channel 1 carrier, in 110 GBaud over 30×80 km SSMF link, net rate per channel 880 Gbit/s. (a): Electrical SNR versus Power per channel in the non-linear channel. (b): GMI versus Power per channel. D-SS$_4^1$ exhibits the same performance as that of an ideal SSFM-based sequence selection.

and 8DMB. It also shows a 0.16 bits/4D in GMI at the optimal power compared to ESS without SS over a single carrier. On the other hand, the E-SS$_4^1$ has no or very limited gain in GMI. Hence, the sign-dependent-based D-SS scheme maintains the GMI advantage in both short-distance and long-haul distance transmission scenarios.

Additionally, we conducted simulations involving sequence selection based on the computation of the actual NLI produced by each sequence using the Split-step Fourier method (SSFM sequence selection or SSFM-SS) to approximate an upper bound for the performance of a sequence selection scheme. We kept the transmitter structure as 4 concatenated DMs, each equipped with 1 flipping bit. This approach is similar to the one outlined in [3]. Each candidate sequence was transmitted over a 30×80 km standard single mode fiber, single-wavelength, single-carrier system operating at 13.75 Gbaud, without the addition of ASE noise. After transmission, we compare the NLI power of candidate sequences and select the sequence that generates the lowest NLI. This process allows us to obtain the optimal performance of sequence selection with 4 serial DMs, and 1 flipping bit when only considering self-phase modulation effects. From Fig. 5.14, we observe that SSFM-SS and D-SS exhibit almost identical performance. This observation indicates that within the analyzed system, selecting sequences based on D-EDI can be a viable alternative to the more complex SSFM approach. In other words, the variations of the waveform shaped by dispersion effects primarily dictate nonlinear distortions in high-power regions. Nonlinearity, in this context, acts as a small perturbation. Consequently, we can feasibly substitute SSFM with the dispersion-aware metric D-EDI to perform sequence selection with reduced complexity.

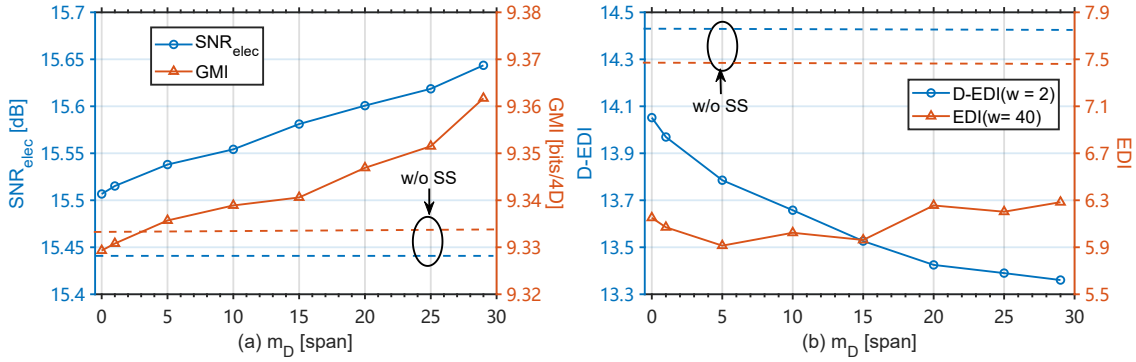


Figure 5.15: Performance analysis of D-SS$^4_1$ (a): GMI versus m_D at optimal power (4.5 dBm) where m_D is the maximum number of spans used by the dispersion applier during sequence selection. Gains are maintained even when we reduce the complexity by computing D-EDI over fewer spans. (b): D-EDI and EDI versus m_D. We add the values for ESS without SS by horizontal dashed lines. Same system configuration as in Fig.5.13. D-EDI demonstrates a negative correlation with SNR_{elec}.

Complexity Reduction and Interpretation of the Results

In the above study, we aimed to improve sequence selection by calculating the EDI at the beginning of each span along the entire transmission link. This resulted in obtaining 30 different EDI values averaged to derive the final D-EDI. Now, we delve into the impact of reducing the number of EDI computations on the achieved performance. To do so, we explore two methods. First, we consider calculating D-EDI for only the first $m_D = \{1, \cdots, 29\}$ spans instead of the entire transmission line. It is worth reminding that when $m_D = 0$, D-EDI simplifies to the original EDI. Second, we consider limiting the number of EDI calculations to $N_D < 30$ by computing EDI only at the beginning of specific spans described in the caption of Fig. 5.16, instead of computing it at the beginning of each span ($N_D = 30$).

In Fig. 5.15, we explore the first method. In Fig. 5.15 (a), we present the performance results in terms of SNR_{elec} and GMI. We include the values for ESS without SS as horizontal dashed lines for comparison. SNR exhibits a steady growth as m_D increases. The peak value is achieved when $m_D = 29$, implying that D-EDI calculations across the entire link yield the maximum benefit. After only three spans, the nonlinear gain brought by D-EDI-based sequence selection overcomes the penalty in increased rate loss, resulting in a higher GMI than the one obtained with ESS without SS. In dispersion unmanaged transmission systems, as analyzed by the GN [96] and EGN models [254], nonlinear penalties are equally distributed and independent across homogeneous spans. Consequently, as m_D increases, the calculation for D-EDI encompasses an increasingly larger number of spans. This inclusion progressively mitigates the NLI across all spans. As a result, the system performance improves linearly with the increase of m_D. Hence, to get the most out of sequence selection, it is crucial to calculate D-EDI for the entire link. Furthermore, it is important to note that the calculation of D-EDI only accounts for the influence of dispersion.

However, in actual optical transmission scenarios, as the transmission distance increases, NLI also escalates, subsequently impacting energy fluctuations. Consequently, the accuracy of D-EDI calculations will be generally higher when computed over a specific number of initial spans compared to a computation over an equivalent number of final spans. This higher accuracy in the early stages of transmission translates into better performance improvements following sequence selection.

Thereafter, in Fig. 5.15 (b), once we have selected all the candidate sequences by D-EDI for different m_D values, we illustrate the EDI and D-EDI values evaluated for the whole transmitted sequence. EDI is evaluated over 40 symbols ($w = 40$), which is an empirical value lower than the total channel memory $2M = 68$ (corresponding to the channel memory of a 13.75 Gbaud transmission over the entire link (30×80 km)) as explained in Sec. 5.3.4, whereas D-EDI is assessed by averaging EDIs computed with a memory of 2 symbols ($w = 2$), corresponding to the memory of a 13.75 Gbaud transmission over a single 80 km span. This D-EDI used for performance evaluation of the selected sequences is calculated along the entire link. Notably, as m_D increases, D-EDI continues to decrease and hence exhibits a negative correlation with SNR_{elec}, while EDI struggles to do so, showing that D-EDI is a better estimator of the performance of the selected sequences for multi-span transmissions. Interestingly, we observe that the EDI metric, computed over sequences selected based on D-EDI, is flat with small fluctuations, while according to Eq. (5.5) it should not. The fluctuations may be due to a statistical error since we evaluate EDI on finite sequences. It is also possible that D-EDI-based sequence selection selects a subset of sequences that have different statistics depending on m_D. A detailed analysis is beyond the scope of this work but could provide further insights on how to design a practical scheme.

Next, we explore the second reduced-complexity D-EDI computation method. Instead of computing EDI at the start of each span, we choose to compute it for fewer spans specified in the caption of Fig. 5.16. The figure shows the performance variations in terms of GMI versus the number of required EDI calculations N_D. We observe that, with decreasing N_D, GMI experiences a gradual decline. Significantly, even with $N_D = 5$ (corresponding to EDI calculations every 800 km), the system still outperforms ESS without SS. In this case, the computational complexity of D-EDI is significantly reduced compared to $N_D = 30$. This finding suggests that, for digital multi-band (DMB) transmissions over long-haul links, we can still achieve gains while using a metric with a reduced computational complexity for sequence selection.

To conclude this subsection, we highlight two key observations from the analyses in Fig. 5.15 and Fig. 5.16 for DMB transmissions over long-haul links: first, a sequence that performs well for D-EDI over fewer spans may not necessarily maintain its effectiveness over longer links. This observation underscores the complexity of sequence selection in optical communications. It is not sufficient to assess the suitability of a sequence based solely on its performance at a single point

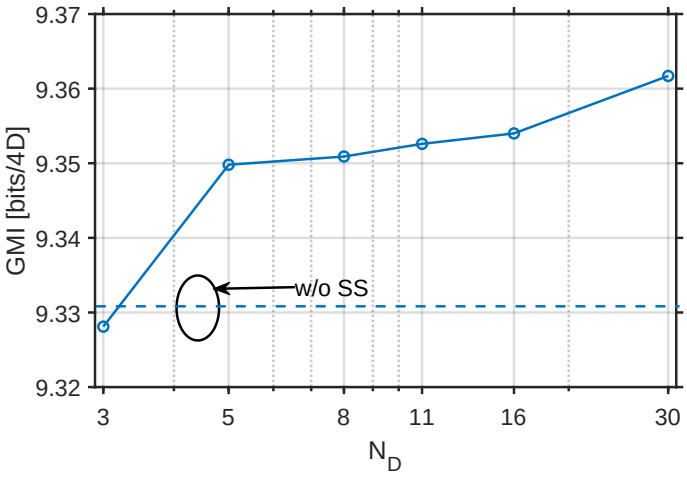


Figure 5.16: GMI versus N_D for D-SS$^4_1$, where N_D is the number of EDI calculations required to compute D-EDI, each N_D value represents the following span indices: $N_D = 3 \rightarrow [0,1,29]; 5 \rightarrow [0,1,9:10:29]; 8 \rightarrow [0,1,4:5:29]; 11 \rightarrow [0,1,2:3:29]; 16 \rightarrow [0,1:2:29]; 30 \rightarrow [0:1:29]$. We add the values for ESS without SS by horizontal dashed lines. Same system configuration as in Fig. 5.13. Gains are maintained even when we reduce the complexity by computing D-EDI over several spans.

in the link, such as the beginning of a link or elsewhere, or at a part of the total distance, which is also proved by Fig. 5.10, in which the EDI for two random sequences exhibits continuous variations over the transmission distance. Second, identifying an optimal sequence does not necessarily require an exhaustive analysis at the start of every span. Although dispersion-induced interference accumulates over longer distances, causing significant power variations and nonlinear phase fluctuations, our goal is to identify sequences that has a lower D-EDI over the entire distance. Since interference typically builds up over an extended period, involving the overlap of several symbols, it is practical to evaluate sequences at intervals spanning several spans rather than continuously or for every span.

The Impact of Block Length and Sequence Length Variations

Up to now, we have used a fixed block length of 108 for the comparison of our proposed schemes with state-of-the-art schemes in similar conditions. It is important to note that in the case of ESS, the block length has a significant impact on the performance (GMI). Decreasing the block length leads to increased nonlinear gain but also results in greater rate loss. Consequently, the optimal block length for ESS strikes a balance between maximizing nonlinear gain and minimizing rate loss. In Fig. 5.17, we plot the performance of ESS schemes for four different block lengths $l = \{60, 108, 200, 300\}$. The considered schemes are ESS without SS alongside D-SS with 16 candidate sequences and D-SS with 256 candidate sequences. With the increase of cascaded DMs and flipping bits, the number of candidate sequences can rapidly increase beyond 16 or 256. For example, when 4 serial DMs, each with 3 flipping bits, are concatenated, there are $2^{4\times3} = 4096$

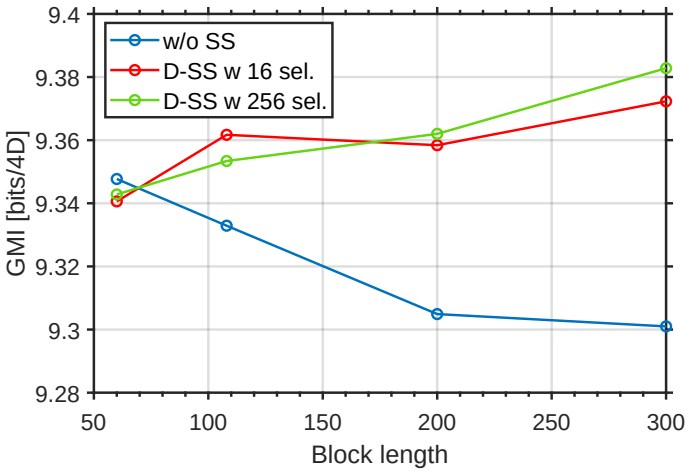


Figure 5.17: GMI versus DM block length. We show the optimal performance for each block length by changing the number of flipping bits and the number of cascaded DMs. Same system configuration as in Fig.5.13. D-SS demonstrates superior performance over different block lengths.

candidate sequences. To limit and maintain a relatively fixed complexity, we randomly select 16 or 256 sequences for selection to examine both low and high complexity schemes. In Fig. 5.17, for each block length and each number of tested candidate sequences, we only show the best D-SS performance found after testing various configurations of cascaded DMs and flipping bits. The main observations are:

- For the ESS without SS, as the block length decreases, the rate loss gradually increases. Concurrently, the shortest block length of 60 provides a nonlinear gain that exceeds the relatively high rate loss, resulting in the maximum achievable GMI.
- For D-SS, at a block length of 60, the rate loss incurred by ESS itself, along with the additional rate loss from flipping bits, becomes too substantial for the achievable rate. Even though D-SS exhibits a larger nonlinear gain, the final GMI falls short of what ESS can offer.
- For longer block lengths, D-SS consistently outperforms ESS due to its ability to maintain a significant nonlinear gain while mitigating excessive rate loss. Among these block lengths, D-SS with a block length of 300 achieves the highest GMI.
- In particular, when the block length is equal to 300, D-SS with 256 candidate sequences shows a 0.08 bits/4D GMI advantage over ESS with the same block length and a 0.04-bits/4D advantage over ESS with a block length of 60. D-SS with 16 selections shows similar trends with slightly lower gains.

For a thorough analysis of the performance of D-SS, we vary the sequence length under identical system configuration and report the performance in Fig. 5.18. We explore D-SS configurations

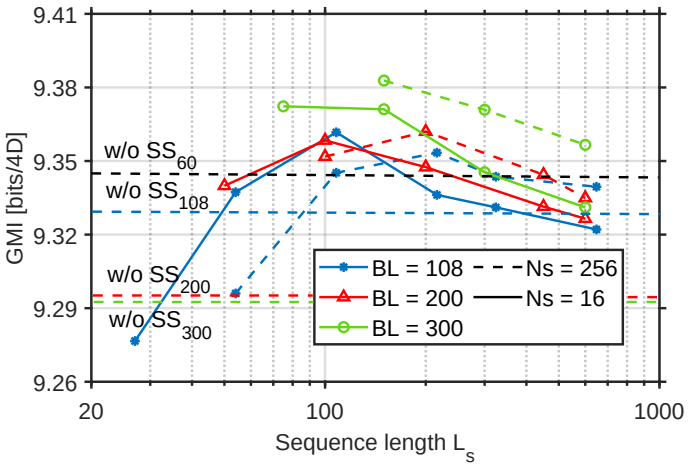


Figure 5.18: GMI versus selected sequence length. We show the optimal performance versus different sequence lengths, for various block lengths, by changing the number of flipping bits and the number of cascaded DMs. Same system configuration as in Fig.5.13. D-SS demonstrates superior performance over different sequence lengths.

with three different block lengths and vary the number of cascaded DMs, resulting in multiple sequence lengths. For each sequence length, we report the optimal performance obtained by adjusting the number of flipping bits and by performing the selection among 16 (full lines) or 256 (dashed lines) candidate sequences. For ease of comparison, the performance of ESS with four block lengths is also depicted on the graph using horizontal dashed lines. When the block length is set to 300, D-SS shows the best performance due to lower rate loss and comparable nonlinear behavior with respect to shorter block lengths. 256 selections tend to perform better for longer sequences because longer sequences necessitate cascading more DMs in series, hence providing more variability in the candidates for selection. With 256 selections, D-SS can more effectively explore and utilize these possibilities. D-SS significantly outperforms ESS across almost the entire range of tested sequence lengths for block lengths of 200 and 300. D-SS with optimal sequence-length range (100 to 200 4D symbols) also surpasses the performance of ESS with block lengths of 60 and 108. The optimal sequence length for all four block lengths falls within the range of 100 to 200 4D symbols. We explain the observed behavior by the fact that overly long sequences capture more channel memory, yet they lead to coarser selections (at a fixed number of selections to limit the complexity). On the other hand, while very short sequences offer finer selections, they only capture a small fraction of the channel memory. Consequently, for the specific transmission scenario discussed in this article, the balance point between channel memory and selection precision lies within the range of 100 to 200 4D symbols.

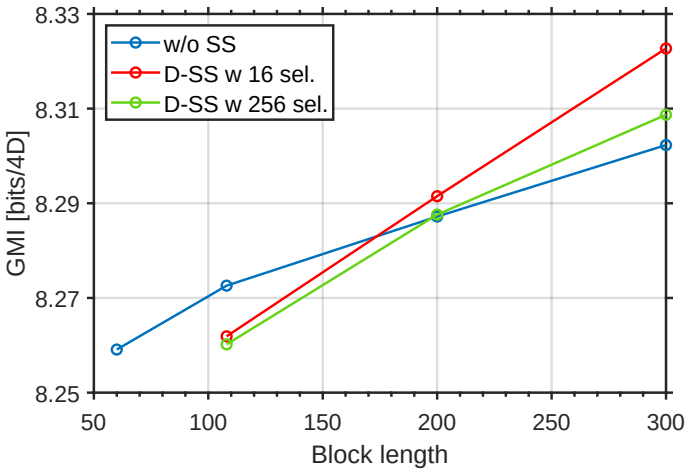


Figure 5.19: GMI versus DM block length. We show the optimal point versus block length by changing the number of flipping bits and the number of cascaded DMs. The same system configuration as in Fig.5.13, however with 5 WDM channels. D-SS demonstrates superior performance in WDM transmission scenario.

WDM Transmission Performance

Finally, we maintain the same system configuration as before but extend it to a 5-channel WDM transmission. The combined net transmission rate for all five channels is 4.4 Tbits/s. In our earlier investigation, we observed poor performance for D-SS with a block length of 60. Consequently, we assess the performance starting with a block length of 108. In Fig. 5.19, we see that the nonlinear gain associated with short block lengths in ESS without SS diminishes because the CPR compensates for the increased inter-channel nonlinearity. There is an observed increase in GMI with increasing block length due to the reduced rate loss, yielding an optimal block length of 300 for ESS. Conversely, D-SS better maintains nonlinear gains as the block length increases and its rate loss continues to decrease. This results in D-SS achieving the maximum GMI at a block length of 300. Particularly, with 16 selections, D-SS with a block length of 300 outperforms ESS by 0.02 bits/4D-symbol, and it achieves a 0.06 bits/4D-symbol improvement compared to ESS with a block length of 60. Lastly, it is important to note that 256 selections, which require more flipping bits, lead to higher rate loss. The additional nonlinear gain does not fully compensate for the performance degradation caused by the increased rate loss, resulting in a lower GMI than with 16 selections. In a WDM configuration, the gains are lower than the ones observed in a single-channel system as a high number of cross-phase modulation components arise and our selection metric does not account for the energy dispersion of the neighboring channels. Finally, since these sequences are first selected for a single digital band and then used for all bands, the sequences selected by SSFM-SS are the same as D-SS and exhibit the same performance.

5.4 Discussion and Summary

We have proposed a novel sign-dependent metric: D-EDI, which can serve as a prediction metric for the generation of NLI because it showed a negative correlation with the measured performance of long-distance transmission systems. By applying D-EDI (resp. EDI) in sequence selection, we defined the D-SS (resp. E-SS) scheme. These two schemes achieved higher gains with ESS used for distribution matching than state-of-the-art ESS schemes over short transmission distances. Noticeably, D-SS showed gains even for longer transmission distances and higher baud rates over single-channel and WDM systems even when CPR is used to reduce the impact of nonlinear phase rotations, achieving the same performance as sequence selection based on a full SSFM simulation [3]. Moreover, D-SS retained a robust performance at moderate complexity (lower number of digital dispersion operations to compute the metric) and outperformed conventional ESS across different block lengths. These results pave the way for further optimizations of shaping schemes for long-distance transmissions and confirm the necessity of using sign-dependent and channel-aware metrics for effectively evaluating nonlinear distortions.

Nevertheless, these findings tell us that there is still room for improvement. For DMB-based transmissions in which the transmission rate per subcarrier is lower than single-carrier transmissions, the impact of channel memory is limited to a few symbols per span or even less than 1 symbol per span, hence only a few dozen symbols for all spans. Thus, the impact of channel memory can be accounted for by concatenating several DMs to minimize the mutual influence of adjacent selected sequences. As the transmission rates continue to increase, for both multi-carrier and single-carrier transmissions, the channel memory can spread over hundreds or thousands of symbols. At the same time, increasing the length of candidate sequences also increases the complexity of the scheme. Hence, modifications and optimizations of D-EDI to achieve excellent performance predictions under high-rate long-distance transmission while maintaining a reduced complexity is still an open question. Other improvements of the metric may address different dispersion management schemes or take into account inter-channel non-linear effects. In this work, we only considered long-haul dispersion-unmanaged systems where the NLI contributions from different spans can be considered independent and D-EDI gives a good performance prediction by summing up EDIs of the sequence at the start of each span.

Ultimately, an extended goal involves conducting an in-depth analysis of the selected “good" sequences. By comprehensively understanding their properties, we aim to design simpler sequence generation solutions.

Conclusions & Outlook

Conclusions

In this thesis, we embarked on a comprehensive journey through the evolution of optical fiber transmission systems, offering significant insights into the modulation technologies shaping modern optical communications. Our exploration commenced with a foundational understanding of the components and DSP in coherent communication systems. This provided a thorough understanding of the technological advancements driving the transition to sophisticated digital communication methods.

We reviewed the progress in modulation and coding technologies, highlighting the transition from conventional modulation techniques to multi-dimensional (MD) modulation and Probabilistic Constellation Shaping (PCS). Our research findings made significant strides in the domain of long-haul optical fiber transmission through the design of a new four-dimensional constellation with an enhanced tolerance to non-ideal components, the combination of MD modulation and PCS, and PCS designs, all optimized for nonlinear fiber channels. These contributions address the inherent nonlinear nature of optical fibers, providing solutions that balance linear and nonlinear gains and mitigate Kerr nonlinear impairments, which limit capacity enhancements.

By introducing two innovative PCS designs and reduced-complexity shaping schemes based on a sign-dependent metric and sequence selection, we have shown that achieving higher gains and robust performance over different transmission distances and baud rates is possible. Our work also emphasizes the importance of accurate and low-complexity sign-dependent metrics in shaping schemes and sequence generation, underlining the need for continuous optimization to enhance long-distance transmission systems further.

Outlook

The work presented in this thesis on MD modulation and PCS merely begins to tap into the potential of these advanced techniques. Future research is imperative to further refine and optimize these methods for even higher spectral efficiency and improved performance in nonlinear regimes. Ongoing innovation in constellation design and the enhancement of shaping techniques hold the

promise of significantly increasing the capacity and efficiency of transmission systems. Particularly for MD modulation, achieving greater spectral efficiency typically necessitates a larger symbol set and more intricate mapping, thereby heightening sensitivity to Mach-Zehnder Modulator (MZM) imbalance and Digital-to-Analog Converter (DAC) variations. The scalability of these solutions, such as employing additional rings for enhanced spectral efficiency, presents a fertile ground for exploration.

Additionally, this thesis underscores the criticality of precise and low-complexity, sign-dependent metrics in shaping schemes and sequence selection. While we have introduced the D-EDI metric, its application is still somewhat limited. The challenge remains to adapt this metric for higher-rate scenarios while also significantly reducing computational complexity. Future research could concentrate on devising more accurate metrics and creating innovative sequence generation methods capable of managing the growing intricacies of future transmission systems that tend to use new spectral windows that bring new challenges. This includes efforts to minimize delays in sequence selection and Forward Error Correction (FEC) matching. Moreover, the quest should not end with sequence selection; instead, we must delve deeper into the underlying causes and physical principles of the beneficial sequences that are identified. By doing so, we can gain a more profound understanding of the fiber channel's fundamental properties and pave the way for more efficient and robust optical communication systems.

Appendix A

Digital Multi-Band Transmission

The introduction of digital sub-carrier multiplexing (DSCM) or digital multi-band (DMB) enables fine data rate adaptation, and water filling (as for OFDM) and can provide robustness against channel effects compared to single-carrier transmissions [102]. This is achieved thanks to water-filling techniques (availability of several operational schemes per subcarrier) and to improved robustness to chromatic-dispersion-compensation induced enhanced phase noise, also referred to as electronically enhanced phase noise (EEPN), especially for high data rates [65, 66].

DMB can also improve the tolerance to intrachannel non-linear effects. In [255], a single-wavelength experimental demonstration using dual-polarization QPSK and 16-QAM formats showed that a nearly optimal number of equal-width subcarriers can extend the transmission reach compared to a single-carrier format. This approach is particularly interesting and has led to many research works on the generation of digital subcarriers and the associated DSP at the receiver side to perform channel equalization. In [256], through simulations of single-wavelength and single-polarization QPSK-modulated DMB transmission, the authors observed that the optimal symbol rate decreased with the transmission distance over SSMF and explained this observation using the four-wave-mixing (FWM) theory and by studying the RF spectra of the intensity fluctuations of the signal along the link. In [257], a similar study is done for an 8-channel WDM transmission and the analysis of the optimum number of bands is done in light of the FWM efficiency [258, 259], which is closely linked to the intensity of the XPM effect (degenerate FWM), showing that a multi-band approach can beat the performance of a single-carrier approach.

In Fig. A.1, simulations of three pumps for an 800-km SSMF system show the evolution of FWM efficiency relative to the frequency gap, which could be interpreted as a subcarrier bandwidth. The results show that if the FWM efficiency drops to a null point, the XPM between adjacent sub-carriers is well reduced, which can be better achieved by reducing the baud rate per subcarrier. At the same time, a lower baud rate reduces the walk-off $\Delta T = 2\pi|\beta_2|BL$ [250] (where B is the frequency spacing between the subcarriers and L is the propagation distance), which in-

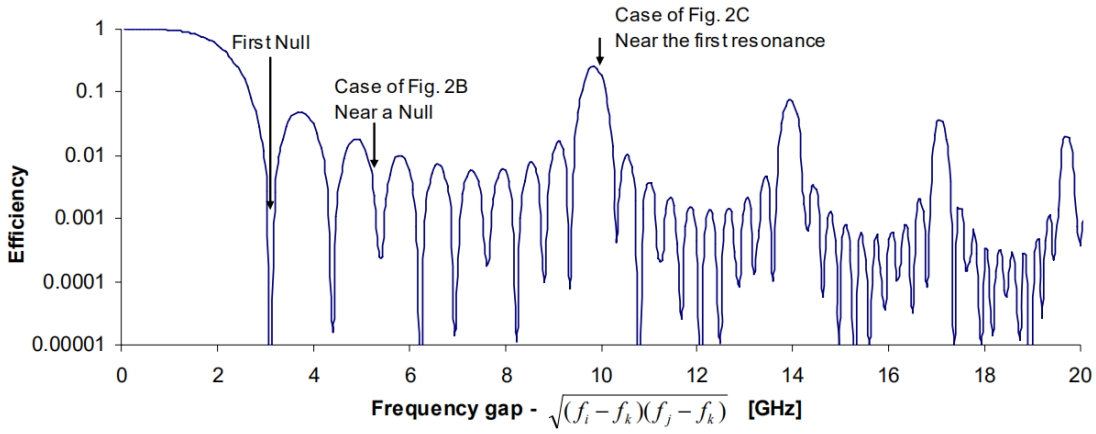


Fig. 3. FWM efficiency versus geometric mean of the frequency separation of three pumps for an 800-km S-SMF system.

Figure A.1: Evolution of FWM efficiency versus frequency gap. Source: [39]

creases the XPM effect. The trade-off between FWM efficiency and walk-off leads to an optimal sub-carrier granularity.

Bibliography

[1] Tobias Fehenberger, Alex Alvarado, Georg Böcherer, and Norbert Hanik. On Probabilistic Shaping of Quadrature Amplitude Modulation for the Nonlinear Fiber Channel. *J. Lightwave Technol.*, 34(21):5063–5073, Nov 2016. Cited pages vii, 40, 88, and 97

[2] Pavel Skvortcov, Ian Phillips, Wladek Forysiak, Toshiaki Koike-Akino, Keisuke Kojima, Kieran Parsons, and David S. Millar. Huffman-Coded Sphere Shaping for Extended-Reach Single-Span Links. *IEEE Journal of Selected Topics in Quantum Electronics*, 27(3):1–15, 2021. Cited pages vii, xxiii, 88, and 96

[3] Stella Civelli, Enrico Forestieri, and Marco Secondini. Sequence-selection-based constellation shaping for nonlinear channels. *Journal of Lightwave Technology*, 2023. Cited pages viii, xxiv, 97, 112, 115, 118, 119, 120, 122, 124, 125, 126, 131, and 138

[4] Yue Zong, Chuan Feng, Yingying Guan, Yejun Liu, and Lei Guo. Virtual network embedding for multi-domain heterogeneous converged optical networks: Issues and challenges. *Sensors*, 20(9):2655, 2020. Cited pages xix and 6

[5] U Cisco. Cisco annual internet report (2018–2023) white paper. *Cisco: San Jose, CA, USA*, 10(1):1–35, 2020. Cited pages xix and 7

[6] Moris. The bandwidth and window of fiber optic cable. https://community.fs.com/article/the-bandwidth-and-window-of-fiber-optic-cable.html. Cited pages xix and 8

[7] Anoop. Types of single mode fibers. https://opticalnetworksforyou.blogspot.com/2016/05/types-of-single-mode-fibers.html. Cited pages xix and 8

[8] Charles Cox, Edward Ackerman, Roger Helkey, and Gary E Betts. Techniques and performance of intensity-modulation direct-detection analog optical links. *IEEE Transactions on Microwave theory and techniques*, 45(8):1375–1383, 1997. Cited pages xix and 9

[9] Kazuro Kikuchi. Fundamentals of coherent optical fiber communications. *Journal of lightwave technology*, 34(1):157–179, 2015. Cited pages xix, 9, 10, 12, and 15

[10] Wikipedia contributors. Root-raised-cosine filter — Wikipedia, the free encyclopedia, 2022. [Online; accessed 25-December-2023]. Cited pages xix and 16

[11] Sina Fazel, Djalal-Falih Bendimerad, Nicola Rossi, Petros Ramantanis, and Yann Frignac. Influence of Modeling Methods on the Estimation of the Nonlinear Noise Statistics Considering Joint PMD and Kerr Effects in Fiber Transmission Systems. *arXiv preprint arXiv:2002.06268*, 2020. Cited pages xix, 20, 48, and 84

[12] mathscinotes. Chromatic dispersion with 10 gigabit optical transports. https://www.mathscinotes.com/2016/09/chromatic-dispersion-with-10-gigabit-optical-transports/. Cited pages xix and 21

[13] Mansoor I. Yousefi. Digital Information Processing Fiber Nonlinearity DCOM1 (L5, L6). Cited pages xix, 25, and 26

[14] Pierluigi Poggiolini, Antonino Nespola, Yanchao Jiang, Gabriella Bosco, Andrea Carena, Luca Bertignono, Syed Muhammad Bilal, Silvio Abrate, and Fabrizio Forghieri. Analytical and Experimental Results on System Maximum Reach Increase Through Symbol Rate Optimization. *J. Lightwave Technol.*, 34(8):1872–1885, Apr 2016. Cited pages xx, 28, and 98

[15] Wikipedia contributors. Phase-shift keying — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Phase-shift_keying&oldid=1185411437`, 2023. [Online; accessed 24-November-2023]. Cited pages xx, 43, and 45

[16] Jiaqi Zhao, Chaoyi Qin, Mengjie Zhang, and Nan Chi. Investigation on performance of special-shaped 8-quadrature amplitude modulation constellations applied in visible light communication. *Photonics Research*, 4(6):249–256, 2016. Cited pages xx and 44

[17] C Thomas, Martin Weidner, and S Durrani. Digital amplitude-phase keying with M-ary alphabets. *IEEE Transactions on Communications*, 22(2):168–180, 1974. Cited pages xx and 44

[18] AD Shiner, M Reimer, A Borowiec, S Oveis Gharan, J Gaudette, P Mehta, D Charlton, K Roberts, and M O'Sullivan. Demonstration of an 8-dimensional modulation format with reduced inter-channel nonlinearities in a polarization multiplexed coherent system. *Optics express*, 22(17):20366–20374, 2014. Cited pages xx, 40, 46, 47, 67, and 75

[19] Djalal F Bendimerad, Huijian Zhang, Ingmar Land, and Hartmut Hafermann. Partially Ordered Statistics Demapping for Multi-Dimensional Modulation Formats. In *Optical Fiber Communication Conference*, pages W3H–5. Optical Society of America, 2019. Cited pages xx, 49, and 50

[20] Ahmed I Abd El-Rahman and John C Cartledge. Evaluating the impact of QAM constellation subset selection on the achievable information rates of multidimensional formats in fully loaded systems. *Journal of Lightwave Technology*, 36(3):712–720, 2017. Cited pages xx, 50, and 51

[21] Z. Qu and I. Djordjevic. On the Probabilistic Shaping and Geometric Shaping in Optical Communication Systems. *IEEE Access*, 7:21454–21464, 2019. Cited pages xx, 51, and 52

[22] Patrick Schulte and Georg Böcherer. Constant composition distribution matching. *IEEE Transactions on Information Theory*, 62(1):430–434, 2015. Cited pages xx, 53, 54, 57, 58, 59, 90, 100, and 111

[23] Laurent Schmalen. Probabilistic constellation shaping: Challenges and opportunities for forward error correction. In *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pages 1–3, 2018. Cited pages xx and 57

[24] Yunus Can Gültekin, Wim J. van Houtum, Arie G. C. Koppelaar, Frans M. J. Willems, and Wim J van Houtum. Enumerative Sphere Shaping for Wireless Communications With Short Packets. *IEEE Transactions on Wireless Communications*, 19(2):1098–1112, 2020. Cited pages xxi, 54, 61, and 62

[25] Tatsuya Nakamura, Emmanuel Le Taillandier de Gabory, Hidemi Noguchi, Wakako Maeda, Jun'ichi Abe, and Kiyoshi Fukuchi. Long haul transmission of four-dimensional 64SP-12QAM signal based on 16QAM constellation for longer distance at same spectral efficiency as PM-8QAM. In *2015 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2015. Cited pages xxi, 67, 68, and 69

[26] Tobias A Eriksson, Martin Sjödin, Pontus Johannisson, Peter A Andrekson, and Magnus Karlsson. Comparison of 128-SP-QAM and PM-16QAM in long-haul WDM transmission. *Optics Express*, 21(16):19269–19279, 2013. Cited pages xxi, 69, 70, and 75

[27] Henning Buelow, Xiaofeng Lu, Laurent Schmalen, Axel Klekamp, and Fred Buchali. Experimental performance of 4D optimized constellation alternatives for PM-8QAM and PM-16QAM. In *Optical Fiber Communication Conference*, pages M2A–6. Optica Publishing Group, 2014. Cited pages xxi, 67, and 70

[28] Keisuke Kojima, Tsuyoshi Yoshida, Toshiaki Koike-Akino, David S Millar, Kieran Parsons, Milutin Pajovic, and Valeria Arlunno. Nonlinearity-tolerant four-dimensional 2A8PSK family for 5–7 bits/symbol spectral efficiency. *Journal of Lightwave Technology*, 35(8):1383–1391, 2017. Cited pages xxi, 68, 71, 75, and 81

[29] Bin Chen, Chigo Okonkwo, Hartmut Hafermann, and Alex Alvarado. Eight-dimensional polarization-ring-switching modulation formats. *IEEE Photonics Technology Letters*, 31(21):1717–1720, 2019. Cited pages xxi, 72, 73, 75, and 78

[30] Marcin Pikus and Wen Xu. Applying bit-level probabilistically shaped coded modulation for high-throughput communications. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–6, 2017. Cited pages xxii, 91, and 92

[31] Mengfan Fu, Qiaoya Liu, Xiaobo Zeng, Yiwen Wu, Lilin Yi, Weisheng Hu, and Qunbi Zhuge. Multi-Dimensional Distribution Matching for Probabilistically Shaped High Order Modulation Format. In *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2020. Cited pages xxii and 92

[32] Mrinmoy Jana, Lutz Lampe, Jeebak Mitra, and Chuandong Li. Multi-dimensional Probabilistic Shaping for Optical Superchannels. In *2021 European Conference on Optical Communication (ECOC)*, pages 1–4, 2021. Cited pages xxii and 92

[33] Tobias Fehenberger and Alex Alvarado. Analysis and Optimisation of Distribution Matching for the Nonlinear Fibre Channel. *arXiv preprint arXiv:1907.02846*, 2019. Cited pages xxiii and 93

[34] Yunus Can Gültekin, Alex Alvarado, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo M. Okonkwo, and Frans M. J. Willems. Kurtosis-Limited Sphere Shaping for Nonlinear Interference Noise Reduction in Optical Channels. *Journal of Lightwave Technology*, 40(1):101–112, 2022. Cited pages xxiii, 93, and 94

[35] Yunus Can Gültekin, Alex Alvarado, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo M. Okonkwo, and Frans M. J. Willems. Mitigating Nonlinear Interference by Limiting Energy Variations in Sphere Shaping. In *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2022. Cited pages xxiii, 94, 95, and 114

[36] Stella Civelli, Enrico Forestieri, Alexey Lotsmanov, Dmitry Razdoburdin, and Marco Secondini. A sequence selection bound for the capacity of the nonlinear fiber channel. In *2021 European Conference on Optical Communication (ECOC)*, pages 1–4. IEEE, 2021. Cited pages xxiii, 112, 114, 117, and 126

[37] Kaiquan Wu, Gabriele Liga, Alireza Sheikh, Yunus Can Gültekin, Frans M. J. Willems, and Alex Alvarado. List-Encoding CCDM: A Nonlinearity-Tolerant Shaper Aided by Energy Dispersion Index. *Journal of Lightwave Technology*, 40(4):1064–1071, 2022. Cited pages xxiii, 115, and 116

[38] Mohammad Taha Askari, Lutz Lampe, and Jeebak Mitra. Probabilistic amplitude shaping and nonlinearity tolerance: Analysis and sequence selection method. *Journal of Lightwave Technology*, 2023. Cited pages xxiv, 112, 115, 116, 117, 122, and 126

[39] Liang B Du and Arthur J Lowery. Improved nonlinearity precompensation for long-haul high-data-rate transmission using coherent optical OFDM. *Optics express*, 16(24):19920–19925, 2008. Cited pages xxvi and 142

[40] K Charles Kao and George A Hockham. Dielectric-fibre surface waveguides for optical frequencies. In *Proceedings of the Institution of Electrical Engineers*, number 7, pages 1151–1158. IET, 1966. Cited page 7

[41] Rene Bonk, Ed Harstead, Robert Borkowski, Vincent Houtsma, Yannick Lefevre, Amitkumar Mahadevan, Dora van Veen, Michiel Verplaetse, and Sheldon Walklin. Perspectives on and the road towards 100 Gb/s TDM PON with intensity-modulation and direct-detection. *Journal of Optical Communications and Networking*, 15(8):518–526, 2023. Cited page 9

[42] Ethan M Liang and Joseph M Kahn. Geometric shaping for distortion-limited intensity modulation/direct detection data center links. *IEEE Photonics Journal*, 2023. Cited pages

[43] Ji Zhou, Haide Wang, Yuanhua Feng, Weiping Liu, Shecheng Gao, Changyuan Yu, and Zhaohui Li. Processing for dispersive intensity-modulation and direct-detection fiber-optic communications. *Optics Letters*, 46(1):138–141, 2021. Cited page 9

[44] Owen E DeLange. Wide-band optical communication systems: Part II—Frequency-division multiplexing. *Proceedings of the IEEE*, 58(10):1683–1690, 1970. Cited page 9

[45] Tadashi Okoshi and Kazuro Kikuchi. Frequency stabilisation of semiconductor lasers for heterodyne-type optical communication systems. *Electronics letters*, 5(16):179–181, 1980. Cited page 9

[46] F Favre and D Le Guen. High frequency stability of laser diode for heterodyne communication systems. *Electronics Letters*, 18(16):709–710, 1980. Cited page 9

[47] Kim Roberts, Sik Heng Foo, Michael Moyer, Michael Hubbard, Andrew Sinclair, Jamie Gaudette, and Charles Laperle. High capacity transport—100G and beyond. *Journal of Lightwave Technology*, 33(3):563–578, 2014. Cited page 11

[48] Kim Roberts, Maurice O'Sullivan, Kuang-Tsan Wu, Han Sun, Ahmed Awadalla, David J Krause, and Charles Laperle. Performance of dual-polarization QPSK for optical transport systems. *Journal of lightwave technology*, 27(16):3546–3559, 2009. Cited page 42

[49] Han Sun, Mehdi Torbatian, Mehdi Karimi, Robert Maher, Sandy Thomson, Mohsen Tehrani, Yuliang Gao, Ales Kumpera, George Soliman, Aditya Kakkar, et al. 800G DSP ASIC design using probabilistic shaping and digital sub-carrier multiplexing. *Journal of lightwave technology*, 38(17):4744–4756, 2020. Cited pages

[50] Stefano Porto, Mohammad Chitgarha, Irene Leung, Robert Maher, Ryan Going, Stefan Wolf, Pavel Studenkov, Jiaming Zhang, Hossein Hodaei, Thomas Frost, et al. Demonstration of a 2× 800 Gb/s/wave coherent optical engine based on an InP monolithic PIC. *Journal of Lightwave Technology*, 40(3):664–671, 2022. Cited page 11

[51] Satoshi Tsukamoto, Yuta Ishikawa, and Kazuro Kikuchi. Optical homodyne receiver comprising phase and polarization diversities with digital signal processing. In *2006 European Conference on Optical Communications*, pages 1–2. IEEE, 2009. Cited page 11

[52] Alan Pak Tao Lau, Yuliang Gao, Qi Sui, Dawei Wang, Qunbi Zhuge, Mohamed H Morsy-Osman, Mathieu Chagnon, Xian Xu, Chao Lu, and David V Plant. Advanced DSP techniques enabling high spectral efficiency and flexible transmissions: Toward elastic optical networks. *IEEE Signal Processing Magazine*, 31(2):82–92, 2014. Cited page 12

[53] Seb J Savory. Digital coherent optical receivers: Algorithms and subsystems. *IEEE Journal of selected topics in quantum electronics*, 16(5):1164–1179, 2010. Cited page 12

[54] P Serena, M Bertolini, and A Vannucci. Optilux toolbox. *Departament of information engineering, University of Palma, Italy*, 2009. Cited page 13

[55] Fred S. Annexstein. Generating de Bruijn sequences: An efficient implementation. *IEEE Transactions on Computers*, 46(2):198–200, 1997. Cited page 13

[56] Hlaing Minn, Vijay K Bhargava, and Khaled Ben Letaief. A robust timing and frequency synchronization for OFDM systems. *IEEE Transactions on Wireless communications*, 2(4):822–839, 2003. Cited page 13

[57] Fabio Pittalà, Amine Mezghani, Fabian N Hauske, Yabin Ye, Idelfonso T Monroy, and Josef A Nossek. Efficient training-based channel estimation for coherent optical communication systems. In *Signal Processing in Photonic Communications*, pages SpTu3A–4. Optical Society of America, 2012. Cited pages 15 and 30

[58] Fabio Pittala, Israa Slim, Amine Mezghani, and Josef A Nossek. Training-aided frequency-domain channel estimation and equalization for single-carrier coherent optical transmission systems. *Journal of Lightwave Technology*, 32(24):4849–4863, 2014. Cited pages 15 and 30

[59] Ian Coddington, William C Swann, Luca Lorini, James C Bergquist, Y Le Coq, Christopher W Oates, Q Quraishi, KS Feder, JW Nicholson, Paul S Westbrook, et al. Coherent optical link over hundreds of metres and hundreds of terahertz with subfemtosecond timing jitter. *Nature Photonics*, 1(5):283–287, 2007. Cited page 17

[60] Arthur L Schawlow and Charles H Townes. Infrared and optical masers. *Physical review*, 112(6):1940, 1958. Cited page 17

[61] Charles Henry. Theory of the linewidth of semiconductor lasers. *IEEE Journal of Quantum Electronics*, 18(2):259–264, 1982. Cited page 17

[62] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019. Cited page 18

[63] Philippe Gallion and Guy Debarge. Quantum phase noise and field correlation in single frequency semiconductor laser systems. *IEEE Journal of Quantum Electronics*, 20(4):343–349, 1984. Cited page 18

[64] R Tkach and A Chraplyvy. Phase noise and linewidth in an InGaAsP DFB laser. *Journal of Lightwave Technology*, 4(11):1711–1716, 1986. Cited page 18

[65] William Shieh and Keang-Po Ho. Equalization-enhanced phase noise for coherent-detection systems using electronic digital signal processing. *Optics Express*, 16(20):15718–15727, 2008. Cited pages 18 and 141

[66] Aymeric Arnould and Amirhossein Ghazisaeidi. Equalization enhanced phase noise in coherent receivers: DSP-aware analysis and shaped constellations. *Journal of Lightwave Technology*, 37(20):5282–5290, 2019. Cited pages 18 and 141

[67] Oleg V Sinkin, Ronald Holzlöhner, John Zweck, and Curtis R Menyuk. Optimization of the split-step Fourier method in modeling optical-fiber communications systems. *Journal of lightwave technology*, 21(1):61, 2003. Cited page 19

[68] Q Zhang and MI Hayee. An SSF scheme to achieve comparable global simulation accuracy in WDM systems. *IEEE photonics technology letters*, 17(9):1869–1871, 2005. Cited pages

[69] Qun Zhang and M Imran Hayee. Symmetrized split-step Fourier scheme to control global simulation accuracy in fiber-optic communication systems. *Journal of Lightwave Technology*, 26(2):302–316, 2008. Cited pages

[70] Simone Musetti, Paolo Serena, and Alberto Bononi. On the accuracy of split-step Fourier simulations for wide-band nonlinear optical communications. *Journal of Lightwave Technology*, 36(23):5669–5677, 2018. Cited pages

[71] Jing Shao, Xiaojun Liang, and Shiva Kumar. Comparison of split-step Fourier schemes for simulating fiber optic communication systems. *IEEE Photonics Journal*, 6(4):1–15, 2014. Cited pages

[72] AO Dal Forno, A Paradisi, R Passy, and JP Von der Weid. Experimental and theoretical modeling of polarization-mode dispersion in single-mode fibers. *IEEE Photonics Technology Letters*, 12(3):296–298, 2000. Cited pages

[73] Dietrich Marcuse, CR Manyuk, and Ping Kong Alexander Wai. Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence. *Journal of Lightwave Technology*, 15(9):1735–1746, 1997. Cited page 20

[74] Cristian Antonelli, Mark Shtaif, and Antonio Mecozzi. Modeling of nonlinear propagation in space-division multiplexed fiber-optic transmission. *Journal of Lightwave Technology*, 34(1):36–54, 2015. Cited pages

[75] Keang-Po Ho, Joseph M Kahn, IP Kaminow, T Li, and AE Willner. Mode coupling and its impact on spatially multiplexed systems. *Optical Fiber Telecommunications VI*, 17:1386–1392, 2013. Cited pages

[76] Peter J Acklam. MATLAB array manipulation tips and tricks. *Notes*, 2003. Cited page 19

[77] Govind P Agrawal. Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century*, pages 195–211. Springer, 2000. Cited pages 19 and 124

[78] Ping Kong Alexander Wai and CR Menyak. Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence. *Journal of Lightwave Technology*, 14(2):148–157, 1996. Cited page 20

[79] Andrea Galtarossa, Luca Palmieri, Marco Schiano, and Tiziana Tambosso. Measurement of birefringence correlation length in long, single-mode fibers. *Optics Letters*, 26(13):962–964, 2001. Cited page 20

[80] Curtis R Menyuk and Brian S Marks. Interaction of polarization mode dispersion and nonlinearity in optical fiber transmission systems. *Journal of Lightwave Technology*, 24(7):2806, 2006. Cited page 20

[81] Ping Kong Alexander Wai, WL Kath, CR Menyuk, and JW Zhang. Nonlinear polarization-mode dispersion in optical fibers with randomly varying birefringence. *JOSA B*, 14(11):2967–2979, 1997. Cited pages

[82] Ping Kong Alexander Wai, WL Kath, CR Menyuk, and J Zhang. Analysis of nonlinear polarization-mode dispersion in optical fibers with randomly varying birefringence. In *Proceedings of Optical Fiber Communication Conference (*, pages 257–258. IEEE, 1997. Cited page 20

[83] Liang Chen and X. Bao. Polarization-dependent loss-induced pulse narrowing in birefringent optical fiber with finite differential group delay. *Journal of Lightwave Technology*, 18:665–667, 2000. Cited page 21

[84] Cristian Francia, Frank Bruyere, Denis Penninckx, and Michel Chbat. PMD second-order effects on pulse propagation in single-mode optical fibers. *Ieee Photonics Technology Letters*, 10(12):1739–1741, 1998. Cited page 22

[85] SJ Ben Yoo. Advanced optical components for next-generation photonic networks. *Active and Passive Optical Components for WDM Communications III*, 5246:224–234, 2003. Cited page 22

[86] Govind P Agrawal. *Fiber-optic communication systems*. John Wiley & Sons, 2012. Cited page 22

[87] Thomas Duthel, CRS Fludger, Jonas Geyer, and Christoph Schulien. Impact of polarisation dependent loss on coherent POLMUX-NRZ-DQPSK. In *OFC/NFOEC 2008-2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference*, pages 1–3. IEEE, 2008. Cited page 22

[88] C Randy Giles and Emmanuel Desurvire. Propagation of signal and noise in concatenated erbium-doped fiber optical amplifiers. *Journal of lightwave technology*, 9(2):147–154, 1991. Cited page 23

[89] Amnon Yariv. Signal-to-noise considerations in fiber links with periodic or distributed optical amplification. *Optics letters*, 15(19):1064–1066, 1990. Cited page 24

[90] René-Jean Essiambre, Gerhard Kramer, Peter J Winzer, Gerard J Foschini, and Bernhard Goebel. Capacity limits of optical fiber networks. *Journal of Lightwave Technology*, 28(4):662–701, 2010. Cited pages 25, 33, and 114

[91] James P Gordon and Linn F Mollenauer. Phase noise in photonic communications systems using linear amplifiers. *Optics letters*, 15(23):1351–1353, 1990. Cited page 26

[92] Keang-Po Ho and Joseph M Kahn. Electronic compensation technique to mitigate nonlinear phase noise. *Journal of Lightwave Technology*, 22(3):779, 2004. Cited page 26

[93] Govind P Agrawal, PL Baldeck, and RR Alfano. Modulation instability induced by cross-phase modulation in optical fibers. *Physical Review A*, 39(7):3406, 1989. Cited page 26

[94] Ronen Dar and Peter J Winzer. Nonlinear interference mitigation: Methods and potential gain. *Journal of Lightwave Technology*, 35(4):903–930, 2017. Cited page 27

[95] Pierluigi Poggiolini, Andrea Carena, Vittorio Curri, Gabriella Bosco, and Fabrizio Forghieri. Analytical modeling of nonlinear propagation in uncompensated optical transmission links. *IEEE Photonics technology letters*, 23(11):742–744, 2011. Cited page 27

[96] Pierluigi Poggiolini, G Bosco, A Carena, V Curri, Y Jiang, and F Forghieri. The gn-model of fiber non-linear propagation and its applications. *Journal of lightwave technology*, 32(4):694–721, 2013. Cited pages 27 and 132

[97] Andrea Carena, Vittorio Curri, Gabriella Bosco, Pierluigi Poggiolini, and F Forghieri. Modeling of the impact of nonlinear propagation effects in uncompensated optical coherent transmission links. *Journal of Lightwave technology*, 30(10):1524–1539, 2012. Cited page 27

[98] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation. *Journal of Lightwave Technology*, 33(5):1044–1053, 2014. Cited page 27

[99] Andrea Carena, Gabriella Bosco, Vittorio Curri, Yanchao Jiang, Pierluigi Poggiolini, and Fabrizio Forghieri. On the accuracy of the GN-model and on analytical correction terms to improve it. *arXiv preprint arXiv:1401.6946*, 2014. Cited pages

[100] Pierluigi Poggiolini, Gabriella Bosco, Andrea Carena, Vittorio Curri, Yanchao Jiang, and Fabrizio Forghieri. A simple and effective closed-form gn model correction formula accounting for signal non-gaussian distribution. *Journal of Lightwave Technology*, 33(2):459–473, 2015. Cited page 27

[101] Zhiwei Liang, Bin Chen, Yi Lei, Gabriele Liga, and Alex Alvarado. Analytical Model of Nonlinear Fiber Propagation for General Dual-Polarization Four-Dimensional Modulation Format. *arXiv preprint arXiv:2302.06831*, 2023. Cited page 27

[102] Han Sun, Mehdi Torbatian, Mehdi Karimi, Robert Maher, Sandy Thomson, Mohsen Tehrani, Yuliang Gao, Ales Kumpera, George Soliman, Aditya Kakkar, Mohammad Osman, Ziad A. El-Sahn, Clayton Doggart, Weikun Hou, Shailesh Sutarwala, Yuejian Wu, Mohammad Reza Chitgarha, Vikrant Lal, Huan-Shang Tsai, Scott Corzine, Jiaming Zhang, John Osenbach, Sanketh Buggaveeti, Zulfikar Morbi, Miguel Iglesias Olmedo, Irene Leung, Xian Xu, Parmijit Samra, Vince Dominic, Steve Sanders, Mehrdad Ziari, Antonio Napoli, Bernhard Spinnler, Kuang-Tsan Wu, and Parthiban Kandappan. 800G DSP ASIC Design Using Probabilistic Shaping and Digital Sub-Carrier Multiplexing. *Journal of Lightwave Technology*, 38(17):4744–4756, 2020. Cited pages 28 and 141

[103] Mansoor I Yousefi and Frank R Kschischang. Information transmission using the nonlinear Fourier transform, Part I: Mathematical tools. *IEEE Transactions on Information Theory*, 60(7):4312–4328, 2014. Cited page 29

[104] Jaroslaw E Prilepsky, Stanislav A Derevyanko, Keith J Blow, Ildar Gabitov, and Sergei K Turitsyn. Nonlinear inverse synthesis and eigenvalue division multiplexing in optical fiber channels. *Physical review letters*, 113(1):013901, 2014. Cited page 29

[105] Ezra Ip and Joseph M Kahn. Compensation of dispersion and nonlinear impairments using digital backpropagation. *Journal of Lightwave Technology*, 26(20):3416–3425, 2008. Cited page 29

[106] Zhenning Tao, Liang Dou, Weizhen Yan, Yangyang Fan, Lei Li, Shoichiro Oda, Yuichi Akiyama, Hisao Nakashima, Takeshi Hoshida, and Jens C Rasmussen. Complexity-reduced digital nonlinear compensation for coherent optical system. In *Next-Generation Optical Communication: Components, Sub-Systems, and Systems II*, volume 8647, pages 115–125. SPIE, 2013. Cited page 29

[107] Kaiquan Wu, Gabriele Liga, Alireza Sheikh, Frans MJ Willems, and Alex Alvarado. Temporal energy analysis of symbol sequences for fiber nonlinear interference modelling via energy dispersion index. *Journal of Lightwave Technology*, 39(18):5766–5782, 2021. Cited pages 29, 96, 106, 112, 114, 122, and 123

[108] Junho Cho, Xi Chen, Greg Raybon, Di Che, Ellsworth Burrows, Samuel Olsson, and Robert Tkach. Shaping lightwaves in time and frequency for optical fiber communication. *Nature communications*, 13(1):785, 2022. Cited pages 29, 106, and 122

[109] Mehrez Selmi, Yves Jaouen, and Philippe Ciblat. Accurate digital frequency offset estimator for coherent PolMux QAM transmission systems. In *2009 35th European Conference on Optical Communication*, pages 1–2. IEEE, 2009. Cited page 29

[110] M Selmi, C Gosset, M Noelle, P Ciblat, and Y Jaouen. Block-wise digital signal processing for PolMux QAM/PSK optical coherent systems. *Journal of lightwave technology*, 29(20):3070–3082, 2011. Cited page 29

[111] D Godard. Passband timing recovery in an all-digital modem receiver. *IEEE Transactions on Communications*, 26(5):517–523, 1978. Cited page 29

[112] Floyd Gardner. A BPSK/QPSK timing-error detector for sampled receivers. *IEEE Transactions on communications*, 34(5):423–429, 1986. Cited page 29

[113] Kurth Mueller and Markus Muller. Timing recovery in digital synchronous data receivers. *IEEE transactions on communications*, 24(5):516–531, 1976. Cited page 29

[114] Dominique Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE transactions on communications*, 28(11):1867–1875, 1980. Cited page 30

[115] Yuliang Gao, Alan Pak Tao Lau, Shuangyi Yan, and Chao Lu. Low-complexity and phase noise tolerant carrier phase estimation for dual-polarization 16-QAM systems. *Optics Express*, 19(22):21717–21729, 2011. Cited page 30

[116] Andrew J Viterbi and Audrey M Viterbi. Nonlinear estimation of PSK-modulated carrier phase with application to burst digital transmission. *IEEE Transactions on Information theory*, 29(4):543–551, 1983. Cited page 30

[117] Timo Pfau, Sebastian Hoffmann, and Reinhold Noé. Hardware-efficient coherent digital receiver concept with feedforward carrier recovery for M-QAM constellations. *Journal of Lightwave Technology*, 27(8):989–999, 2009. Cited pages 30, 74, and 81

[118] Yan Li, Quanyan Ning, Lei Yue, Honghang Zhou, Chao Gao, Yuyang Liu, Jifang Qiu, Wei Li, Xiaobin Hong, and Jian Wu. Post-FEC Performance of Pilot-Aided Carrier Phase Estimation over Cycle Slip. *Applied Sciences*, 9(13):2749, 2019. Cited page 30

[119] Arni F Alfredsson, Erik Agrell, Henk Wymeersch, Benjamin J Puttnam, Georg Rademacher, Ruben S Luís, and Magnus Karlsson. Pilot-aided joint-channel carrier-phase estimation in space-division multiplexed multicore fiber transmission. *Journal of Lightwave Technology*, 37(4):1133–1142, 2018. Cited page 30

[120] Tobias Fehenberger et al. Compensation of XPM interference by blind tracking of the nonlinear phase in WDM systems with QAM input. In *2015 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2015. Cited page 30

[121] Alex Alvarado, Tobias Fehenberger, Bin Chen, and Frans MJ Willems. Achievable information rates for fiber optics: Applications and computations. *Journal of Lightwave Technology*, 36(2):424–439, 2018. Cited pages 31 and 33

[122] Telecommunication Standardization Sector ITU. Forward error correction for submarine systems. *Tech. Recommendation G*, 975:G709, 2000. Cited page 32

[123] Svitlana Matsenko, Sandis Spolitis, and Vjaceslavs Bobrovs. Integration of FEC Channel-coding Schemes Based on the Bose-Chaudhuri-Hocquenghem (BCH) Code for WDM Fiber Optical Communication Systems. In *2019 Photonics & Electromagnetics Research Symposium-Fall (PIERS-Fall)*, pages 2960–2964. IEEE, 2019. Cited page 32

[124] Laurent Schmalen, Vahid Aref, Junho Cho, Detlef Suikat, Detlef Rösener, and Andreas Leven. Spatially coupled soft-decision error correction for future lightwave systems. *Journal of Lightwave Technology*, 33(5):1109–1116, 2014. Cited page 32

[125] Andreas Leven and Laurent Schmalen. Status and recent advances on forward error correction technologies for lightwave systems. *Journal of Lightwave Technology*, 32(16):2735–2750, 2014. Cited page 32

[126] ETSI ETSI. Digital video broadcasting (dvb); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications. *Part II: S2-Extensions (DVB-S2X)*, pages 22–27, 2005. Cited pages 32 and 85

[127] Junho Cho, Chongjin Xie, and Peter J Winzer. Analysis of soft-decision FEC on non-AWGN channels. *Optics Express*, 20(7):7915–7928, 2012. Cited page 32

[128] Laurent Schmalen, Alex Alvarado, and Rafael Rios-Müller. Performance prediction of nonbinary forward error correction in optical transmission experiments. *Journal of Lightwave Technology*, 35(4):1015–1027, 2016. Cited page 32

[129] Alex Alvarado, Erik Agrell, Domaniç Lavery, Robert Maher, and Polina Bayvel. Replacing the soft-decision fec limit paradigm in the design of optical communication systems. *Journal of Lightwave Technology*, 33(20):4338–4352, 2015. Cited page 32

[130] Fred Buchali, Fabian Steiner, Georg Böcherer, Laurent Schmalen, Patrick Schulte, and Wilfried Idler. Rate adaptation and reach increase by probabilistically shaped 64-qam: An experimental demonstration. *Journal of lightwave technology*, 34(7):1599–1609, 2015. Cited pages 32 and 40

[131] Sethumadhavan Chandrasekhar, Borui Li, Junho Cho, Xi Chen, Ellsworth Burrows, Gregory Raybon, and Peter Winzer. High-spectral-efficiency transmission of PDM 256-QAM with parallel probabilistic shaping at record rate-reach trade-offs. In *ECOC 2016-Post Deadline Paper; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016. Cited pages

[132] Junho Cho, Xi Chen, S Chandrasekhar, Gregory Raybon, Ronen Dar, Laurent Schmalen, E Burrows, Andrew Adamiecki, Stephen Corteselli, Yin Pan, et al. Trans-Atlantic field trial using probabilistically shaped 64-QAM at high spectral efficiencies and single-carrier real-time 250-Gb/s 16-QAM. In *Optical Fiber Communication Conference*, pages Th5B–3. Optica Publishing Group, 2017. Cited page 32

[133] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. Cited page 33

[134] Georg Böcherer. Achievable rates for probabilistic shaping. *arXiv preprint arXiv:1707.01134*, 2017. Cited page 35

[135] Junho Cho, Laurent Schmalen, and Peter J Winzer. Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited page 35

[136] Tobias Fehenberger, David S. Millar, Toshiaki Koike-Akino, Keisuke Kojima, and Kieran Parsons. Multiset-Partition Distribution Matching. *IEEE Transactions on Communications*, 67(3):1885–1893, 2019. Cited page 36

[137] Anxu Zhang, Junjie Li, Lipeng Feng, Kai Lv, Fei Yan, Yusen Yang, Haiqiang Wang, Qifang Yang, Lingquan Wang, Xiaolei Zhang, et al. Field trial of 24-Tb/s (60× 400Gb/s) DWDM transmission over a 1910-km G. 654. E fiber link with 6-THz-bandwidth C-band EDFAs. *Optics Express*, 29(26):43811–43818, 2021. Cited page 39

[138] Ihtesham Khan, Lorenzo Tunesi, Muhammad Umar Masood, Enrico Ghillino, Paolo Bardella, Andrea Carena, and Vittorio Curri. Optimized management of ultra-wideband photonics switching systems assisted by machine learning. *Optics Express*, 30(3):3989–4004, 2022. Cited page 39

[139] Erik Agrell and Magnus Karlsson. Power-efficient modulation formats in coherent transmission systems. *Journal of Lightwave Technology*, 27(22):5115–5126, 2009. Cited pages 39 and 70

[140] Djalal F Bendimerad, Hartmut Hafermann, and Huijian Zhang. Nonlinearity-tolerant 8D modulation formats by set-partitioning PDM-QPSK. In *2018 Optical Fiber Communications Conference and Exposition (OFC)*, pages 1–3. IEEE, 2018. Cited pages 40, 49, 67, and 75

[141] Magnus Karlsson and Erik Agrell. Which is the most power-efficient modulation format in optical links? *Optics express*, 17(13):10814–10819, 2009. Cited pages 40, 50, 67, and 75

[142] Zhen Qu and Ivan B Djordjevic. Geometrically shaped 16QAM outperforming probabilistically shaped 16QAM. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited page 40

[143] Shaoliang Zhang, Fatih Yaman, Eduardo Mateo, Takanori Inoue, Kohei Nakamura, and Yoshihisa Inada. Design and performance evaluation of a GMI-optimized 32QAM. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited page 74

[144] Bin Chen, Chigo Okonkwo, Hartmut Hafermann, and Alex Alvarado. Increasing achievable information rates via geometric shaping. In *2018 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2018. Cited page 74

[145] Bin Chen, Chigo Okonkwo, Domaniç Lavery, and Alex Alvarado. Geometrically-shaped 64-point constellations via achievable information rates. In *2018 20th International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2018. Cited pages 40 and 74

[146] Georg Böcherer, Fabian Steiner, and Patrick Schulte. Fast probabilistic shaping implementation for long-haul fiber-optic communication systems. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited page 40

[147] Fred Buchali, Wilfried Idler, Roman Dischler, Tobias Eriksson, and Laurent Schmalen. Spectrally efficient probabilistically shaped square 64QAM to 256 QAM. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited pages

[148] Robert Maher, Kevin Croussore, Matthias Lauermann, Ryan Going, Xian Xu, and Jeff Rahn. Constellation shaped 66 GBd DP-1024QAM transceiver with 400 km transmission over standard SMF. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited pages

[149] Abdelkerim Amari, Sebastiaan Goossens, Yunus Can Gültekin, Olga Vassilieva, Inwoong Kim, Tadashi Ikeuchi, Chigo M Okonkwo, Frans MJ Willems, and Alex Alvarado. Introducing enumerative sphere shaping for optical communication systems with short blocklengths. *Journal of Lightwave Technology*, 37(23):5926–5936, 2019. Cited pages 40, 54, 61, 90, 111, and 112

[150] Leszek Szczecinski and Alex Alvarado. *Bit-interleaved coded modulation: fundamentals, analysis and design*. John Wiley & Sons, 2015. Cited pages 40 and 55

[151] Fabian Steiner and Georg Böcherer. Comparison of geometric and probabilistic shaping with application to ATSC 3.0. In *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*, pages 1–6. VDE, 2017. Cited page 40

[152] Hussam G Batshon, Matthew V Mazurczyk, J-X Cai, Oleg V Sinkin, Milen Paskov, Carl R Davidson, Ding Wang, M Bolshtyansky, and D Foursa. Coded modulation based on 56apsk with hybrid shaping for high spectral efficiency transmission. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited pages 41 and 91

[153] J-X Cai, HG Batshon, Matthew V Mazurczyk, Oleg V Sinkin, D Wang, Milen Paskov, W Patterson, Carl R Davidson, P Corbett, G Wolter, 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*, pages Th5B–2. Optica Publishing Group, 2017. Cited pages 41 and 91

[154] Han Sun, Kuang-Tsan Wu, and Kim Roberts. Real-time measurements of a 40 Gb/s coherent system. *Optics express*, 16(2):873–879, 2008. Cited page 42

[155] Markus Nölle, Felix Frey, Robert Elschner, Carsten Schmidt-Langhorst, Antonio Napoli, and Colja Schubert. Performance comparison of different 8QAM constellations for the use in flexible optical networks. In *Optical Fiber Communication Conference*, pages W3B–2. Optical Society of America, 2014. Cited pages 44 and 75

[156] Xiang Zhou, Lynn Nelson, Peter Magill, R Isaac, B Zhu, DW Peckham, P Borel, and K Carlson. 8× 450-Gb/s, 50-GHz-Spaced, PDM-32QAM transmission over 400km and one 50GHz-grid ROADM. In *Optical Fiber Communication Conference*, page PDPB3. Optica Publishing Group, 2011. Cited page 45

[157] Wei-Ren Peng, Takehiro Tsuritani, and Itsuro Morita. Transmission of high-baud PDM-64QAM signals. *Journal of lightwave technology*, 31(13):2146–2162, 2013. Cited page 45

[158] Xi Chen, Ken Tanizawa, Peter Winzer, Po Dong, Junho Cho, Fumio Futami, Kentaro Kato, Argishti Melikyan, and KW Kim. Experimental demonstration of a 4,294,967,296-QAM-based Y-00 quantum stream cipher template carrying 160-Gb/s 16-QAM signals. *Optics Express*, 29(4):5658–5664, 2021. Cited page 45

[159] V. Neskorniuk, A. Carnio, D. Marsella, S. Turitsyn, J. Prilepsky, and V. Aref. Model-Based Deep Learning of Joint Probabilistic and Geometric Shaping for Optical Communication. *2022 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2, 2022. Cited page 52

[160] G David Forney and L-F Wei. Multidimensional constellations. I. Introduction, figures of merit, and generalized cross constellations. *IEEE journal on selected areas in communications*, 7(6):877–892, 1989. Cited page 52

[161] Hussam G Batshon, Ivan B Djordjevic, Lei Xu, and Ting Wang. Iterative polar quantization-based modulation to achieve channel capacity in ultrahigh-speed optical communication systems. *IEEE Photonics Journal*, 2(4):593–599, 2010. Cited page 52

[162] Shaoliang Zhang, Fatih Yaman, Eduardo Mateo, Takanori Inoue, Kohei Nakamura, and Yoshihisa Inada. A generalized pairwise optimization for designing multi-dimensional modulation formats. In *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3. IEEE, 2017. Cited page 53

[163] RJ Essiambre, J Karout, E Agrell, and A Tulino. Achievable rates of multidimensional multisphere distributions. In *Optical Fiber Communication Conference (OFC)*, 2017. Cited page 53

[164] Frank R Kschischang et al. Optimal nonuniform signaling for Gaussian channels. *IEEE Transactions on Information Theory*, 39(3):913–929, 1993. Cited pages 53, 55, and 90

[165] Georg Böcherer and Rudolf Mathar. Matching dyadic distributions to channels. In *2011 Data Compression Conference*, pages 23–32. IEEE, 2011. Cited page 53

[166] Rana Ali Amjad and Georg Böcherer. Fixed-to-variable length distribution matching. In *2013 IEEE International Symposium on Information Theory*, pages 1511–1515. IEEE, 2013. Cited page 53

[167] Sebastian Baur and Georg Böcherer. Arithmetic distribution matching. In *SCC 2015; 10th International ITG Conference on Systems, Communications and Coding*, pages 1–6. VDE, 2015. Cited page 53

[168] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. On shaping gain in the nonlinear fiber-optic channel. In *2014 IEEE International Symposium on Information Theory*, pages 2794–2798. IEEE, 2014. Cited page 54

[169] Omri Geller, Ronen Dar, Meir Feder, and Mark Shtaif. A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems. *Journal of Lightwave Technology*, 34(16):3884–3889, 2016. Cited pages

[170] Rajiv Laroia, Nariman Farvardin, and Steven A Tretter. On optimal shaping of multidimensional constellations. *IEEE Transactions on Information Theory*, 40(4):1044–1056, 1994. Cited pages 54 and 60

[171] Georg Böcherer, Fabian Steiner, and Patrick Schulte. Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation. *IEEE Transactions on Communications*, 63(12):4651–4665, 2015. Cited pages 55, 57, and 58

[172] Georg Böcherer. *Principles of coded modulation*. PhD thesis, Technische Universität München, 2018. Cited page 55

[173] F Willems and J Wuijts. A pragmatic approach to shaped coded modulation. In *IEEE 1st Symp. on Commun. and Veh. Technol. in the Benelux*, 1993. Cited page 60

[174] Y Can Gültekin, WJ Van Houtum, and FMJ Willems. On constellation shaping for short block lengths. In *Proc. Symp. Inf. Theory Signal Process. Benelux*, pages 1–11, 2018. Cited page 60

[175] Patrick Schulte and Fabian Steiner. Divergence-optimal fixed-to-fixed length distribution matching with shell mapping. *IEEE Wireless Communications Letters*, 8(2):620–623, 2019. Cited page 61

[176] Stella Civelli, Enrico Forestieri, and Marco Secondini. Interplay of Probabilistic Shaping and Carrier Phase Recovery for Nonlinearity Mitigation. In *2020 European Conference on Optical Communications (ECOC)*, pages 1–4, 2020. Cited pages 62 and 98

[177] Yunus Can Gültekin, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo Okonkwo, and Alex Alvarado. On Optimum Enumerative Sphere Shaping Blocklength at Different Symbol Rates for the Nonlinear Fiber Channel. In *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*, pages 1–3, 2022. Cited pages

[178] Yunus Can Gültekin, Kaiquan Wu, and Alex Alvarado. Temporal Properties of Enumerative Shaping: Autocorrelation and Energy Dispersion Index. In *Optica Advanced Photonics Congress 2022*, page SpM2I.3. Optica Publishing Group, 2022. Cited pages 62 and 98

[179] Yunus Can Gültekin. Enumerative sphere shaping techniques for short blocklength wireless communications. 2020. Cited page 64

[180] T Cover. Enumerative source encoding. *IEEE Transactions on Information Theory*, 19(1):73–77, 1973. Cited page 64

[181] G Welti and Jhong Lee. Digital transmission with coherent four-dimensional modulation. *IEEE Transactions on Information Theory*, 20(4):497–502, 1974. Cited page 67

[182] Martin Sjodin, Erik Agrell, and Magnus Karlsson. Subset-optimized polarization-multiplexed PSK for fiber-optic communications. *IEEE communications letters*, 17(5):838–840, 2013. Cited page 67

[183] Rafael Rios-Müller, Jeremie Renaudier, Laurent Schmalen, and Gabriel Charlet. Joint coding rate and modulation format optimization for 8QAM constellations using BICM mutual information. In *Optical Fiber Communication Conference*, pages W3K–4. Optica Publishing Group, 2015. Cited pages

[184] Qunbi Zhuge, Michael Reimer, Andrew D Shiner, Andrzej Borowiec, Douglas Charlton, Fangyuan Zhang, Meng Qiu, Wei Wang, David V Plant, and Maurice O'Sullivan. Flexible modulation and nonlinear tolerance for coherent transceivers. In *2015 Opto-Electronics and Communications Conference (OECC)*, pages 1–3. IEEE, 2015. Cited page 67

[185] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. Properties of nonlinear noise in long, dispersion-uncompensated fiber links. *Optics Express*, 21(22):25685–25699, 2013. Cited pages 68 and 84

[186] Keisuke Kojima, Toshiaki Koike-Akino, David S Millar, and Kieran Parsons. BICM capacity analysis of 8QAM-alternative modulation formats in nonlinear fiber transmission. In *2015 Tyrrhenian International Workshop on Digital Communications (TIWDC)*, pages 57–59. IEEE, 2015. Cited pages 68 and 71

[187] Bin Chen, Chigo Okonkwo, Hartmut Hafermann, and Alex Alvarado. Polarization-ring-switching for nonlinearity-tolerant geometrically shaped four-dimensional formats maximizing generalized mutual information. *Journal of Lightwave Technology*, 37(14):3579–3591, 2019. Cited pages 68 and 81

[188] Eric Sillekens, Gabriele Liga, Domaniç Lavery, Polina Bayvel, and Robert I Killey. High-cardinality geometrical constellation shaping for the nonlinear fibre channel. *Journal of Lightwave Technology*, 40(19):6374–6387, 2022. Cited page 68

[189] Sebastiaan Goossens, Yunus Can Gültekin, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo Okonkwo, and Alex Alvarado. Introducing 4D Geometric Shell Shaping for Mitigating Nonlinear Interference Noise. *Journal of Lightwave Technology*, 41(2):599–609, 2022. Cited page 68

[190] J Renaudier, A Voicila, O Bertran-Pardo, O Rival, Magnus Karlsson, G Charlet, and S Bigo. Comparison of set-partitioned two-polarization 16QAM formats with PDM-QPSK and PDM-8QAM for optical transmission systems with error-correction coding. In *European Conference and Exhibition on Optical Communication*, pages We–1. Optica Publishing Group, 2012. Cited page 70

[191] Toshiaki Koike-Akino, David S Millar, Keisuke Kojima, and Kieran Parsons. Eight-dimensional modulation for coherent optical communications. In *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*, pages 1–3. IET, 2013. Cited page 70

[192] Keisuke Kojima, David S Millar, Toshiaki Koike-Akino, and Kieran Parsons. Constant modulus 4D optimized constellation alternative for DP-8QAM. In *2014 The European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2014. Cited pages 71 and 78

[193] Keisuke Kojima, Tsuyoshi Yoshida, Toshiaki Koike-Akino, David S Millar, Kieran Parsons, and Valeria Arlunno. 5 and 7 bit/symbol 4D Modulation Formats Based on 2A8PSK. In *ECOC 2016; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016. Cited page 71

[194] Tsuyoshi Yoshida, Keisuke Matsuda, Keisuke Kojima, Hiroshi Miura, Keisuke Dohi, Milutin Pajovic, Toshiaki Koike-Akino, David S Millar, Kieran Parsons, and Takashi Sugihara. Hardware-efficient precise and flexible soft-demapping for multi-dimensional complementary APSK signals. In *ECOC 2016; 42nd European Conference on Optical Communication*, pages 1–3. VDE, 2016. Cited page 71

[195] Alex Alvarado, Erik Agrell, Domaniç Lavery, and Polina Bayvel. LDPC codes for optical channels: Is the "FEC limit" a good predictor of post-FEC BER? In *Optical Fiber Communication Conference*, pages Th3E–5. Optica Publishing Group, 2015. Cited page 72

[196] Alex Alvarado and Erik Agrell. Four-dimensional coded modulation with bit-wise decoders for future optical communications. *Journal of Lightwave Technology*, 33(10):1993–2003, 2015. Cited page 74

[197] Sylvain Almonacil, Fabien Boitier, and Patricia Layec. Performance model and design rules for optical systems employing low-resolution DAC/ADC. *Journal of Lightwave Technology*, 38(11):3007–3014, 2020. Cited pages 74 and 81

[198] Steven A. Jacobs and Robert A. Marsland. Constellation recovery and impairment evaluation through minimization of the blind EVM. *Opt. Express*, 24(19):21708–21721, Sep 2016. Cited pages 74 and 80

[199] Chris R. S. Fludger and Theo Kupfer. Transmitter Impairment Mitigation and Monitoring for High Baud-Rate, High Order Modulation Systems. In *ECOC 2016; 42nd European Conference on Optical Communication*, pages 1–3, 2016. Cited page 74

[200] Junpeng Liang, Yangyang Fan, Zhenning Tao, Xiaofei Su, and Hisao Nakashima. Transceiver Imbalances Compensation and Monitoring by Receiver DSP. *Journal of Lightwave Technology*, 39(17):5397–5404, 2021. Cited pages 74 and 80

[201] Bin Chen, Yi Lei, Gabriele Liga, Zhiwei Liang, Wei Ling, Xuwei Xue, and Alex Alvarado. Geometrically-shaped multi-dimensional modulation formats in coherent optical transmission systems. *Journal of Lightwave Technology*, 41(3):897–910, 2022. Cited page 75

[202] Bin Chen, Alex Alvarado, Sjoerd van der Heide, Menno van den Hout, Hartmut Hafermann, and Chigo Okonkwo. Analysis and experimental demonstration of orthant-symmetric four-dimensional 7 bit/4D-sym modulation for optical fiber communication. *Journal of Lightwave Technology*, 39(9):2737–2753, 2021. Cited page 75

[203] Andre Tkacenko and PP Vaidyanathan. Generalized kurtosis and applications in blind equalization of MIMO channels. In *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat. No. 01CH37256)*, volume 1, pages 742–746. IEEE, 2001. Cited page 76

[204] Fremont OFI Internetworking Forum. Implementation Agreement 400ZR, 2022. Cited page 80

[205] Charles Laperle. Advances in High-Speed ADCs, DACs, and DSP for Optical Transceivers. In *Optical Fiber Communication Conference/National Fiber Optic Engineers Conference 2013*, page OTh1F.5. Optica Publishing Group, 2013. Cited page 82

[206] Xi Chen, Sethumadhavan Chandrasekhar, Sebastian Randel, Wenjun Gu, and Peter Winzer. Experimental Quantification of Implementation Penalties from Limited ADC Resolution for Nyquist Shaped Higher-Order QAM. In *Optical Fiber Communication Conference*, page W4A.3. Optica Publishing Group, 2016. Cited page 82

[207] Olga Vassilieva, Inwoong Kim, and Paparao Palacharla. Practical Considerations for Geometric and Probabilistic Shaping in Optical Communication Systems. In *2023 European Conference on Optical Communication (ECOC)*, page M.B.3.1. IEEE, 2023. Cited page 82

[208] Metodi Yankov, Ognjen Jovanovic, Darko Zibar, and Francesco Da Ros. Rate Adaptive Geometric Constellation Shaping Using Autoencoders and Many-To-One Mapping. In *2023 European Conference on Optical Communication (ECOC)*, page M.B.3.4. IEEE, 2023. Cited page 82

[209] Kadir Gumus, Bin Chen, Thomas Bradley, and Chigo Oknonkwo. A Simplified Method for Optimising Geometrically Shaped Constellations of Higher Dimensionality. In *2023 European Conference on Optical Communication (ECOC)*, page We.A.6.3. IEEE, 2023. Cited page 82

[210] Tobias Fehenberger, Metodi P Yankov, Luca Barletta, and Norbert Hanik. Compensation of xpm interference by blind tracking of the nonlinear phase in wdm systems with qam input. In *2015 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2015. Cited pages 84 and 122

[211] Alex Alvarado, Tobias Fehenberger, Bin Chen, and Frans MJ Willems. Achievable information rates for fiber optics: Applications and computations. *Journal of Lightwave Technology*, 36(2):424–439, 2017. Cited pages 90 and 111

[212] Miao Kong, Junting Shi, Bohan Sang, Junjie Ding, Kaihui Wang, Weiping Li, Feng Wang, Cuiwei Liu, Yanyi Wang, Yiran Wei, et al. 800-Gb/s/carrier WDM Coherent Transmission Over 2000 km Based on Truncated PS-64QAM Utilizing MIMO Volterra Equalizer. *Journal of Lightwave Technology*, 40(9):2830–2839, 2022. Cited page 91

[213] Xueyang Li, Zhenping Xing, Md Samiul Alam, Maxime Jacques, Stéphane Lessard, and David V Plant. NN-based PCS distribution optimization for practical channels. In *2020 European Conference on Optical Communications (ECOC)*, pages 1–4. IEEE, 2020. Cited page 91

[214] Zhongliang Sun, Du Tang, Yingjie Jiang, Zhen Wu, and Yaojun Qiao. Impact of Super-Gaussian Distribution on System Gain of Probabilistic Shaping 64QAM. In *2023 Opto-Electronics and Communications Conference (OECC)*, pages 1–3. IEEE, 2023. Cited page 91

[215] Mengfan Fu, Qiaoya Liu, Yicheng Xu, Hexun Jiang, Xiaobo Zeng, Yiwen Wu, Lilin Yi, Weisheng Hu, and Qunbi Zhuge. Multi-Dimensional Distribution Matching With Bit-Level Shaping for Probabilistically Shaped High Order Modulation Formats. *Journal of Lightwave Technology*, 40(9):2870–2879, 2022. Cited page 91

[216] Yunus Can Gültekin, Alex Alvarado, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo M Okonkwo, and Frans MJ Willems. Kurtosis-limited sphere shaping for nonlinear interference noise reduction in optical channels. *Journal of Lightwave Technology*, 40(1):101–112, 2021. Cited pages 91, 95, 105, 106, 107, 112, 120, and 121

[217] Yunus Can Gültekin, Alex Alvarado, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo M. Okonkwo, and Frans M. J. Willems. Mitigating nonlinear interference by limiting energy variations in sphere shaping. In *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2022. Cited pages 91, 105, 106, 107, 112, and 121

[218] Marcin Pikus and Wen Xu. Bit-Level Probabilistically Shaped Coded Modulation. *IEEE Communications Letters*, 21(9):1929–1932, 2017. Cited page 92

[219] Georg Böcherer, Patrick Schulte, and Fabian Steiner. High Throughput Probabilistic Shaping with Product Distribution Matching, 2017. Cited pages

[220] Maxim Goukhshtein, Stark C. Draper, and Jeebak Mitra. Probabilistic Shaping Using a Block-Based Bit-Level Distribution Matcher. In *2022 17th Canadian Workshop on Information Theory (CWIT)*, pages 68–73, 2022. Cited page 92

[221] Mrinmoy Jana, Lutz Lampe, Jeebak Mitra, and Chuandong Li. Jointly Shaped Dual Polarization Systems. In *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2021. Cited page 92

[222] Zelin Gan, Xiang Li, and Seb J. Savory. Nonlinearity aware bisection-based sphere shaping for optical digital subcarrier multiplexing systems. *Opt. Express*, 30(24):44118–44131, Nov 2022. Cited pages 95 and 97

[223] Mengfan Fu, Qiaoya Liu, Huazhi Lun, Hexun Jiang, Yiwen Wu, Xiaomin Liu, Zhiyuan Yang, Lilin Yi, Weisheng Hu, and Qunbi Zhuge. Parallel bisection-based distribution matching for nonlinearity-tolerant probabilistic shaping in coherent optical communication systems. *Journal of Lightwave Technology*, 39(20):6459–6469, 2021. Cited page 95

[224] Pavel Skvortcov, Ian Phillips, Wladek Forysiak, Toshiaki Koike-Akino, Keisuke Kojima, Kieran Parsons, and David S Millar. Huffman-coded sphere shaping for extended-reach single-span links. *IEEE Journal of Selected Topics in Quantum Electronics*, 27(3):1–15, 2021. Cited pages 95, 120, 122, and 125

[225] Tobias Fehenberger, David S Millar, Toshiaki Koike-Akino, Keisuke Kojima, Kieran Parsons, and Helmut Griesser. Analysis of nonlinear fiber interactions for finite-length constant-composition sequences. *Journal of Lightwave Technology*, 38(2):457–465, 2019. Cited pages 95, 111, and 120

[226] Ronen Dar, Meir Feder, Antonio Mecozzi, and Mark Shtaif. On shaping gain in the nonlinear fiber-optic channel. In *2014 IEEE International Symposium on Information Theory*, pages 2794–2798. IEEE, 2014. Cited page 97

[227] Omri Geller, Ronen Dar, Meir Feder, and Mark Shtaif. A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems. *Journal of Lightwave Technology*, 34(16):3884–3889, 2016. Cited page 97

[228] Junho Cho, Xi Chen, Greg Raybon, Di Che, Ellsworth Burrows, Samuel Olsson, and Robert Tkach. Shaping lightwaves in time and frequency for optical fiber communication. *Nature communications*, 13(1):785, 2022. Cited pages 97 and 98

[229] Yunus Can Gültekin, Wim J van Houtum, Arie GC Koppelaar, Frans MJ Willems, J Wim, et al. Enumerative sphere shaping for wireless communications with short packets. *IEEE Transactions on Wireless Communications*, 19(2):1098–1112, 2019. Cited pages 106, 111, and 122

[230] Yunus Can Gültekin, Kaiquan Wu, and Alex Alvarado. Temporal Properties of Enumerative Shaping: Autocorrelation and Energy Dispersion Index. In *Signal Processing in Photonic Communications*, pages SpM2I–3. Optica Publishing Group, 2022. Cited page 106

[231] Manuel S Neves, Andrea Carena, Antonino Nespola, Paulo P Monteiro, and Fernando P Guiomar. Joint carrier-phase estimation for digital subcarrier multiplexing systems with symbol-rate optimization. *Journal of Lightwave Technology*, 39(20):6403–6412, 2021. Cited page 109

[232] Tobias Fehenberger, Helmut Griesser, and Jörg-Peter Elbers. Mitigating fiber nonlinearities by short-length probabilistic shaping. In *Optical Fiber Communication Conference*, pages Th1I–2. Optica Publishing Group, 2020. Cited page 111

[233] Yunus Can Gültekin, Olga Vassilieva, Inwoong Kim, Paparao Palacharla, Chigo Okonkwo, and Alex Alvarado. On optimum enumerative sphere shaping blocklength at different symbol rates for the nonlinear fiber channel. In *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*, pages 1–3. IEEE, 2022. Cited page 111

[234] Stella Civelli, Enrico Forestieri, and Marco Secondini. Interplay of probabilistic shaping and carrier phase recovery for nonlinearity mitigation. In *2020 European Conference on Optical Communications (ECOC)*, pages 1–4. IEEE, 2020. Cited page 111

[235] Reza Rafie Borujeny and Frank R Kschischang. Why constant-composition codes reduce nonlinear interference noise. *Journal of Lightwave Technology*, 2023. Cited page 111

[236] Jingtian Liu, Elie Awwad, and Yves Jaouen. Multi-dimensional Energy Limitation in Sphere Shaping for Nonlinear Interference Noise Mitigation. In *Asia Communications and Photonics Conference*. IEEE, 2023. Cited pages 112 and 114

[237] Kaiquan Wu, Gabriele Liga, Alireza Sheikh, Yunus Can Gültekin, Frans MJ Willems, and Alex Alvarado. List-encoding CCDM: A nonlinearity-tolerant shaper aided by energy dispersion index. *Journal of Lightwave Technology*, 40(4):1064–1071, 2022. Cited pages 112, 115, and 124

[238] Askari, Mohammad Taha and Lampe, Lutz and Mitra, Jeebak. Nonlinearity tolerant shaping with sequence selection. In *2022 European Conference on Optical Communication (ECOC)*, pages 1–4. IEEE, 2022. Cited page 112

[239] Marco Secondini, Stella Civelli, Enrico Forestieri, and Lareb Zar Khan. New lower bounds on the capacity of optical fiber channels via optimized shaping and detection. *Journal of Lightwave Technology*, 40(10):3197–3209, 2022. Cited pages 112 and 126

[240] Marco Secondini and Enrico Forestieri. Scope and limitations of the nonlinear Shannon limit. *Journal of Lightwave Technology*, 35(4):893–902, 2016. Cited page 113

[241] Erik Agrell. Capacity bounds in optical communications. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited pages

[242] Marco Secondini. Information capacity of optical channels. In *Optical Fiber Telecommunications VII*, pages 867–920. Elsevier, 2020. Cited page 113

[243] Gerhard Kramer, Mansoor I Yousefi, and Frank R Kschischang. Upper bound on the capacity of a cascade of nonlinear and noisy channels. In *2015 IEEE Information Theory Workshop (ITW)*, pages 1–4. IEEE, 2015. Cited page 113

[244] Armin Splett, Christian Kurtzke, and Klaus Petermann. Ultimate transmission capacity of amplified optical fiber communication systems taking into account fiber nonlinearities. In *Proc. ECOC*, volume 2, pages 41–44, 1993. Cited page 114

[245] Partha P Mitra and Jason B Stark. Nonlinear limits to the information capacity of optical fibre communications. *Nature*, 411(6841):1027–1030, 2001. Cited page 114

[246] Antonio Mecozzi and René-Jean Essiambre. Nonlinear Shannon limit in pseudolinear coherent systems. *Journal of Lightwave Technology*, 30(12):2011–2024, 2012. Cited page 114

[247] Ronen Dar, Mark Shtaif, and Meir Feder. New bounds on the capacity of the nonlinear fiber-optic channel. *Optics letters*, 39(2):398–401, 2014. Cited pages

[248] Marco Secondini, Erik Agrell, Enrico Forestieri, and Domenico Marsella. Fiber nonlinearity mitigation in WDM systems: Strategies and achievable rates. In *2017 European Conference on Optical Communication (ECOC)*, pages 1–3. IEEE, 2017. Cited pages

[249] Francisco Javier García-Gómez and Gerhard Kramer. Mismatched models to lower bound the capacity of dual-polarization optical fiber channels. *Journal of Lightwave Technology*, 39(11):3390–3399, 2021. Cited page 114

[250] Kaiquan Wu, Gabriele Liga, Alireza Sheikh, Frans M. J. Willems, and Alex Alvarado. Temporal Energy Analysis of Symbol Sequences for Fiber Nonlinear Interference Modelling via Energy Dispersion Index. *Journal of Lightwave Technology*, 39(18):5766–5782, 2021. Cited pages 115 and 141

[251] Stella Civelli, Emanuele Parente, Enrico Forestieri, and Marco Secondini. On the nonlinear shaping gain with probabilistic shaping and carrier phase recovery. *Journal of Lightwave Technology*, 41(10):3046–3056, 2023. Cited page 117

[252] Georg Böcherer. Probabilistic signal shaping for bit-metric decoding. In *2014 IEEE International Symposium on Information Theory*, pages 431–435, 2014. Cited page 122

[253] Abel Lorences-Riesgo, Manuel S Neves, Celestino S Martins, Sami Mumtaz, Pedro A Loureiro, Yann Frignac, Paulo P Monteiro, Gabriel Charlet, Fernando P Guiomar, and Stefanos Dris. Improving nonlinearity tolerance of pcs-qam digital multi-carrier systems through symbol rate optimization. In *European Conference and Exhibition on Optical Communication*, pages We1C–3. Optica Publishing Group, 2022. Cited page 122

[254] Andrea Carena, Gabriella Bosco, Vittorio Curri, Yanchao Jiang, Pierluigi Poggiolini, and Fabrizio Forghieri. Egn model of non-linear fiber propagation. *Optics express*, 22(13):16335–16362, 2014. Cited page 132

[255] Meng Qiu, Qunbi Zhuge, Mathieu Chagnon, Yuliang Gao, Xian Xu, Mohamed Morsy-Osman, and David V Plant. Digital subcarrier multiplexing for fiber nonlinearity mitigation in coherent optical communication systems. *Optics Express*, 22(15):18770–18777, 2014. Cited page 141

[256] Liang B Du and Arthur J Lowery. Optimizing the subcarrier granularity of coherent optical communications systems. *Optics express*, 19(9):8079–8084, 2011. Cited page 141

[257] William Shieh and Yan Tang. Ultrahigh-speed signal transmission over nonlinear and dispersive fiber optic channel: the multicarrier advantage. *IEEE Photonics Journal*, 2(3):276–283, 2010. Cited page 141

[258] AD Ellis and WA Stallard. Four wave mixing in ultra long transmission systems incorporating linear amplifiers. In *IEE Colloquium on Non-Linear Effects in Fibre Communications*, pages 6–1. IET, 1990. Cited page 141

[259] RW Tkach, AR Chraplyvy, Fabrizio Forghieri, AH Gnauck, and RM Derosier. Four-photon mixing and high-speed WDM systems. *Journal of Lightwave Technology*, 13(5):841–849, 1995. Cited page 141

Publications

Conference proceedings

- **Liu, Jingtian**, Élie Awwad, and Yves Jaouën. "Probabilistic Shaping over Multi-Dimensional Constellations for Optical Fiber Transmissions: Trade-offs and Insights." 2023 Opto-Electronics and Communications Conference (OECC). IEEE, 2023.
- **Liu, Jingtian**, Élie Awwad, and Yves Jaouën. "Multi-dimensional Energy Limitation in Sphere Shaping for Nonlinear Interference Noise Mitigation." 2023 Asia Communications and Photonics Conference/2023 International Photonics and Optoelectronics Meetings (ACP/POEM). IEEE, 2023.
- **Liu, Jingtian**, Élie Awwad, and Yves Jaouën. "Nonlinear Tolerant Sphere Shaping with Sequence Selection." 2024 Opto-Electronics and Communications Conference (OECC). IEEE, 2024.

Journal papers

- **Liu, Jingtian**, Élie Awwad, and Yves Jaouën. "Fault-tolerant four-dimensional constellation for coherent optical transmission systems." Optics Express 31.26 (2023): 43449-43461.
- **Liu, Jingtian**, Élie Awwad, Hartmut Hafermann, and Yves Jaouën. "Sequence Selection with Dispersion-Aware Metric for Long-Haul Transmission Systems." Journal of Lightwave Technology, 10.1109/JLT.2024.3385109.

Titre : Stratégies de mise en forme pour minimiser les effets non linéaires dans les communications par fibres optiques

Mots clés : Communications sur Fibres Optiques, Multiplexage en Polarisation, Non-linéaire Effets Kerr, Mise en forme de constellation probabiliste, Modulations multi-dimensionnelles, Sélection de séquence.

Résumé : Le principal obstacle aux communications longue distance par fibre optique est l'interférence non-linéaire (NLI), résultant des effets non linéaires des fibres optiques. Bien que les algorithmes de traitement numérique des signaux offrent une réduction partielle d'NLI, la nature non-linéaire inhérente à la fibre, associée à des effets de dispersion prédominants, continue à entraver l'accroissement des débits de transmission. Le traitement de ce problème à la source par la conception de nouvelles modulations intrinsèquement moins sensibles aux distorsions non-linéaires est au cœur de notre recherche. À mesure que l'efficacité spectrale augmente, les schémas de modulation traditionnels, tels que les formats QAM, deviennent de plus en plus sensibles à la NLI alors que leur distance Euclidienne minimale diminue. Bien que les modulations multidimensionnelles (MD) permettent d'améliorer les gains linéaires et d'apporter des gains non linéaires partiels, elles n'ont pas encore démontré d'avantages tangibles. D'autre part, l'émergence de la mise en forme de constellation probabiliste (PCS pour Probabilistic Constellation Shaping), préférée pour son gain linéaire accru et sa compatibilité avec la chaîne de traitement numérique utilisée pour les formats de modulation conventionnels, introduit une NLI supplémentaire. Par conséquent, la conception de schémas PCS tolérants à la non-linéarité apparaît comme un axe de recherche crucial. Notre thèse commence par la proposition d'une nouvelle modulation MD pour les signaux uniformément distribués. Ensuite, nous proposons une nouvelle approche combinant les modulations MD et la PCS pour examiner les variations de performance. En ce qui concerne les schémas PCS, nous étudions différentes solutions de transformation de la distribution des signaux (DM pour Distribution Matcher), en tenant compte des dimensions multiples, et concevons un DM optimisé pour la tolérance aux effets non-linéaires sur les liaisons de courte distance. Ensuite, nous définissons une nouvelle technique de mesure de la NLI, qui tient compte des effets de dispersion accumulés. En intégrant cette technique au cadre de sélection des séquences de symboles PCS, nous parvenons à une transmission longue distance réussie avec des gains-non linéaires notables.

Title : Shaping Strategies to Embrace Nonlinear Effects in Optical Fiber Communications

Keywords : Optical Fiber Communications, Polarization Division Multiplexing, Nonlinear Kerr effects, Probabilistic constellation shaping, Multi-dimensional modulations, Sequence selection.

Abstract : The main impediment in long-haul optical fiber communications is nonlinear interference (NLI), stemming from nonlinear effects in optical fibers. While digital signal processing algorithms offer partial mitigation of NLI, the inherent nonlinear nature of the fiber, coupled with predominant dispersion effects, continues to challenge the increase of transmission throughputs. Addressing nonlinearity at the information source through signal modulation technology is at the heart of our research. As spectral efficiency climbs, traditional modulation schemes, such as QAM formats, become increasingly susceptible to NLI while their minimum squared Euclidean distance diminishes. While multi-dimensional (MD) modulations yield improved linear and partial nonlinear gains, it has not yet demonstrated tangible benefits. On the other hand, the emergence of probabilistic constellation shaping (PCS), preferred for its enhanced linear gain and compatibility with conventional modulation hardware and software, introduces additional NLI. Consequently, the design of nonlinear-tolerant PCS schemes is emerging as a pivotal research direction. Our thesis begins with the proposal of a novel MD modulation for uniformly distributed signals. Then, we propose a novel approach combining MD modulations with PCS to examine performance variations. Delving into PCS schemes, we investigate the enumerative sphere shaping distribution matcher (DM), initially from a multi-dimensional stance, and design a DM optimized for nonlinear tolerance over short-distance links. Subsequently, we introduce a new NLI measurement technique, accounting for accumulated dispersion effects. Integrating this with the sequence selection framework of PCS, we achieve successful long-distance transmission with notable nonlinear gains.